NCC-WCH

Version 2.0

Preterm labour and birth

Full guideline

NICE Guideline <…>

Methods, evidence and recommendations

1 June 2015

Draft for Consultation

Commissioned by the National Institute for Health and Care Excellence

Disclaimer
Healthcare professionals are expected to take NICE clinical guidelines fully into account when exercising their clinical judgement. However, the guidance does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of each patient, in consultation with the patient and/or their guardian or carer.

Funding
Registered charity no. 213280

Contents

Introduction

Preterm birth is the single biggest cause of neonatal mortality and morbidity in the UK, affecting over 52,000 babies (around 7.3% of live births) in England and Wales in 2012. There has been no decline in the UK preterm birth rate over the last 10 years.

Babies born preterm have high rates of early, late and postneonatal mortality, with the risk of mortality being inversely proportional to gestational age at birth. Babies who survive have increased rates of disability. Recent UK studies comparing cohorts born in 1995 and 2006 have shown improved rates of survival (from 40% to 53%) for extreme preterm births (born between 22 and 26 weeks). Rates of disability amongst survivors are largely unchanged over this time period, with rates of bronchopulmonary dysplasia, major cerebral scan abnormality, and weight and/or head circumference less than 2 standard deviations being maintained at 68%, 13% and 44% and 23% respectively, although there has been an increase in the proportion treated for retinopathy of prematurity from 13% to 22% (Costeloe 2012).

The major long term consequence of prematurity is neurodevelopmental disability. This can range from severe motor abnormalities such as cerebral palsy, through to less severe cognitive abnormalities (MacKay DF 2010). Although the risk for the individual child is greatest for those born at the earliest gestations, the global burden of neurodevelopmental disabilities is driven by the number of babies born at each of these gestations, and is therefore greatest for babies born between 32 and 36 weeks, less for those born between 28 and 31 weeks, and least for those born at less than 28 weeks gestation (Blencowe 2013) .

Around 75% of women delivering preterm do so after preterm labour. In the majority of women with preterm labour, a “cause” is not found, although it is known that a significant proportion of preterm labours are associated with infection. Preterm labour may or may not be preceded by preterm prelabour membrane rupture. The remaining women delivering preterm have undergone elective or iatrogenic preterm delivery when this is thought to be in the fetal or maternal interest (e.g. because of extreme growth restriction in the baby, or because of maternal conditions such as pre-eclampsia).

“Treatments” for preterm labour include strategies to reduce the risk in women who are at high risk, tocolytics to delay preterm delivery, and additional antenatal strategies to improve outcomes for babies who will be born preterm.

Risk reduction strategies include the use of progesterone prophylaxis (for which there has been an increase in interest since the early part of this century) and cervical cerclage.

Tocolytics are used to stop uterine contractions. However, there is considerable variation in practice, and there is little agreement whether an attempt to delay delivery will improve outcomes for the baby.

Antenatal strategies to improve outcomes for babies who will be born preterm include the common practice of using prenatal steroids to improve lung maturation. A more recent development is the use of magnesium sulfate administered to the mother for neuroprotection of the baby. Again, there is considerable variation of practice around this latter agent, and little consensus about the subgroup of babies who might benefit.

This guideline reviews the evidence for the care of women who present with signs and symptoms of preterm labour, and those who are scheduled to have a preterm birth. It also reviews how preterm birth can be optimally diagnosed in symptomatic women, given that many women thought to be in preterm labour on a clinical assessment will not deliver preterm. Optimal diagnosis can facilitate transfer to a place where appropriate neonatal intensive care can be provided, a strategy known to improve rates of survival for the baby. Additional areas that will be covered by the guidance (such as information needs for women

who presents with signs and symptoms of preterm labour) are outlined in the guideline scope.

The guideline does not cover who should and should not have medically indicated preterm delivery, nor diagnostic or predictive tests in asymptomatic women. These issues have been reviewed in other NICE guidelines (Diabetes in Pregnancy, Hypertensive Disease in Pregnancy and Antenatal Care) or will be covered in the forthcoming NICE guideline on High Risk Intrapartum care.

1 Guideline summary

1.1 Guideline Committee membership, NCC-WCH staff and acknowledgements

1.1.1 Guideline Committee membership

Name	Role
Judi Barratt	Clinical midwife specialist, Worcester Royal Hospital
Paul Eunson	Consultant Paediatric Neurologist & Honorary Senior Lecturer, Royal Hospital for Sick Children, Edinburgh
Jane Hawdon	Consultant Neonatologist, Barts Health NHS Trust
Jane Norman (Chair)	Professor of Maternal and Fetal Health, Director of the Tommy's Centre for Maternal and Fetal Health, University of Edinburgh MRC Centre for Reproductive Health Queen's Medical Research Institute
Philip Owen	Consultant Obstetrician and Gynaecologist, North Glasgow NHS Trust
Jane Plumb	Lay member
Farrah Pradhan	Lay member
Marianne Rowntree	Midwife, Plymouth NHS Hospitals Trust
Meekai To	Consultant in fetal medicine and obstetrics, Kings College Hospital
Martin Ward Platt	Consultant Paediatrician (neonatal medicine), The Newcastle upon Tyne Hospitals
Louise Weaver-Lowe	Neonatal Nurse, Central Manchester University Hospitals NHS Trust

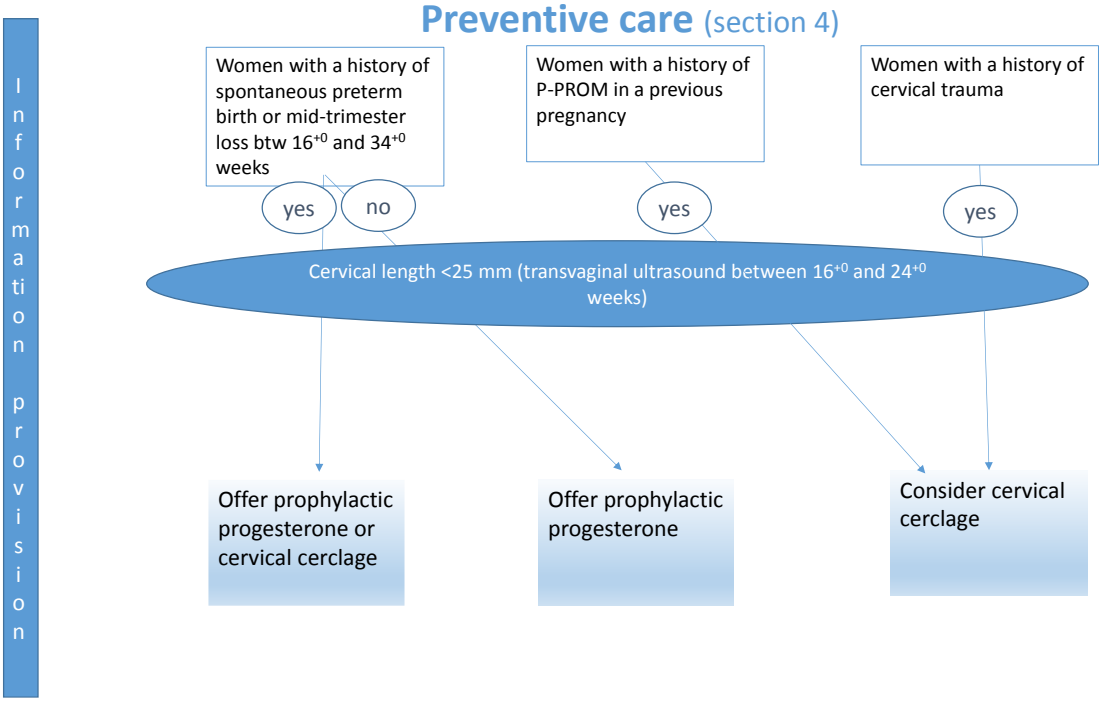
1.1.2 NCC-WCH staff

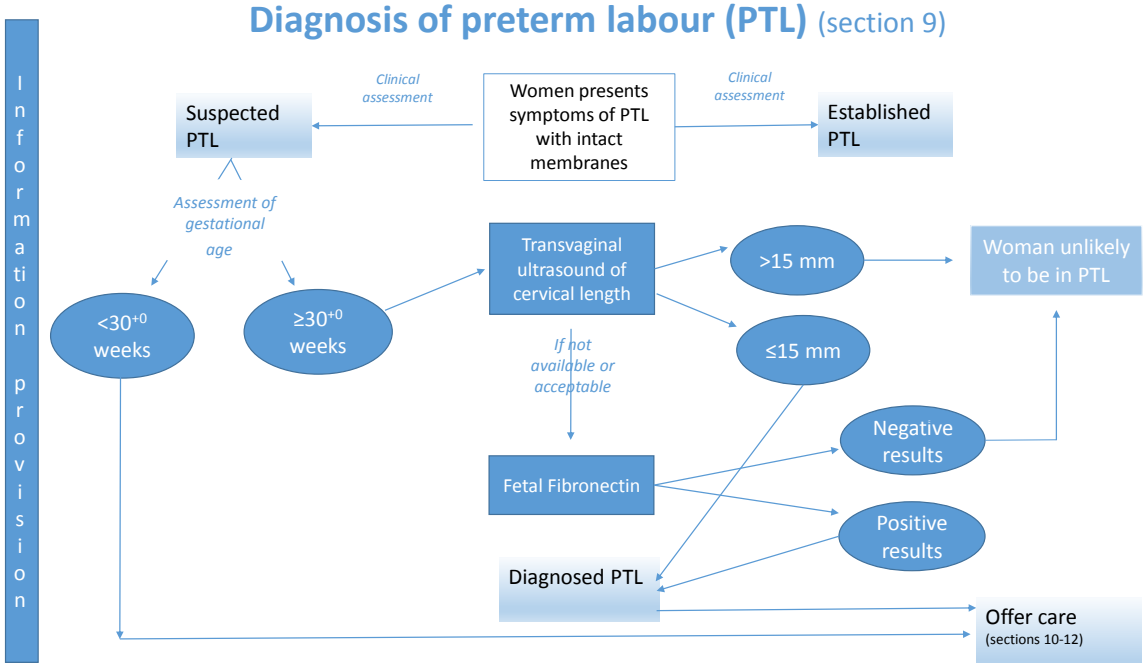
Name	Role
Ebenezer Ademisoye	Health Economist (from February 2015)
Shona Burman-Roy	Senior Research Fellow
Zosia Beckles	Information Scientist (from October 2014)
Liz Bickerdike	Research Assistant (until September 2013)
Anne Carty	Project Manager (from March 2015)
Melanie Davies	Clinical Director for Women's Health (from December 2014)
Maryam Gholitabar	Research Fellow
Paul Jacklin	Senior Research Fellow – Health Economist
David James	Clinical Director for Women's Health (until November 2014)
Juliet Kenny	Project Manager (until March 2015)
Rosalind Lai	Information Scientist (until October 2014)
Hugo Pedder	Statistician (from September 2014)
Grammati Sarri	Senior Research Fellow – Guideline Lead (from October 2014)
Roz Ullman	Senior Research Fellow and Clinical Lead for Midwifery (until May 2014)
Amy Wang	Research Fellow (from January 2015)

1.1.3 Acknowledgements

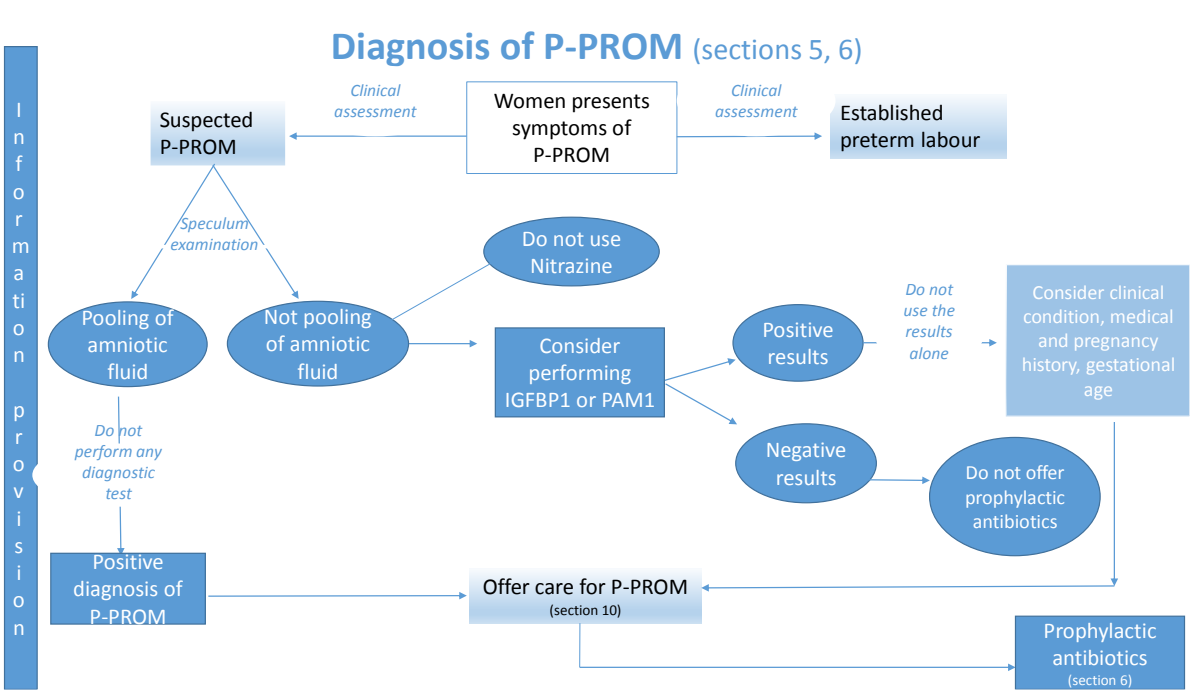
Additional support was received from Ed Peston, Wahab Bello, Sarah Bailey, and Sofia Dias.

1.2 Care pathway/Algorithm



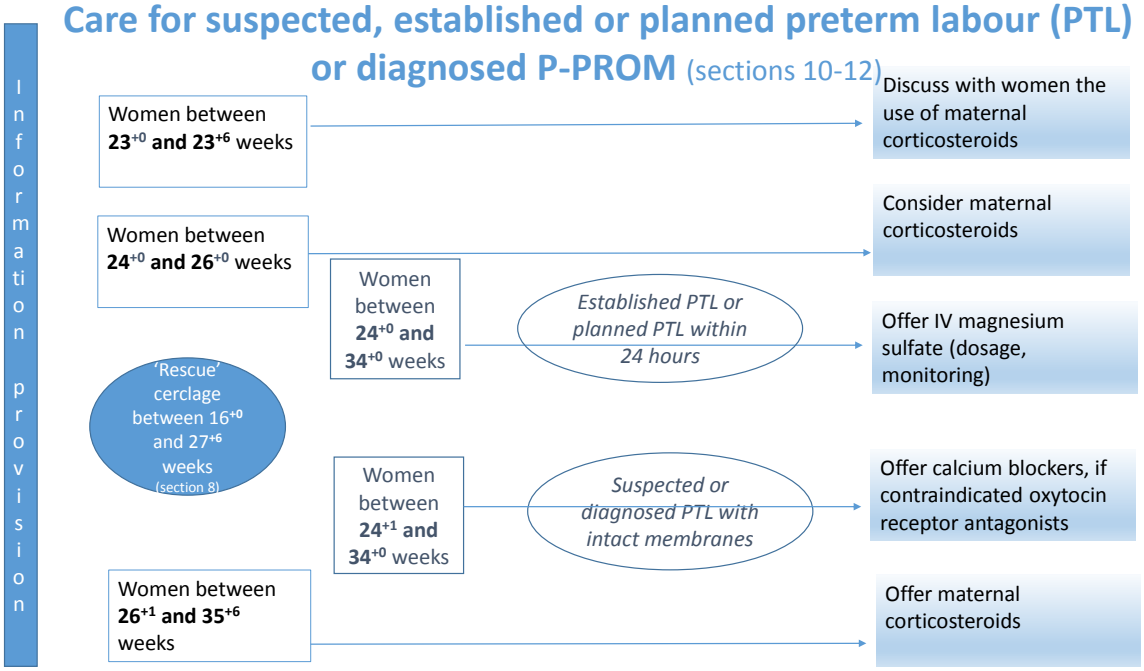


1

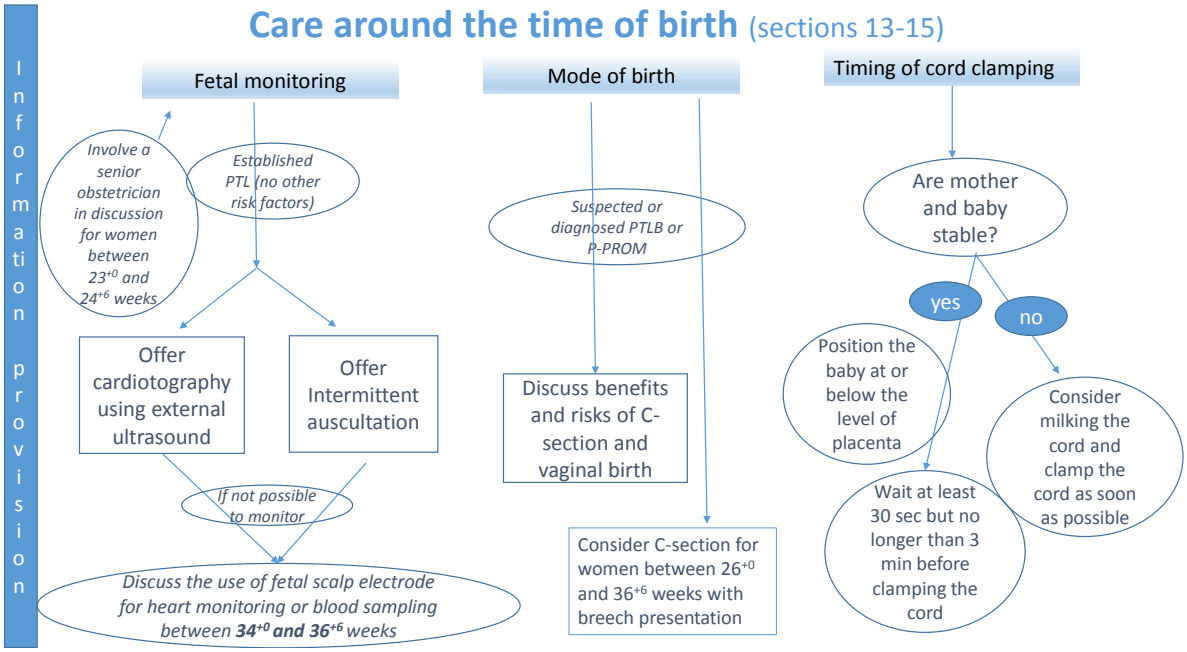


2

3



4



1.3 Recommendations

1. When giving information and support to women at increased risk of preterm labour, with suspected, diagnosed or established preterm labour, or having a planned preterm birth (and their family members or carers as appropriate):
    - give this information and support as early as possible, taking into account the likelihood of preterm birth and the status of labour
    - follow the principles in the NICE guideline on patient experience in adult NHS services
    - give both oral and written information
    - describe the symptoms and signs of preterm labour
    - explain to the woman about the care she might be offered.
2. For women who are having a planned preterm birth or are offered treatment for preterm labour in line with recommendations 30 – 43 (and their family members or carers as appropriate), provide information and support that includes:
    - information about the likelihood of the baby surviving and other outcomes (including long-term outcomes) and risks for the baby, giving values as natural frequencies (for example, 1 in 100)
    - explaining about the neonatal care of preterm babies, including location of care
    - explaining about the immediate problems that can arise when a baby is born preterm
    - explaining about the possible long-term consequences of prematurity for the baby (how premature babies grow and develop)

- ongoing opportunities to talk about and state their wishes about resuscitation of the baby
- an opportunity to tour the neonatal unit
- an opportunity to speak to a neonatologist or paediatrician.

3. Offer a choice of either prophylactic vaginal progesterone or prophylactic cervical cerclage to women:
   - with a history of spontaneous preterm birth or mid-trimester loss between 16^{+0} and 34^{+0} weeks of pregnancy and
   - in whom a transvaginal ultrasound scan has been carried out between 16^{+0} and 24^{+0} weeks of pregnancy that reveals a cervical length of less than 25 mm.

   Discuss the benefits and risks of prophylactic progesterone and cervical cerclage with the woman and take her preferences into account.
4. Offer prophylactic vaginal progesterone to women with no history of spontaneous preterm birth or mid-trimester loss in whom a transvaginal ultrasound scan has been carried out between 16^{+0} and 24^{+0} weeks of pregnancy that reveals a cervical length of less than 25 mm.
5. Consider prophylactic cervical cerclage for women in whom a transvaginal ultrasound scan has been carried out between 16^{+0} and 24^{+0} weeks of pregnancy that reveals a cervical length of less than 25 mm and who have either:
   - had preterm prelabour rupture of membranes (P-PROM) in a previous pregnancy or
   - a history of cervical trauma.
6. In a woman reporting symptoms suggestive of preterm prelabour rupture of membranes (P-PROM), offer a speculum examination to look for pooling of amniotic fluid and:
   - if pooling of amniotic fluid is observed, do not perform any diagnostic test but offer care consistent with the woman having P-PROM (see Chapters 6, 7 and 10)
   - if pooling of amniotic fluid is not observed, consider performing an insulin-like growth factor binding protein-1 test or placental alpha-microglobulin-1 test of vaginal fluid.
7. If the results of the insulin-like growth factor binding protein-1 or placental alpha-microglobulin-1 test are positive, do not use the test results alone to decide what care to offer the woman, but also take into account her clinical condition, her medical and pregnancy history and gestational age, and either:
   - offer care consistent with the woman having P-PROM (see Chapters 6, 7 and 10) or
   - re-evaluate the woman's diagnostic status at a later time point.
8. If the results of the insulin-like growth factor binding protein-1 or placental alpha-microglobulin-1 test are negative and no amniotic fluid is observed:
   - do not offer antenatal prophylactic antibiotics
   - explain to the woman that it is unlikely that she has P-PROM, but that she should return if she has any further symptoms suggestive of P-PROM or preterm labour.

9. Do not use nitrazine to diagnose P-PROM.
10. Do not perform diagnostic tests for P-PROM if labour becomes established in a woman reporting symptoms suggestive of P-PROM.
11. Offer women with P-PROM oral erythromycin 250 mg 4 times a day[a] for a maximum of 10 days or until the woman is in established labour (whichever is sooner).
12. For women with P-PROM who cannot tolerate erythromycin or in whom erythromycin is contraindicated, consider oral penicillin for a maximum of 10 days or until the woman is in established labour (whichever is sooner).
13. Do not offer women with P-PROM co-amoxiclav as prophylaxis for intrauterine infection.
14. For guidance on the use of intrapartum antibiotics, see the NICE guideline on Antibiotics for early-onset neonatal infection.
15. Use a combination of clinical assessment and biomedical tests to diagnose intrauterine infection in women with P-PROM.
16. Do not use any of the following in isolation to confirm or exclude intrauterine infection in women with P-PROM:
    - a single test of C-reactive protein
    - white blood cell count
    - cardiotocography.
17. If the results of the clinical assessment or any of the biomedical tests are not consistent with each other, continue to observe the woman and consider repeating the tests.
18. Consider 'rescue' cervical cerclage for women between 16^{+0} and 27^{+6} weeks of pregnancy with a dilated cervix and exposed, unruptured fetal membranes.
19. Do not offer 'rescue' cervical cerclage to women with signs of infection, active vaginal bleeding or uterine contractions.
20. When deciding whether to offer 'rescue' cervical cerclage:
    - take into account gestational age and the extent of cervical dilatation
    - discuss with a consultant obstetrician and consultant paediatrician.
21. Explain to women for whom 'rescue' cervical cerclage is being considered (and their family members or carers as appropriate):
    - about the risks of the procedure
    - that it aims to delay the birth, and so increase the likelihood of the baby surviving and of reducing serious neonatal morbidity.
22. Explain to women reporting symptoms of preterm labour who have intact membranes (and their family members or carers as appropriate):

a At the time of consultation (June 2015), erythromycin did not have a UK marketing authorisation for use in pregnancy. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Prescribing guidance: prescribing unlicensed medicines for further information. The summaries of product characteristics for oral erythromycin recommend different dosages. The evidence reviewed for the guideline supports a dosage of 250 mg 4 times a day for prophylaxis in women with P-PROM.

- about the clinical assessment and diagnostic tests that are available
- how the clinical assessment and diagnostic tests are carried out
- what the benefits, risks and possible consequences of the clinical assessment and diagnostic tests are, including the consequences of false positive and false negative test results taking into account gestational age.

23. Offer a clinical assessment to women reporting symptoms of preterm labour who have intact membranes. This should include:
    - clinical history taking
    - the observations described for the initial assessment of a woman in labour in recommendation 1.4.2 of the NICE guideline on intrapartum care
    - a speculum examination (followed by a digital vaginal examination[b] if the extent of cervical dilatation cannot be assessed).
24. If the clinical assessment suggests that the woman is in suspected preterm labour and she is 29^{+6} weeks pregnant or less, advise treatment for preterm labour as described in Chapters 10 – 12.
25. If the clinical assessment suggests that the woman is in suspected preterm labour and she is 30^{+0} weeks pregnant or more, consider transvaginal ultrasound measurement of cervical length as a diagnostic test to determine likelihood of birth within 48 hours. Act on the results as follows:
    - if cervical length is more than 15 mm, explain to the woman that it is unlikely that she is in preterm labour and:
        - discuss with her the benefits and risks of going home compared with continued monitoring and treatment in hospital
        - advise her that if she does decide to go home, she should return if symptoms suggestive of preterm labour recur
    - if cervical length is 15 mm or less, view the woman as being in diagnosed preterm labour and offer treatment as described in Chapters 10–12.
26. Consider fetal fibronectin testing as a diagnostic test to determine likelihood of birth within 48 hours for women who are 30^{+0} weeks pregnant or more if transvaginal ultrasound measurement of cervical length is indicated but is not available or not acceptable. Act on the results as follows:
    - if fetal fibronectin testing is negative, explain to the woman that it is unlikely that she is in preterm labour and:
        - discuss with her the benefits and risks of going home compared with continued monitoring and treatment in hospital
        - advise her that if she does decide to go home, she should return if symptoms suggestive of preterm labour recur

b Be aware that if a swab for fetal fibronectin testing is anticipated (see recommendation 26), the swab should be taken before any digital vaginal examination.

- if fetal fibronectin testing is positive, view the woman as being in diagnosed preterm labour and offer treatment as described in Chapters 10 – 12.

27. If a woman in suspected preterm labour who is 30^{+0} weeks pregnant or more does not have transvaginal ultrasound measurement of cervical length or fetal fibronectin testing to exclude preterm labour, offer treatment consistent with her being in diagnosed preterm labour (see Chapters 10 - 12).
28. Do not use transvaginal ultrasound measurement of cervical length and fetal fibronectin testing in combination to diagnose preterm labour.
29. Ultrasound scans should be performed by healthcare professionals with training in, and experience of, transvaginal ultrasound measurement of cervical length.
30. For women between 23^{+0} and 23^{+6} weeks of pregnancy who are in suspected or established preterm labour, are having a planned preterm birth or have P-PROM (see Chapter 5), discuss with the woman (and her family members or carers as appropriate) the use of maternal corticosteroids in the context of her individual circumstances.
31. Consider maternal corticosteroids for women between 24^{+0} and 26^{+0} weeks of pregnancy who are in suspected or established preterm labour, are having a planned preterm birth or have P-PROM (see Chapter 5).
32. Offer maternal corticosteroids to women between 26^{+1} and 35^{+6} weeks of pregnancy who are in suspected, diagnosed or established preterm labour, are having a planned preterm birth or have P-PROM.
33. When offering or considering maternal corticosteroids, discuss with the woman (and her family members or carers as appropriate):
    - how corticosteroids may help
    - the potential risks associated with them.
34. Do not routinely offer repeat courses of maternal corticosteroids, but take into account:
    - whether the interval since the end of last course is more than 10 weeks
    - gestational age
    - the likelihood of birth within 48 hours.
35. Offer intravenous magnesium sulfate for neuroprotection of the baby to women between 24^{+0} and 34^{+0} weeks of pregnancy who are:
    - in established preterm labour or
    - having a planned preterm birth within 24 hours.
36. Give a 4 g intravenous bolus of magnesium sulfate over 15 minutes, followed by an intravenous infusion of 1 g per hour until the birth or for 24 hours (whichever is sooner).
37. For women on magnesium sulfate, monitor for clinical signs of magnesium toxicity at least every 4 hours by recording pulse, blood pressure, respiratory rate and deep tendon (for example, patellar) reflexes.
38. If a woman has or develops oliguria or other signs of renal failure:
    - monitor more frequently for magnesium toxicity
    - think about reducing the dose of magnesium sulfate.

39. Take the following factors into account when making a decision about whether to start tocolysis:
    - whether the woman is in suspected or diagnosed preterm labour
    - other clinical features (for example, bleeding or infection) which might suggest that stopping labour is contraindicated
    - gestational age at presentation
    - likely benefit of maternal corticosteroids (see Chapter 10)
    - availability of neonatal care (need for transfer to another unit)
    - the preference of the woman.
40. Offer calcium channel blockers for tocolysis[c] to women between 24^{+1} and 34^{+0} weeks of pregnancy who have intact membranes and are in suspected or diagnosed preterm labour.
41. If calcium channel blockers are contraindicated, offer oxytocin receptor antagonists for tocolysis.
42. Be aware that there is an absence of evidence about all tocolytic medicines before 26^{+0} weeks of pregnancy.
43. Do not offer betamimetics for tocolysis.
44. Discuss with women in suspected, diagnosed or established preterm labour (and their family members or carers as appropriate):
    - the purpose of fetal monitoring
    - the clinical decisions it informs at different gestational ages
    - if appropriate, the option not to monitor the fetal heart rate (for example, at the threshold of viability).
45. Involve a senior obstetrician in discussions about whether and how to monitor the fetal heart rate in women between 23^{+0} and 24^{+6} weeks of pregnancy.
46. Explain the different fetal monitoring options to the woman (and her family members or carers as appropriate), being aware that:
    - there is limited evidence about the usefulness of specific cardiotocography features suggestive of hypoxia or acidosis in preterm babies
    - the available evidence is broadly consistent with that for babies born at term (see section 1.10 in the NICE guideline on intrapartum care)
    - a normal cardiotocography trace is reassuring and indicates that the baby is coping well with labour, but an abnormal trace does not necessarily indicate that fetal hypoxia or acidosis is present.
47. Explain to the woman (and her family members or carers as appropriate) that there is an absence of evidence that using cardiotocography improves the outcomes of preterm labour for the woman or the baby compared with intermittent auscultation.

c Although this use is common in UK clinical practice, at the time of consultation (June 2015), calcium channel blockers did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Prescribing guidance: prescribing unlicensed medicines further information

48. Offer women in established preterm labour but with no other risk factors (see section 1.10 in the NICE guideline on intrapartum care) fetal heart rate monitoring using either:
    - cardiotocography using external ultrasound or
    - intermittent auscultation.

    Take the woman's preferences into account when deciding on choice of monitoring option.
49. For guidance on using intermittent auscultation for fetal heart rate monitoring, see recommendation 1.10.1 in the NICE guideline on intrapartum care.
50. Do not use a fetal scalp electrode for fetal heart rate monitoring if the woman is less than 34^{+0} weeks pregnant unless all of the following apply:
    - it is not possible to monitor the fetal heart rate using either external cardiotocography or intermittent auscultation
    - it has been discussed with a senior obstetrician
    - the benefits are likely to outweigh the potential risks
    - the alternatives (immediate birth, intermittent ultrasound and no monitoring) have been discussed with the woman and are unacceptable to her.
51. Discuss with the woman (and her family members or carers as appropriate) the possible use of a fetal scalp electrode between 34^{+0} and 36^{+6} weeks of pregnancy if it is not possible to monitor the fetal heart rate using either external cardiotocography or intermittent auscultation.
52. Do not carry out fetal blood sampling if the woman is less than 34^{+0} weeks pregnant.
53. Discuss with the woman the possible use of fetal blood sampling between 34^{+0} and 36^{+6} weeks of pregnancy if the benefits are likely to outweigh the potential risks.
54. When offering fetal blood sampling, discuss this with the woman as described in recommendation 1.10.41 in the NICE guideline on intrapartum care, and advise her that if a blood sample cannot be obtained a caesarean section is likely.
55. Discuss the general benefits and risks of caesarean section and vaginal birth with women in suspected or diagnosed preterm labour and women with P-PROM (and their family members or carers as appropriate) - see recommendation 1.1.2.1 in the NICE guideline on caesarean section.
56. Explain to women in suspected or diagnosed preterm labour and women with P-PROM about the benefits and risks of caesarean section that are specific to gestational age. In particular, highlight the difficulties associated with performing a caesarean section for a preterm birth, especially the increased likelihood of a vertical uterine incision and the implications of this for future pregnancies.
57. Explain to women in suspected or diagnosed preterm labour that there are no known benefits or harms for the baby from caesarean section, but the evidence is very limited.
58. Consider caesarean section for women presenting in suspected or diagnosed preterm labour between 26^{+0} and 36^{+6} weeks of pregnancy with breech presentation, and explain to the woman that:

- caesarean section for breech presentation for preterm babies is common but not universal practice
- this practice is based on an extrapolation of evidence of best management for breech presentation for babies born at term
- there is some evidence that there may be a large reduction in perinatal mortality associated with caesarean section for preterm babies with breech presentation, but overall the evidence is inconclusive.

59. If a preterm baby needs to be moved away from the mother for resuscitation, or there is significant maternal bleeding:
    - consider milking the cord and
    - clamp the cord as soon as possible.
60. Wait at least 30 seconds, but no longer than 3 minutes, before clamping the cord of preterm babies if the mother and baby are stable.
61. Position the baby at or below the level of the placenta before clamping the cord.

1.4 Key research recommendations

1.4.1 Prophylactic cervical cerclage compared with prophylactic vaginal progesterone for preventing preterm birth

What is the clinical effectiveness of prophylactic cervical cerclage alone compared with prophylactic vaginal progesterone alone and with both strategies together for preventing preterm birth in women with a short cervix and a history of spontaneous preterm birth?

Why this is important

Preterm birth causes significant neonatal morbidity and mortality, as well as long-term disability. Therefore strategies for preventing preterm birth are important. There are recognised risk factors for preterm birth, and so interventions can be offered to women with these risk factors. Both prophylactic cervical cerclage and prophylactic vaginal progesterone are effective in preventing preterm birth in women with a short cervix and a history of preterm birth, but there is limited evidence on which is more effective, and the relative risks and benefits (including costs) of each. More randomised research is needed to compare the relative effectiveness of prophylactic cervical cerclage and prophylactic vaginal progesterone in improving both neonatal and maternal outcomes. This will help women and healthcare professionals to make an informed decision about which is the most effective prophylactic option.

1.4.2 Diagnosing preterm prelabour rupture of membranes (P-PROM)

What is the diagnostic accuracy and utility of tests (placental alpha-microglobulin-1, insulin-like growth factor binding protein-1, fetal fibronectin, panty-liner with polymer-embedded strip) for diagnosing P-PROM?

Why this is important

P-PROM is relatively common. In the absence of clear pooling of amniotic fluid in the vagina, clinical assessment cannot be conclusive about the diagnosis. There is limited evidence about the accuracy of diagnostic tests (placental alpha-microglobulin-1, insulin-like growth factor binding protein-1, fetal fibronectin, panty-liner with polymer-embedded strip), and the

results of available studies are inconclusive. Making the correct diagnosis is important, because women with a true positive diagnosis or a false negative diagnosis could benefit from prophylactic antibiotics, whereas women with a false positive diagnosis (who have intact fetal membranes) could be harmed by inappropriate use of prophylactic antibiotics. More research on the diagnostic accuracy of the various tests should evaluate both the performance of the tests themselves and their impact on management and outcome. Studies should include subgroup analysis broken down by different gestational ages.

1.4.3 Identifying infection in women with preterm prelabour rupture of membranes (P-PROM)

What is the diagnostic accuracy of serial C-reactive protein testing to identify chorioamnionitis in women with P-PROM?

Why this is important

Identifying infection in women with P-PROM is needed to allow appropriate management. Early diagnosis of infection allows consideration of therapeutic strategies (including antibiotics and/or early birth). Effective treatment of infection is particularly important given that sepsis is a common direct cause of maternal death. There is currently limited evidence that serial C-reactive protein testing might be useful, but the Committee is aware that this strategy is in common practice. Evidence from diagnostic studies is needed about the accuracy of serial C-reactive protein testing for identifying chorioamnionitis, which is one of the most common and serious infective complications of P-PROM.

1.4.4 Effectiveness of 'rescue' cerclage

What is the clinical effectiveness of 'rescue' cerclage in improving outcomes for women at risk of preterm birth?

Why this is important

There is some evidence from randomised studies that 'rescue' cerclage might be effective in improving neonatal outcomes in women with a dilated cervix and exposed, unruptured fetal membranes. However, there is uncertainty about the magnitude of this effect. The full consequences of this strategy and the subgroups of women at risk of preterm labour who might particularly benefit are not known. A randomised controlled trial would best address this question, but a national registry of the most critical outcomes (neonatal mortality and morbidity, maternal morbidity) could also be considered for women who did not want to participate in a randomised trial but who opted for 'rescue' cerclage.

1.4.5 Magnesium sulfate for neuroprotection: bolus plus infusion compared with bolus alone

What is the clinical effectiveness of a bolus plus infusion of magnesium sulfate compared with a bolus alone for preventing neurodevelopmental injury in babies born preterm?

Why this is important

There is evidence from randomised studies that magnesium sulfate has neuroprotective properties for the baby when given to women who will deliver preterm up to 34^{+0} weeks of pregnancy. However, there is uncertainty about the best method of administering magnesium sulfate for this purpose, with different studies using different strategies. There are significant advantages for the woman and for reducing healthcare costs if a bolus is as effective as a bolus plus infusion, because magnesium sulfate has side effects for the woman, and more monitoring is needed for infusion, with additional associated healthcare costs. A randomised

controlled trial would best address this question by assessing the effects of each method on neonatal and maternal outcomes.

1.5 Research recommendations

1. What is the clinical effectiveness of prophylactic cervical cerclage alone compared with prophylactic vaginal progesterone alone and with both strategies together for preventing preterm birth in women with a short cervix and a history of spontaneous preterm birth?
2. What is the diagnostic accuracy and utility of tests (placental alpha-microglobulin-1, insulin-like growth factor binding protein-1, fetal fibronectin, panty-liner with polymer-embedded strip) for diagnosing P-PROM?
3. What is the diagnostic accuracy of serial C-reactive protein testing to identify chorioamnionitis in women with P-PROM?
4. What is the clinical effectiveness of 'rescue' cerclage in improving outcomes for women at risk of preterm birth?
5. What is the clinical effectiveness of a bolus plus infusion of magnesium sulfate compared with a bolus alone for preventing neurodevelopmental injury in babies born preterm?
6. Is intermittent auscultation or electronic fetal monitoring effective in the preterm fetus?
7. Is there any advantage to preterm babies from delayed versus early cord clamping, or cord milking?

2 Guideline development methodology

2.1 Development of the guideline

2.1.1 What is a NICE clinical guideline?

NICE clinical guidelines are recommendations for the care of individuals in specific clinical conditions or circumstances within the NHS – from prevention and self-care through primary and secondary care to more specialised services. We base our clinical guidelines on the best available research evidence, with the aim of improving the quality of health care. We use predetermined and systematic methods to identify and evaluate the evidence relating to specific review questions.

NICE clinical guidelines can:

- provide recommendations for the treatment and care of people by health professionals
- be used to develop standards to assess the clinical practice of individual health professionals
- be used in the education and training of health professionals
- help patients to make informed decisions
- improve communication between patient and health professional

While guidelines assist the practice of healthcare professionals, they do not replace their knowledge and skills. We produce our guidelines using the following steps:

- guideline topic is referred to NICE from the Department of Health
- stakeholders register an interest in the guideline and are consulted throughout the development process.
- the scope is prepared by the National Collaborating Centre for Women's and Children's Health (NCC-WCH).
- the NCC-WCH establishes a Guideline Committee.
- a draft guideline is produced after the Committee assesses the available evidence and makes recommendations.
- there is a consultation on the draft guideline.
- the final guideline is produced.

The NCC-WCH and NICE produce a number of versions of this guideline:

- the 'full guideline' contains all the recommendations, together with details of the methods used and the underpinning evidence
- the 'short guideline' lists the recommendations
- 'information for the public' is written using suitable language for people without specialist medical knowledge
- NICE Pathways brings together all connected NICE guidance.

2.1.2 Remit

NICE received the remit for this guideline from the Department of Health. They commissioned the NCC-WCH to produce the guideline.

The remit for this guideline is to develop a clinical guideline on preterm labour and birth.

2.1.3 Who developed this guideline?

A multidisciplinary Guideline Committee comprising of health professionals and researchers as well as lay members developed this guideline (see the list of Guideline Committee members).

The National Institute for Health and Care Excellence (NICE) funds the NCC-WCH and thus supported the development of this guideline. The Guideline Committee was convened by the NCC-WCH and chaired by Professor Jane Norman in accordance with guidance from NICE.

The group met every 4 to 6 weeks during the development of the guideline. At the start of the guideline development process all Guideline Committee members declared interests including consultancies, fee-paid work, share-holdings, fellowships and support from the healthcare industry. At all subsequent Guideline Committee meetings, members declared arising conflicts of interest.

Members were either required to withdraw completely or for part of the discussion if their declared interest made it appropriate. The details of declared interests and the actions taken are shown in Appendix C.

Staff from the NCC-WCH provided methodological support and guidance for the development process. The team working on the guideline included a project manager, systematic reviewers, health economists and information scientists. They undertook systematic searches of the literature, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate and drafted the guideline in collaboration with the Guideline Committee.

2.1.4 What this guideline covers

2.1.4.1 Groups that will be covered

- Pregnant women who are considered to be at risk of preterm labour and birth because they have a history of:
  - spontaneous preterm birth
  - preterm pre-labour rupture of membranes
  - mid-trimester loss
  - cervical trauma (including surgery – for example, previous cone biopsy [cold knife or laser], large loop excision of the transformation zone [LLETZ – any number] and radical diathermy)
- pregnant women who are considered to be at risk of preterm labour and birth because they have a short cervix that has been identified on ultrasound scan and/or bulging membranes in the current pregnancy.
- pregnant women with preterm pre-labour rupture of membranes
- pregnant women clinically suspected to be in preterm labour
- women diagnosed to be in spontaneous preterm labour
- women having a planned preterm birth

2.1.4.2 Key clinical issues that will be covered

- prophylactic use of vaginal progesterone for women considered to be at risk of preterm labour and birth because they have any of the factors listed in 2.1.4.1
- prophylactic use of cervical cerclage for women considered to be at risk of preterm labour and birth because they have any of the factors listed in 2.1.4.1
- non-prophylactic ('rescue') cervical cerclage for women in suspected preterm labour

- diagnosis of preterm pre-labour rupture of membranes using biochemical tests
- diagnosis of preterm labour by clinical assessment, biochemical testing and cervical ultrasound (alone or in combination)
- routine surveillance (temperature monitoring and cardiotocography) of women with suspected or diagnosed preterm pre-labour rupture of membranes
- antenatal antibiotic prophylaxis for women diagnosed with preterm pre-labour rupture of membranes
- use of progesterone/progestogens for women with suspected or diagnosed preterm labour to improve the outcomes of preterm labour
- use of tocolytic agents (beta-sympathomimetics, oxytocin receptor antagonists, calcium channel blockers, cyclo-oxygenase enzyme inhibitors, non-steroidal anti-inflammatory drugs, nitroglycerin, magnesium sulfate) for women with suspected or diagnosed preterm labour to improve the outcomes of preterm labour
- pharmacological interventions to improve neonatal outcomes including:
    - maternal corticosteroids for fetal lung maturation
    - magnesium sulfate for preterm neonatal neuroprotection
- information giving and support for women at risk of preterm labour, or who are suspected or diagnosed to be in preterm labour, and women having a planned preterm birth
- fetal monitoring for women suspected to be in or diagnosed to be in preterm labour
- mode of birth for women suspected to be in or diagnosed to be in spontaneous preterm labour
- timing of cord clamping

For further details please refer to the scope in Appendix A and review questions in Appendix D.

2.1.5 What this guideline does not cover

2.1.5.1 Groups that will not be covered

- women in labour at term
- women with a multiple pregnancy

2.1.5.2 Clinical issues that will not be covered

- routine screening for preterm labour in all pregnant women, including fibronectin testing
- risk factors for preterm labour
- Laparoscopic cerclage.
- Indications for planned preterm birth.
- methods of induction of preterm labour
- mode of birth where this is planned antenatally (for women not in suspected or diagnosed preterm labour)
- use of intrapartum analgesia
- care of preterm neonates including resuscitation
- additional care that is specific to women with co-existing conditions such as hypertension, diabetes or blood-borne viruses

2.1.6 Relationships between the guideline and other NICE guidance

2.1.6.1 Related NICE guidance

Diabetes in pregnancy (2015) NICE clinical guideline NG3

Antibiotics for early-onset neonatal infection (2012) NICE clinical guideline 149

Drainage, irrigation and fibrinolytic therapy (DRIFT) for post-haemorrhagic hydrocephalus in preterm infants (2011) NICE interventional procedure guidance 412

Multiple pregnancy (2011) NICE clinical guideline 129

Quitting smoking in pregnancy and following childbirth (2010) NICE public health guidance 26

Pregnancy and complex social factors (2010) NICE clinical guideline 110

Hypertension in pregnancy (2010). NICE clinical guideline 107

Neonatal jaundice (2010) NICE clinical guideline 98

Induction of labour (2008) NICE clinical guideline 70

Antenatal care (2008) NICE clinical guideline 62

Antenatal and postnatal mental health (2007) NICE clinical guideline 45

Laparoscopic cerclage for prevention of recurrent pregnancy loss due to cervical incompetence (2007) NICE interventional procedure guidance 228

Postnatal care (2014) NICE clinical guideline 37

Endovascular closure of patent ductus arteriosus (2004) NICE interventional procedure guidance 97

Cerebral Palsy update (under development) NICE clinical guideline

Vision Amniotic Leak Detector to assess unexplained vaginal wetness in pregnancy (2013) NICE medical technology guidance MTG15

2.2 Methods

This chapter sets out in detail the methods used to review the evidence and to generate the recommendations that are presented in subsequent chapters. This guidance was developed in accordance with the methods outlined in the NICE guidelines manual 2012 for the stages up to and including guideline development and moved to the updated NICE guidelines manual 2014 since consultation stage.

2.2.1 Developing the review questions and outcomes

Review questions were framed according to the type of question:

- intervention - PICO (patient, intervention, comparison and outcome)
- dianostic test accuracy - population, index tests, reference standard and target condition for reviews of diagnostic test accuracy
- qualitative - population, area of interest, outcomes.

These frameworks guided the literature searching process, critical appraisal and synthesis of evidence and facilitated the development of recommendations by the Guideline Committee.

The review questions were drafted by the NCC-WCH technical team and were refined and validated by the Guideline Committee. The questions were based on the key clinical areas identified in the scope (Appendix D).

A total of 18 review questions were identified.

Full literature searches, critical appraisals and evidence reviews were completed for all the specified review questions.

Table 1: Review questions

Chapter	Type of review	Review questions	Outcomes
3	Qualitative	What additional information and support should be given to women (antenatally or during labour) and their families where the woman is at increased risk of preterm labour, or is suspected or diagnosed to be in preterm labour, or has a planned preterm birth?	**Maternal/family outcomes** • psychological outcomes • satisfaction/views of care • knowing choices • experience of childbirth • established breastfeeding/breastfeeding on discharge from hospital • breastfeeding in longer term (as defined by research authors) • "bonding" with the baby • plans to have/not to have any more children • impact on family/siblings – any reported **Neonatal outcomes** • any longer term outcomes
4.2	Interventional	What is the clinical effectiveness of prophylactic progesterone (vaginal or oral) in preventing preterm labour in pregnant women considered to be at risk of preterm labour and birth?	**Maternal outcomes** • maternal mortality • side effects/adverse effects • emotional/psychological impact/effect **Neonatal outcomes** • all mortality • number/proportion of babies born preterm • time from intervention to birth (delay to birth was selected as a surrogate) • bronchopulmonary dysplasia/chronic lung disease • neonatal sepsis • neurodevelopmental disability • congenital abnormalities
4.3	Interventional	What is the clinical effectiveness of prophylactic cervical cerclage in preventing preterm labour in women considered to be at risk of preterm labour and birth?	**Maternal outcomes** • maternal mortality • maternal adverse effects including infection requiring intervention, cervical trauma requiring repair • maternal emotional/psychological impact **Neonatal outcomes** • mortality up to 1 year

Chapter	Type of review	Review questions	Outcomes
			• interval between procedure and delivery • preterm birth • serious neonatal morbidity • sepsis • chronic lung disease/ bronchopulmonary dysplasia • long-term infant neurodevelopmental outcomes/neurodevelopmental disability
5	Diagnostic	What is the diagnostic accuracy of the following tests to identify preterm pre-labour rupture of membranes: • placental alpha-microglobulin-1 • nitrazine (pH) • insulin-like growth factor binding protein-1 • fetal fibronectin	• sensitivity / specificity • positive / negative likelihood ratio
6	Interventional	What is the clinical effectiveness of antenatal prophylactic antibiotics given to women with diagnosed preterm pre-labour rupture of membranes to improve outcomes of preterm labour?	**Maternal** • maternal mortality • any infection including chorioamnionitis • major adverse drug reaction **Neonatal** • infant/neonatal/perinatal mortality • number/proportion of babies born preterm • interval between intervention and delivery (delay to birth was selected as a surrogate) • brain injury • necrotising enterocolitis • any neonatal infection (including. neonatal sepsis) • cerebral palsy (CP) or other neurodevelopmental disability • any composite neurological outcomes
7	Diagnostic accuracy	What is the diagnostic value of temperature, pulse, white cell count, C-reactive protein and cardiotocography (CTG) to identify infection in women with preterm pre-labour rupture of membranes (P-PROM)?	• sensitivity / specificity • positive / negative likelihood ratio

Chapter	Type of review	Review questions	Outcomes
8	Interventional	What is the clinical effectiveness of non-prophylactic 'rescue' cervical cerclage in preventing preterm birth for women in suspected preterm labour?	**Maternal** • maternal mortality • maternal adverse effects • maternal emotional/psychological impact **Neonatal** • mortality up to 1 year • interval between procedure and delivery (delay to birth was selected as a surrogate) • preterm birth • serious neonatal morbidity • sepsis • chronic lung disease/ bronchopulmonary dysplasia • long-term infant neurodevelopmental outcomes
10	Diagnostic accuracy	What is the diagnostic accuracy of the following (alone or in combination) in women with intact membranes to identify preterm labour leading to preterm birth: • clinical assessment • biochemical testing for markers for preterm labour namely cervicovaginal fetal fibronectin and IGF-BP1 insulin-like growth factor binding protein 1 • cervical ultrasound features (such as cervical length and funnelling)?	• sensitivity / specificity • positive / negative likelihood ratio
10	Interventional	What is the clinical effectiveness of a single course of maternal corticosteroids for fetal lung maturation given at different gestations in improving preterm neonatal outcomes? What is the clinical effectiveness of repeat courses of maternal corticosteroids for fetal lung maturation in	Maternal outcomes • maternal mortality • adverse events– all Neonatal outcomes • all death up to 1 year • need for mechanical ventilation • bronchopulmonary dysplasia/chronic lung disease • intraventricular haemorrhage • neonatal sepsis • neurodevelopmental disability

Chapter	Type of review	Review questions	Outcomes
		improving preterm neonatal outcomes?	
11	Interventional	What is the clinical and cost effectiveness of magnesium sulfate given to women at high risk of giving birth preterm (defined as those suspected to be in preterm labour or diagnosed as being in preterm labour and those having planned preterm birth) for preventing cerebral palsy and other neurological disorders in babies born at different preterm gestations?	**Maternal outcomes** • mortality • side effects **Neonatal outcomes** • stillbirth • neonatal/perinatal mortality • apgar score < 7 at 5 minutes • need for transfer to NICU • need for mechanical ventilation • infant feeding at 1 and 6 weeks of age (breastfeeding or exclusive formula) • longer term outcomes (record any that are reported) • major neonatal morbidity (any reported)
13	Interventional	What is the clinical and cost effectiveness of tocolytics given to women with suspected or diagnosed preterm labour to improve outcomes: • progesterone/progestogens • beta-sympathomimetics • oxytocin receptor antagonists • calcium channel blockers • cyclo-oxygenase enzyme inhibitors • non-steroidal anti-inflammatory drugs • nitric oxide donors • magnesium sulfate	**Maternal outcomes** • maternal mortality • adverse events-discontinuation of treatment • maternal infection **Neonatal outcomes** • perinatal mortality • neonatal mortality • time from administration to birth; (delay of birth by 48h or more was selected as a surrogate) • mean gestational age at birth • respiratory distress syndrome • chronic lung disease/bronchopulmonary dysplasia • intraventricular haemorrhage • white matter injury/periventricular leucomalacia • neonatal infection/sepsis • neurodevelopmental disability • quality of life
13.2	Predictive accuracy	What are the criteria for best interpreting the preterm fetal heart rate trace at different gestational ages for unborn babies whose mothers are in suspected or diagnosed preterm labour?	**Maternal outcomes** • mortality • mode of birth (and indication if operative delivery, and type of CS incision) **Neonatal outcomes** • mortality (all death up to 1 year - includes stillbirth, perinatal mortality, neonatal mortality and infant death) • trauma/injury to infant (specify type) • intraventricular haemorrhage/periventricular leucomalacia (PVL)/white matter

Chapter	Type of review	Review questions	Outcomes
			injury (and sub-group analysis needed where reported separately) • neonatal sepsis • need for mechanical ventilation • length of stay in neonatal intensive care unit or neonatal unit • cord blood gas values at birth
13.3	Diagnostic accuracy	What is the clinical effectiveness of electronic fetal monitoring compared with intermittent auscultation at different gestational ages for unborn babies whose mothers are in suspected or diagnosed preterm labour?	**Maternal** • mortality • mode of birth (and indication if operative delivery, and type of CS incision) **Neonatal** • mortality (all death up to 1 year - includes stillbirth, perinatal mortality, neonatal mortality and infant death) • trauma/injury to infant (specify type) • intracranial or interventricular haemorrhage/periventricular leucomalacia (PVL)/white matter injury (and sub-group analysis needed where reported separately) • respiratory distress syndrome • neonatal sepsis • need for mechanical ventilation • length of stay in neonatal intensive care unit or neonatal unit • cord blood gas values at birth
13.3	Interventional	What is the utility of fetal blood sampling (FBS) as an adjunct to intrapartum fetal heart rate monitoring at different gestational ages	**Maternal** • mortality • mode of birth **Neonatal** • mortality • trauma/injury to infant • intraventricular haemorrhage • neonatal sepsis • need for mechanical ventilation • length of stay in neonatal intensive care unit or neonatal unit • cord blood gas values at birth
14	Interventional	For women who present in suspected or diagnosed preterm labour (who have not planned antenatally to give birth by caesarean section (CS) and for whom there are no other known indications for CS birth), what is the clinical effectiveness of deciding to carry out a	**Maternal outcomes** • maternal mortality • hysterectomy/postpartum haemorrhage • infection Sepsis **Neonatal outcomes** • mortality up to 1 year • chronic lung disease/ bronchopulmonary dysplasia • respiratory distress syndrome

Chapter	Type of review	Review questions	Outcomes
		CS compared with deciding to allow vaginal birth?	• intracranial haemorrhage • long-term infant neurodevelopmental outcomes
15	Interventional	In preterm birth, does later or delayed cord clamping in active management of third stage improve maternal and neonatal outcomes compared to earlier or immediate cord clamping?	**Maternal outcomes** • mortality • primary postpartum haemorrhage • retained placenta **Neonatal outcomes** • neonatal or infant mortality • anaemia requiring transfusion • respiratory distress • brain injury • treatment for hyperbilirubinaemia with blood exchange transfusion • blood counts at 6 and 12 hours • apgar score

2.2.2 Searching for evidence

2.2.2.1 Clinical literature search

During the scoping stage, a search was conducted for guidelines and reports on websites of organisations relevant to the topic. Searching for grey literature or unpublished literature was not undertaken. Searches for electronic, ahead of print or "online early" publications were not routinely undertaken.

Systematic literature searches were undertaken to identify all published clinical evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE guidelines manual 2012.

Databases were searched using relevant medical subject headings, free-text terms and study type filters where appropriate. Studies published in languages other than English were not reviewed. Where possible, searches were restricted to retrieve articles published in English. All searches were conducted in Medline, Embase, and The Cochrane Library. All searches were updated on March 2015 with the exception of the search for the review question that included the Network Meta-Analysis (NMA) which was last updated on January 2015. Any studies added to the databases after this date (even those published prior to this date) were not included unless specifically stated in the text.

Search strategies were quality assured by cross checking reference lists of highly relevant papers, analysing search strategies in systematic reviews (SR) and asking the Guideline Committee members to highlight any additional studies. The questions, the study types applied, the databases searched and the years covered can be found in Appendix E.

The titles and abstracts of records retrieved by the searches were sifted for relevance, with potentially significant publications obtained in full text. These were assessed against the inclusion criteria specified in the protocols (Appendix D).

2.2.2.2 Health economic literature search

Systematic literature searches were also undertaken to identify health economic evidence within published literature relevant to the review questions. The evidence was identified by conducting a broad search relating to pneumonia in the NHS Economic Evaluation Database

(NHS EED), the Health Economic Evaluations Database (HEED) and Health Technology Assessment (HTA) databases with no date restrictions. Additionally, the search was run on Medline and Embase using a specific economic filter, from 2011 to ensure recent publications that had not yet been indexed by the economic databases were identified. This was supplemented by additional searches that looked for economic papers specifically relating to gas exchange management as this was an area identified for original economic modelling. Studies published in languages other than English were not reviewed. Where possible, searches were restricted to articles published in English.

The search strategies for the health economics literature search are included in Appendix E. All searches were updated on March 2015. No papers published after this date were considered.

2.2.3 Evidence of effectiveness

The evidence was reviewed following the steps shown:

- Potentially relevant studies were identified for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Full papers were reviewed against pre-specified inclusion and exclusion criteria to identify studies that addressed the review question in the appropriate population and reported on outcomes of interest (review protocols are included in Appendix D).
- Relevant studies were critically appraised using the appropriate checklist as specified in the NICE guidelines manual 2012. For diagnostic questions the QUADAS-2 checklist was followed.
- Key information was extracted on the study's methods, PICO factors and results. These were presented in summary tables in each chapter and evidence tables (in Appendix H).
- Summaries of evidence were generated by outcome and were presented in Guideline Committee meetings:
    - randomised studies - data were meta-analysed where appropriate and reported in GRADE profiles (for interventional reviews)
    - diagnostic / predictive accuracy studies were presented as measures of diagnostic / predictive test accuracy (sensitivity, specificity, positive and negative predictive value). A meta-analysis was only conducted when the included studies were not heterogeneous.
    - qualitative studies - the themes of the studies were organised in a modified version of a GRADE profile, where possible, along with quality assessment otherwise presented in a narrative form
- 50% of all data extracted was quality assured by a second reviewer. 50% of the GRADE quality assessment was quality assured by a second reviewer to minimise any potential risk of reviewer bias or error.

2.2.3.1 Inclusion and exclusion criteria

The inclusion and exclusion of studies was based on the review protocols, which can be found in Appendix D:. Excluded studies by review question (with the reasons for their exclusion) are listed in Appendix G:. The Guideline Committee was consulted about any uncertainty regarding inclusion or exclusion.

2.2.4 Methods of combining clinical studies

2.2.4.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted to combine the results of studies for each review question using Cochrane Review Manager (RevMan5) software or STATA. Fixed-

effects (Mantel-Haenszel) techniques were used to calculate risk ratios (relative risk) for the binary outcomes.

For the continuous outcomes, measures of central tendency (mean) and variation (standard deviation) were required for meta-analysis. A generic inverse variance option in RevMan5 was used if any studies reported solely the summary statistics and 95% confidence interval (95% CI) or standard error; this included any hazard ratios reported. However, in cases where standard deviations were not reported per intervention group, the standard error (SE) for the mean difference was calculated from other reported statistics (p values or 95% CIs) if available; meta-analysis was then undertaken for the mean difference and SE using the generic inverse variance method in RevMan5. When the only evidence was based on studies that summarised results by presenting medians (and interquartile ranges), or only p values were given, this information was assessed in terms of the study's sample size and was included in the GRADE tables without calculating the relative or absolute effects or as a narrative summary. Consequently, aspects of quality assessment such as imprecision of effect could not be assessed for this evidence and this has been recorded in the footnotes of the GRADE tables. When more than 2 studies reported a continuous outcome, the presentation of mean (SD) per comparison group was taken by averaging the means of included studies.

In instances where multiple scales were reported for a single outcome, mean differences were standardised (divided by their SD) before pooling, giving meta-analysed results that were reported as standardised mean differences (SMD), with a standard deviation of 1.

Where reported, time-to-event data were presented as a hazard ratio or results from a Cox hazard proportion model were given as a result from a multivariate analysis.

Statistical heterogeneity was assessed by visually examining the forest plots, and by considering the chi-squared test for significance at $p < 0.1$ or an I-squared inconsistency statistic (with an I-squared value of 50-74.99% indicating serious inconsistency and I-squared value of over 75% indicating very serious inconsistency). If the heterogeneity still remained, a random-effects (DerSimonian and Laird) model was employed to provide a more conservative estimate of the effect. Where considerable heterogeneity was present, we set out to perform predefined subgroup analyses based on the following factors:

- different gestational age of fetus
- inclusion of studies with mixed populations of women with single and multiple pregnancies
- different groups of women at high risk of preterm labour.

2.2.4.1.1 Data synthesis for diagnostic test accuracy review

For diagnostic test accuracy studies, the following outcomes were reported: sensitivity, specificity, positive likelihood ratio, negative likelihood ratio.

The assessment of usefulness of the diagnostic or predictive accuracy of tests followed the terms and thresholds below:

Sensitivity and specificity:

- high – 90% and above
- moderate – 75% to 89.9%
- low – 74.9% or below

Positive likelihood ratio:

- very useful – more than 10
- moderately useful – 5 to 10
- not useful – less than 5

Negative likelihood ratio:

- very useful – 0 to 0.1
- moderately useful – more than 0.1 to 0.5
- not useful – more than 0.5

2.2.4.1.2 Data synthesis for qualitative review

For the qualitative review in the guideline, results were presented in 2 ways:

- NICE checklists on assessing qualitative studies were used to assess the quality assessment of individual studies.
- results were reported narratively by individual study when appropriate

2.2.4.1.3 Data synthesis using a Network Meta-Analysis (NMA)

A NMA was formulated to synthesise direct and indirect evidence of treatments' efficacy to determine which treatments are most effective at delaying preterm birth to improve the outcomes for the baby with least harm to/adverse effects for the woman whilst preserving randomisation within primary studies for the outcomes of:

- neonatal mortality
- perinatal mortality
- respiratory distress syndrome (RDS)
- intraventricular haemorrhage (IVH)
- adverse events requiring discontinuation of treatment
- delay of birth by at least 48 hours
- neonatal sepsis
- gestational age at birth.

Hierarchical Bayesian network meta-analyses (NMAs) was performed using the software WinBUGS version 1.4. These models was based on original work from the University of Bristol (https://www.bris.ac.uk/cobm/research/mpes/mtc.html).

A class effect model was adopted for the new NMA because it was hypothesised that treatments within class would borrow similar clinical characteristics and mechanisms of effect. In other words, results for one member of the class in relation to efficacy and side effects were considered to be generalisable to other members of that same class. Since there was no evidence of within-class variability for any of the outcomes considered, all the results presented assume that all treatments in a class have the same relative effect (see Appendix K).

Trials with non UK licensed interventions were included in the NMA to allow the maximum use of available evidence and borrow strength of loops in the network only if there was at least one trial that included licensed (for preterm labour or for other conditions) interventions for the same class. Some other considerations in the design of the NMA:

- The Guideline Committee discussed that although dosage, mode of administration and timing of treatment may influence the effectiveness of different tocolytics interventions, it was considered unlikely for this factor to change the direction of relative effect for the different interventions tested in the analysis. Therefore, the Committee decided not to consider any confounding effect of these factors in the NMA.
- Some of the included studies examined drugs that are not licensed as tocolytics for use in pregnancy (including nylidrin and barusiban). These drugs were included in the NMA to increase the size of the network, and because it is not uncommon for drugs that are not licenced for pregnancy indications to be prescribed for use in this context.
- The Guideline Committee chose to have separate classes for alcohol/ethanol and combination treatments (classed as 'other') in the new NMA.

Standard deviations (SD) were imputed where they were not reported for 5 studies assessing estimated gestational age. Imputed values were based on the median SD for each of these treatments from other included studies. A sensitivity analysis using the upper quartile of the reported SD was carried out. Apart from increased uncertainty in estimates the main results were not affected.

2.2.4.1.4 Assessment of consistency

Consistency was assessed by checking the agreement of direct and indirect evidence using a node-split model (Dias 2010). Bayesian p-values for agreement between direct and indirect evidence were calculated. When these were lower than 0.05, included trials were inspected to help determine reasons for the potential inconsistency, bearing in mind that multiple probabilities of disagreement are being calculated and there is the potential to find spurious results.

Consistency was considered as part of the quality appraisal of the evidence for the NMA (see below).

2.2.4.1.5 Model evaluation

For all the networks set up in the NMA, models for fixed and random effects were developed and then these were compared based on residual deviance and deviance information criteria (DIC). The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset which has the same structure as that currently observed. A small difference in DIC between the fixed and random effects models (3-5 points) implies that the better fit obtained by adding random effects does not justify the additional complexity. However, if the difference in DIC between a fixed and random effect model was less than 5 points, and the models make very similar inferences, then we would report the results from a fixed effects model results as it does not make as many assumptions as the random effect model, contains fewer parameters and it is easier for clinical interpretation than the random effects model.

2.2.5 Type of studies

Systematic reviews (SR) with or without meta-analyses were considered the highest quality evidence to be selected for inclusion. Individual patient data (IPD) meta-analyses are considered the gold standard type of meta-analysis and was prioritised for inclusion in the evidence base of this guideline when appropriate.

Randomised trials and observational studies (including diagnostic or prognostic studies) were included in the evidence reviews as appropriate..

Literature reviews, posters, letters, editorials, comment articles, conference abstracts, unpublished studies and studies not in English were excluded.

For intervention reviews in this guideline, randomised controlled trials (RCTs) were included because they are considered the most robust study design for unbiased estimation of intervention effects.

If the Guideline Committee believed RCT data were not appropriate or there was limited evidence from RCTs, well-conducted non-randomised comparative studies were included. For diagnostic reviews, cross-sectional, retrospective studies and case series were included. Please refer to Appendix D for full details on the study design of studies selected for each review question.

The Guideline Committee defined women's and babies' mortality, birth within 48 hours and 7 days as primary outcomes and long-term infant neurodevelopmental outcomes, birth events (mode of birth, complications of birth, perineal trauma), newborn events (condition at birth, birth injuries, admission to neonatal units), women's assessment of birth experience, as

secondary outcomes. The Guideline Committee considered other outcomes when they were relevant to specific questions.

2.2.6 Appraising the quality of evidence by outcomes

The evidence for outcomes from the included RCTs and, where appropriate, observational studies was evaluated and presented using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group (http://www.gradeworkinggroup.org/). The software developed by the GRADE working group (GRADEpro) was used to assess the quality of each outcome, taking into account individual study quality factors and the meta-analysis results. The 'Clinical evidence profile' table includes details of the quality assessment and pooled outcome data, where appropriate, an absolute measure of intervention effect and the summary of quality of evidence for that outcome. In this table, the columns for intervention and control indicate summary measures and measures of dispersion (such as mean and standard deviation or median and range) for continuous outcomes and frequency of events (n/N: the sum across studies of the number of patients with events divided by sum of the number of completers) for binary outcomes..

The evidence for each outcome was examined separately for the quality elements listed and defined in Table 1. Each element was graded using the quality levels listed in Table 2.

The main criteria considered in the rating of these elements are discussed below. Footnotes were used to describe reasons for grading a quality element as having serious or very serious limitations. The ratings for each component were summed to obtain an overall assessment for each outcome (Table 1)

The GRADE toolbox is currently designed only for randomised trials and observational studies but we adapted the quality assessment elements and outcome presentation for diagnostic accuracy and prognostic studies subject to data availability.

Table 1: Description of quality elements in GRADE for intervention studies

Quality element	Description of quality elements in GRADE for intervention studies
Risk of bias (study limitations)	Limitations in the study design and implementation may bias the estimates of the treatment effect. High risk of bias for the majority of the evidence decreases confidence in the estimate of the effect.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of results.
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question, or recommendation made, such that the effect estimate is changed.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect. Imprecision results if the confidence interval includes the clinically important threshold.
Publication bias	Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies.

Table 2: Levels of quality elements in GRADE Level

Levels of quality elements in GRADE Level	Description
None	There are no serious issues with the evidence.
Serious	The issues are serious enough to downgrade the outcome evidence by 1 level.

Levels of quality elements in GRADE Level	Description
Very serious	The issues are serious enough to downgrade the outcome evidence by 2 levels.

Table 3: Overall quality of outcome evidence in GRADE Level

Overall quality of outcome evidence in GRADE Level	Description
High	Further research is very unlikely to change our confidence in the estimate of effect.
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low	Any estimate of effect is very uncertain.

2.2.6.1 Grading the quality of clinical evidence

After results were pooled, the overall quality of evidence for each outcome was considered. The following procedure was adopted when using GRADE:

- A quality rating was assigned, based on the study design. RCTs start as high, observational studies as moderate, and uncontrolled case series as low or very low.
- The rating was then downgraded for the specified criteria: risk of bias (study limitations), inconsistency, indirectness, imprecision and publication bias. These criteria are detailed below. Evidence from observational studies (which had not previously been downgraded) was upgraded if there was: a large magnitude of effect, a dose-response gradient, and if all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results showed no effect. Each quality element considered to have 'serious' or 'very serious' risk of bias was rated down by 1 or 2 points respectively.
- The downgraded/upgraded ratings were then summed and the overall quality rating was revised. For example, all RCTs started as High and the overall quality became Moderate, Low or Very low if 1, 2 or 3 points were deducted respectively.
- The reasons or criteria used for downgrading were specified in the footnotes.

The details of the criteria used for each of the main quality elements are discussed further in the following sections 2.2.6.2 to 2.2.6.6.

2.2.6.2 Risk of bias

Bias can be defined as anything that causes a consistent deviation from the truth. Bias can be perceived as a systematic error, for example, if a study was carried out several times and there was a consistently wrong answer, the results would be inaccurate.

The risk of bias for a given study and outcome is associated with the risk of over- or underestimation of the true effect.

The risks of bias are listed in Table 4.

A study with a poor methodological design does not automatically imply high risk of bias; the bias is considered individually for each outcome and it is assessed whether this poor design will impact on the estimation of the intervention effect.

Table 4: Risk of bias in randomised controlled trials

Risk of bias	Explanation
Allocation concealment	Those enrolling patients are aware of the group to which the next enrolled patient will be allocated (this is a major problem in "pseudo" or "quasi" randomised trials with allocation by for example, day of week, birth date, chart number).
Lack of blinding	Patient, caregivers, those recording outcomes, those adjudicating outcomes, or data analysts are aware of the arm to which patients are allocated. As mortality is the most critical outcome for this guideline and its effect is not biased by lack of blinding, unblinded studies were not automatically downgraded for this outcome.
Incomplete accounting of patients and outcome events	Missing data not accounted for and failure of the trialists to adhere to the intention to treat principle when indicated.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results.
Other risks of bias	For example: • stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules • use of unvalidated patient-reported outcomes • recruitment bias in cluster randomised trials

2.2.6.3 Diagnostic studies

For diagnostic accuracy studies, the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklist was used (see Appendix F in The guidelines manual 2012). Risk of bias and applicability in primary diagnostic accuracy studies in QUADAS-2 consists of 4 domains (see Figure 1)

- patient selection
- index test
- reference standard
- flow and timing

Figure 1: Description of QUADAS-2 domains

DOMAIN	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING
Description	Describe methods of patient selection: Describe included patients (prior testing, presentation, intended use of index test and setting):	Describe the index test and how it was conducted and interpreted:	Describe the reference standard and how it was conducted and interpreted:	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram): Describe the time interval and any interventions between index test(s) and reference standard:
Signalling questions (yes/no/unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias: High/low/unclear	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: High/low/unclear	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

2.2.6.4 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results. When estimates of the treatment effect across studies differ widely (that is when there is heterogeneity or variability in results), this suggests true differences in underlying treatment effect.

Heterogeneity in meta-analyses was examined and sensitivity and subgroup analyses performed as pre-specified in the protocols (Appendix D).

When heterogeneity existed (chi-squared $p < 0.1$, I-squared inconsistency statistic of between 50-74.99% or I-squared > 50% or evidence from examining forest plots), but no plausible explanation was found (for example duration of intervention or different follow-up periods) the quality of evidence was downgraded by 1 or 2 levels, depending on the extent of uncertainty to the results contributed by the inconsistency in the results. In addition to the I-squared and chi-squared values, the decision for downgrading was also dependent on factors such as whether the intervention is associated with benefit in all other outcomes or whether the uncertainty about the magnitude of benefit (or harm) of the outcome showing heterogeneity would influence the overall judgment about net benefit or harm (across all outcomes).

When outcomes are derived from a single trial, inconsistency is not an issue for downgrading the quality of evidence. However, “no inconsistency” is nevertheless used to reflect the decision not to downgrade the evidence for this quality assessment domain.

2.2.6.5 Indirectness

Directness refers to the extent to which the populations, intervention, comparisons and outcome measures are similar to those defined in the inclusion criteria for the reviews. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention.

2.2.6.6 Imprecision

Imprecision in guidelines concerns whether the uncertainty (confidence interval) around the effect estimate means that it is not clear whether there is a clinically important difference between interventions or not. Therefore, imprecision differs from the other aspects of evidence quality in that it is not really concerned with whether the point estimate is accurate or correct (has internal or external validity) instead it is concerned with the uncertainty about what the point estimate is. This uncertainty is reflected in the width of the confidence interval.

The 95% confidence interval (95% CI) is defined as the range of values that contain the population value with 95% probability. The larger the trial, the smaller the 95% CI and the more certain the effect estimate.

Imprecision in the evidence reviews was assessed by considering whether the width of the 95% CI of the effect estimate was relevant to decision-making, considering each outcome in isolation.

When the confidence interval of the effect estimate is wholly contained in 1 of the 3 zones (clinically important benefit, clinically important harm, no clinically important benefit or harm) we are not uncertain about the size and direction of effect (whether there is a clinically important benefit, or the effect is not clinically important, or there is a clinically important harm), so there is no imprecision (Figure 2)).

When a wide confidence interval lies partly in each of 2 zones, it is uncertain in which zone the true value of effect estimate lies, and therefore there is uncertainty over which decision to make (based on this outcome alone). The confidence interval is consistent with 2 decisions and so this is considered to be imprecise in the GRADE analysis and the evidence is downgraded by 1 level (‘serious imprecision’).

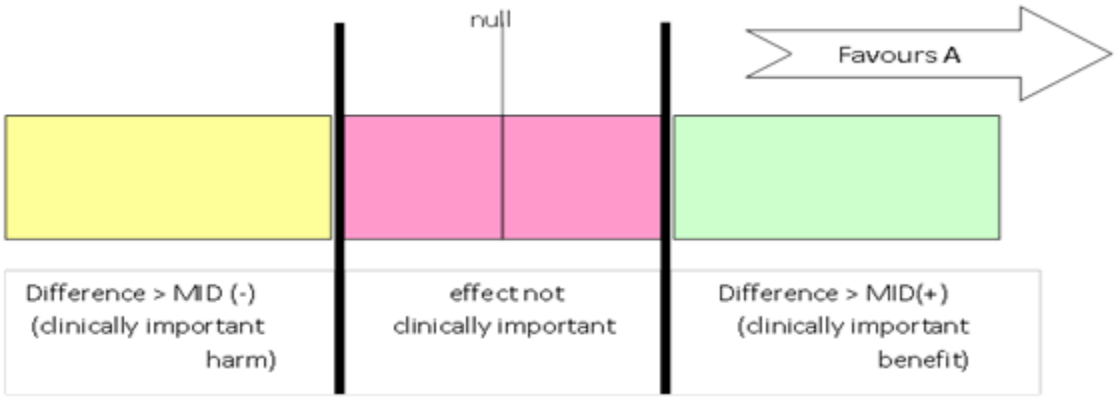
If the confidence interval of the effect estimate crosses into 3 zones, this is considered to be very imprecise evidence because the confidence interval is consistent with 3 clinical decisions and there is a considerable lack of confidence in the results. The evidence is therefore downgraded by 2 levels in the GRADE analysis (‘very serious imprecision’).

Implicitly, assessing whether the confidence interval is in, or partially in, a clinically important zone, requires the Committee to estimate an MID or to say whether they would make different decisions for the 2 confidence limits.

The Committee considered it clinically acceptable to use the GRADE default MID to assess imprecision: a 25% relative risk reduction or relative risk increase was used, which corresponds to clinically important thresholds for a risk ratio of 0.75 and 1.25 respectively. This default MID was used for all the dichotomous outcomes in the interventions evidence reviews. For continuous outcomes, a MID was calculated by adding or subtracting 0.5 times standard deviations. For outcomes that were meta-analysed using the standardised mean difference approach (SMD), the MID was calculated by adding or subtracting 0.5 (given SD equals 1).

For the diagnostic questions, we assessed imprecision on the outcome of positive or negative likelihood ratio because these were prioritised by the Committee as the most important diagnostic outcomes for their decision making. The assessment of imprecision for the results on positive or negative likelihood ratio followed the same concept as used in interventional reviews. For example, if the 95% confidence interval of the positive likelihood ratio crossed 2 zones (from moderately useful [5 to 10] to very useful [>10]) then imprecision was downgraded by 1, or if crossed 3 zones (not useful (<5), moderately useful (5 to 10) and very useful (>10) then imprecision was downgraded by 2.

Figure 2: Illustration of imprecision assessment



2.2.7 Quality assessment of NMA

The quality of evidence from NMA was assessed using a modified GRADE appraisal process.

Risk of bias was assessed using the quality assessment undertaken by Haas 2012 and for all additional studies using the check list developed by the Technical Support Unit (TSU), commissioned by NICE.

For more information about the check list developed by the TSU see here: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0092508

Indirectness was assessed using information about the study population and imprecision based on credible intervals in line with standard GRADE methodology. Inconsistency was assessed by comparing estimates based on direct and indirect data included in the network. Where there was evidence of inconsistency (see Appendix J) then quality was downgraded.

Imprecision was assessed based on the credible interval within each comparison. Data were downgraded if credible interval crossed the two default MIDs or majority of the comparisons.

2.2.8 Assessing clinical importance

The Guideline Committee assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. To facilitate this, binary outcomes were converted into absolute risk differences (ARDs) using GRADEpro software: the median control group risk across studies was used to calculate the ARD and its 95% CI from the pooled risk ratio.

2.2.9 Evidence statements

Evidence statements are summary statements that are presented after the GRADE profiles, summarising the key features of the clinical evidence presented. The wording of the evidence statements reflects the certainty or uncertainty in the estimate of effect. The evidence statements are presented by comparison (for intervention reviews) or by outcome and encompass the following key features of the evidence:

- the number of studies and the number of participants for a particular outcome
- a brief description of the participants
- an indication of the direction of effect (if 1 treatment is beneficial or harmful compared with the other, or whether there is no difference between the 2 tested treatments)

- a description of the overall quality of evidence (GRADE overall quality)

2.3 Evidence of cost effectiveness

The Guideline Committee is required to make decisions based on the best available evidence of both clinical and cost effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost effectiveness') rather than the total implementation cost. Thus, if the evidence suggests that a strategy provides significant health benefits at an acceptable cost per patient treated, it should be recommended even if it would be expensive to implement across the whole population.

Evidence on cost effectiveness related to the key clinical issues being addressed in the guideline was sought.

- a SR of the published economic literature was undertaken
- new cost-effectiveness analysis was conducted in priority areas

2.3.1 Literature review

The health economist:

- identified potentially relevant studies for each review question from the economic search results by reviewing titles and abstracts. Full papers were then obtained
- reviewed full papers against pre-specified inclusion/exclusion criteria to identify relevant studies (see 2.3.1.1 for details)
- critically appraised relevant studies using the economic evaluations checklist as specified in the guidelines manual
- extracted key information about study methods and results into evidence tables (included in Appendix H)
- generated summaries of the evidence in NICE economic evidence profiles (included in the relevant chapter for each review question) – see 2.3.1.2 for details

2.3.1.1 Inclusion and exclusion criteria

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost–utility, cost-effectiveness, cost-benefit and cost-consequence analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially includable as economic evidence.

2.3.1.2 NICE economic evidence profiles

The NICE economic evidence profile has been used to summarise cost and cost-effectiveness estimates. The economic evidence profile shows, for each economic study, an assessment of applicability and methodological quality for each economic evaluation. These assessments were made by the health economist using the economic evaluation checklist from The Guidelines Manual. It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and the incremental cost-effectiveness ratio for the base-case analysis in the evaluation, as well as information about the assessment of uncertainty in the analysis. See Table 5 for more details.

Table 5: Content of NICE economic evidence profile

Content of NICE economic evidence profile item	Description
Study	First author name, reference, date of study publication and country perspective.
Applicability	An assessment of applicability of the study to the clinical guideline, the current NHS situation and NICE decision-making(a): Directly applicable – the study meets all applicability criteria, or fails to meet 1 or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness. Partially applicable – the study fails to meet 1 or more applicability criteria and this could change the conclusions about cost effectiveness. Not applicable – the study fails to meet 1 or more applicability criteria and this is likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Limitations	An assessment of methodological quality of the study(a): • Minor limitations – the study meets all quality criteria, or fails to meet 1 or more quality criteria, but this is unlikely to change the conclusions about cost effectiveness. • Potentially serious limitations – the study fails to meet 1 or more quality criteria, and this could change the conclusion about cost effectiveness. • Very serious limitations – the study fails to meet 1 or more quality criteria and this is highly likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.
Other comments	Particular issues that should be considered when interpreting the study.
Incremental cost	The mean cost associated with one strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with one strategy minus the mean QALYs of a comparator strategy.
Cost effectiveness	Incremental cost-effectiveness ratio (ICER): the incremental cost divided by the incremental effects.
Uncertainty	A summary of the extent of uncertainty about the ICER reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

a. *Applicability and limitations were assessed using the economic evaluation checklist from the guidelines manual.*

2.3.2 Undertaking new health-economic analysis

As well as reviewing the published economic literature for each most review questions, as described above, new economic analysis was undertaken by the health economist in selected areas. Priority areas for new health economic analysis were agreed by the Guideline Committee after formation of the review questions and consideration of the available health economic evidence.

2.3.2.1 Cost-effectiveness criteria

NICE's report 'Social value judgements: principles for the development of NICE guidance' sets out the principles that Guideline Committees should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be

cost effective if either of the following criteria applied (given that the estimate was considered plausible).

- The intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- The intervention cost less than £20,000 per QALY gained compared with the next best strategy.

If the Guideline Committee recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in the 'Recommendations and link to evidence' section of the relevant chapter with reference to issues regarding the plausibility of the estimate or to the factors set out in 'Social value judgements: principles for the development of NICE guidance' guidance'.

If a study reported the cost per life year gained but not QALYs, the cost per QALY gained was estimated by multiplying by an appropriate utility estimate to aid interpretation. The estimated cost per QALY gained is reported in the economic evidence profile with a footnote detailing the life-years gained and the utility value used. When QALYs or life years gained are not used in the analysis, results are difficult to interpret unless one strategy dominates the others with respect to every relevant health outcome and cost.

2.3.3 In the absence of economic evidence

When no relevant published studies were found, and a new analysis was not prioritised, the Guideline Committee made a qualitative judgement about cost effectiveness by considering expected differences in resource use between options and relevant UK NHS unit costs alongside the results of the clinical review of effectiveness evidence.

2.4 Developing recommendations

Over the course of the guideline development process, the Guideline Committee was presented with:

- evidence tables of the clinical and economic evidence reviewed from the literature. All evidence tables are in Appendix H
- summary of clinical and economic evidence and quality assessment (as presented in Chapters 3 to 17)
- forest plots (Appendix I); and
- a description of the methods and results of the cost-effectiveness analysis undertaken for the guideline (Chapter 16)

Recommendations were drafted on the basis of the Guideline Committee's interpretation of the available evidence, taking into account the balance of benefits, harms and costs between different courses of action. Firstly, the net benefit over harm (clinical effectiveness) was considered, focusing on the prioritised outcomes and taking into account the clinical benefits and harms when 1 intervention was compared with another. The assessment of net benefit was moderated by the importance placed on the outcomes (the Guideline Committee's values and preferences), and the confidence the Guideline Committee had in the evidence (evidence quality). Secondly, it was assessed whether the net benefit justified any differences in costs.

In areas where no substantial clinical research evidence was identified, the Guideline Committee considered other NICE relevant guidelines and consensus statements or used their collective experience to identify good practice. The health economics justification in areas of the guideline where the use of NHS resources (interventions) was considered was

based on Guideline Committee consensus in relation to the likely cost effectiveness implications of the recommendations. The Guideline Committee also identified areas where evidence to answer their review questions was lacking and used this information to formulate recommendations for future research. When clinical and economic evidence was of poor quality, conflicting or absent, the Guideline Committee drafted recommendations based on their expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs or implications compared with the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues.

The wording of recommendations was agreed by the Guideline Committee and focused on the following factors:

- the actions health professionals need to take
- the information readers need to know
- the strength of the recommendation (for example the word 'offer' was used for strong recommendations and 'consider' for weak recommendations)
- the involvement of patients (and their carers if needed) in decisions on treatment and care
- consistency with NICE's standard advice on recommendations about drugs, waiting times and ineffective interventions

The main considerations specific to each recommendation are outlined in the 'Evidence to Recommendations' sections within each chapter.

2.4.1 Research recommendations

When areas were identified for which good evidence was lacking, the Guideline Committee considered making recommendations for future research. Decisions about inclusion were based on factors such as:

- the importance to patients or the population
- national priorities
- potential impact on the NHS and future NICE guidance
- ethical and technical feasibility

2.4.2 Validation process

This guidance is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the document. All comments received from registered stakeholders are responded to in turn and posted on the NICE website when the pre-publication check of the full guideline occurs.

2.4.3 Updating the guideline

Following publication, and in accordance with the NICE guidelines manual, NICE will undertake a review of whether the evidence base has progressed significantly to alter the guideline recommendations and warrant an update.

2.4.4 Disclaimer

Health care providers need to use clinical judgement, knowledge and expertise when deciding whether it is appropriate to apply guidelines. The recommendations cited here are a guide and may not be appropriate for use in all situations. The decision to adopt any of the recommendations cited here must be made by practitioners in light of individual patient circumstances, the wishes of the patient, clinical expertise and resources.

The National Collaborating Centre for Women and Children's Health disclaims any responsibility for damages arising out of the use or non-use of these guidelines and the literature used in support of these guidelines.

2.4.5 Funding

The National Collaborating Centre for Women and Children's Health was commissioned by NICE to undertake the work on this guideline.

3 Information and support

3.1 Introduction

Although many preterm babies arrive early without warning, some pregnancies are known to be at high risk of ending in preterm birth – when this is the case, there may be opportunities for steps to be taken to try to reduce that risk. The earlier a baby is born, the more severe his or her health problems are likely to be. Some babies born preterm require special care and may spend weeks or months hospitalised in neonatal intensive care unit.

At such a vulnerable time in their lives, families need to be given information and support to meet their needs, without causing unnecessary anxiety. This section considers the evidence for providing information and support to pregnant women and their families whose babies are at high risk of arriving early.

3.1.1 Review question

What additional information and support should be given to women (antenatally or during labour) and their families where the woman is at increased risk of preterm labour, or is suspected or diagnosed to be in preterm labour, or has a planned preterm birth?

This review question has two sections; the first section aims to identify themes of additional information given prior to birth that would be considered as important for women at increased risk or is suspected or diagnosed in preterm labour or has planned preterm labour. The second section examines whether interventions or packages of care designed to provide additional information prior to or during birth compared to usual care could result in better maternal, family and/or neonatal outcomes.Description of included studies

Seven studies were included in this review, 5 qualitative studies (Gauche & Payot 2011, Gupton & Heaman 1994, Sawyer 2013, Young 2012, Griffin 1997) were included in the first section of the review, which aims to identify the key areas of antenatal and intrapartum information and support needs of women and their families at increased risk for preterm labour. 2 RCTs (Oakley 1990, Villar 1992) were included in the second section of the review, which aimed to assess the effectiveness of interventions or packages of care with regard to maternal, family, and/or neonatal outcomes. The gestational age of women in included studies ranged from 15 to 36 weeks.

More information on the study characteristics and population characteristics are given in Table 6.

Table 6: Study and population characteristics

Included Studies	Type of study/methods/comparison groups (if applicable)	Population characteristics	Outcomes
Section 1 of review; key areas of information for families at risk for preterm birth			
Gaucher & Payot 2011	Qualitative study/ face to face interviews using grounded theory	5 mothers Range of gestational age: 26 to 30 2/7 weeks 3/5 had full term pregnancies after hospital discharge	Explored the women's concerns regarding possible preterm labour and their expectations of the prenatal consultation and of the neonatologist
Gupton & Heaman 1994	Qualitative study using Preterm Birth Learning	Convenience sample of 34 women	Ranking ordering of priorities for learning needs of hospitalised

Included Studies	Type of study/methods/comparison groups (if applicable)	Population characteristics	Outcomes
	Needs Questionnaire (PBLNQ)	Range of gestational age: 26-36 weeks All high risk pregnancies	women at risk of preterm birth
Sawyer 2013	Qualitative study/ interview with open ended questions	25 mothers and 7 couples Range of gestational age: 24-32	Experiences and satisfaction with care during preterm birth
Young 2012	Qualitative study/ face to face semi-structured interviews using ethnography methods	Preterm labour between 23 to 26 weeks 10 families 80% high risk pregnancies	Exploring the areas of importance for counselling for extreme prematurity
Griffin 1997	Qualitative study/ face to face interviews with open ended questions	Convenience sample of 13 parents (10 mothers) All high risk pregnancies	Evaluating the experience of a prenatal tour of the neonatal intensive care unit during high risk pregnancy
Section 2 of review: effectiveness of interventions or packages of antenatal care for pregnancies at risk of preterm birth			
Oakley 1990	RCT/ intervention: a minimum, 3 home visits from a midwife - at 14, 20, and 28 weeks' gestation - plus 2 telephone contacts	Intervention: 255/ control: 254 At risk pregnancies	Postnatal depression
Villar 1992	RCT/ intervention: a minimum of 4 home visits from specially trained female social workers or obstetrical nurses and had access to a special 'drop in' support office at each study hospital	Intervention: 1115/ control: 1120 At risk pregnancies	Satisfaction with antenatal care

Section 1; themes of additional information given prior to birth

All qualitative studies employed appropriate study methods:

Preterm Birth Learning Needs Questionnaire (PBLNQ) was included in 1 study to elicit both qualitative and quantitative data (Gupton & Heaman 1994), grounded theory (Gaucher & Payot 2011) and qualitative ethnography using semi-structured interviews (Young 2012) was employed in another study and the last one used semi-structured interviews (Sawyer 2013).

Two qualitative studies examined cases of extreme prematurity (Sawyer 2013, Young 2012). The first study assessed parents' positive and negative experiences and satisfaction with care during very preterm births (<32 gestational weeks), identifying determinants of their experience of care (Sawyer 2013). The second study interviewed parents with babies born at 23-26 gestational weeks to ascertain retrospectively how pre-delivery counselling could be improved (Young 2012).

One study (Griffin 1990) qualitatively examined a specific intervention (prenatal tour of the neonatal intensive care unit during high-risk pregnancy) to elicit parents' experiences regarding this.

Section 2; interventions or packages of care designed to provide additional information prior to or during birth compared to usual care

Two RCTs (Oakley 1990, Villar 1992) included in a Cochrane review (Hodnett 2010) that compared routine care provision with provision of additional support to reduce the likelihood of preterm birth or low birth weight in pregnant women matched the second part of this review protocol. One RCT (Oakley 1990) was conducted in the UK and usual antenatal care was compared with the addition of a social support intervention consisting of, at a minimum, 3 home visits from a midwife - at 14, 20, and 28 weeks' gestation - plus 2 telephone contacts or brief home visits between these times. The midwife was also on-call to provide support to mothers if necessary for 24 hours/day.

The second RCT (Villar 1992) was an international multicentre study that aimed at increasing social support and reducing stress and anxiety in pregnancy. Women in the control group received standard antenatal care whereas women in the intervention group received a minimum of 4 home visits from specially trained female social workers or obstetrical nurses and had access to a special 'drop in' support office at each study hospital. The purpose of these was to strengthen the woman's social network, and provide direct emotional support and health education.

The Committee anticipated that information and support needs might vary for women in different clinical scenarios and hence that specific recommendations might be needed for women:

- at increased risk of preterm labour (the risk could be either known prior to conception, early in pregnancy or later in pregnancy) and who may be having a planned preterm birth
- who are suspected or diagnosed to be in preterm labour (where preterm birth had not been expected)

However, the information included in the selected studies did not allow for further stratified analysis based on these different clinical scenaria.

3.1.2 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The following tables are presented:

Section 1

- Table 7: Quality assessment of qualitative studies
- Table 8: Information needs of hospitalised women at high risk of preterm birth (Gupton & Heaman 1994, N = 34)
- Table 9: Aspects of hospitalised women's' stressful experience of their possible preterm labour (Gaucher & Payot 2011, N = 5)
- Table 10: The expectations of women hospitalised for preterm labour regarding the prenatal consultation (Gaucher & Payot, 2011, N = 5)
- Table 11: The expectations of women hospitalised for preterm labour regarding the neonatologist (Gaucher & Payot, 2011, N = 5)
- Table 12: Parents' views of a prenatal tour of a NICU unit during a high-risk pregnancy (Griffin 1997, N = 13)
- Table 13: Information and support needs of women during the birth of their preterm baby (Sawyer 2013, N = 39)
- Table 14: Pre-delivery counselling experiences and information and support needs of parents with babies born between 23-26 gestational weeks (Young 2012, N = 10)

Section 2

- Table 15: GRADE findings for the comparison of antenatal information/support intervention with routine care in women with a high risk of preterm birth

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.A summary quality assessment for each qualitative study is given in Table 7.

Table 7: Quality assessment of qualitative studies

Study	Population	Methods	Analysis	Relevance to guideline population	Results transferrable to the population specified in the protocol
Gupton & Heaman 1994	Well reported[1]	Well reported	Well reported	Canada. The majority of women were white, married and had completed high school education. The mean gestational age was 31^{+3} weeks (range 26-36 weeks). The majority of women were hospitalised for spontaneous premature rupture of membranes (35%), twin pregnancy with cervical dilation and/or contractions (18%) or antepartum haemorrhage (12%)	Likely to be transferrable
Gaucher & Payot 2011	Well reported[2]	Well reported	Well reported	Canada. Women aged between 24 and 36 years, and gestational age of $26\text{-}30^{+2}$ weeks, admitted for preterm labour who had diverse reasons for hospitalisation and diverse social backgrounds	Very unlikely
Griffin 1997	Well reported	Well reported	Well reported	USA. Women aged between 20-42 years, with high-risk pregnancies and with the majority married and completed high school education.	Unlikely
Sawyer 2013	Well reported	Well reported	Well reported	UK. Women whose babies were born at 24 to 32 gestational weeks. The majority of women were white European, married, had completed secondary education, were employed and had 1 previous birth.	Unlikely
Young 2012	Well reported	Poorly reported [3,4]	Well reported	Canada. Women aged between 22 and 37 years, with high-risk pregnancies of 24-26 gestational weeks, educated to college or university level.	Likely to be transferrable

1. *A convenience sample was used,*
2. *5 of 7 women who agreed to participate were interviewed. Women were enrolled until no additional themes were identified. 3 of the women went on to have a term birth.*
3. *Interviews were conducted a long time after the birth (recall bias). All but 1 were conducted within 4 years, and the mean was 3.2 years after the birth.*
4. *Interviews are stated to be semi-structured but no further details of the questions asked are presented.*

The evidence from the qualitative studies that explored the areas of additional information and support needs for women at increased risk are presented in the following Tables 3 to 6. As the nature of this review was explorative, details of the main themes identified in each of these studies are given in the following tables along with direct quotations from studies' participants when necessary. Given that each of these qualitative studies explored different aspects of information needs, results are presented separately by study.

Table 8: Information needs of hospitalised women at high risk of preterm birth (Gupton & Heaman 1994, N = 34)

Italics represent direct quotations of women.

Non-italics represent field-workers' reporting of women's words.

What information is the most important for a mother who is at risk for preterm birth to know?

The Preterm Birth Learning Needs Questionnaire consists of 18 topics commonly included in educational programs for women at high risk of preterm birth. Participants scored each topic applying a score range from 0-20. The 5 most important topics (estimated using mean score (SD) were:

- the consequences of prematurity for the baby (mean 19.38 [1.65])
- problems of the newborn associated with preterm birth (mean 19.29 [1.66])
- how premature babies are cared for at home (mean 19.21 [1.82])
- how premature babies grow and develop (mean 18.71 [3.40])
- the signs and symptoms of preterm labour (mean 18.53 [2.60])

- 22/34 (67%) women indicated a need to know the possible risks or complications to the baby and the baby's chance of survival if *premature* birth should become a reality.
- 11/34 (32%) women indicated a need for reassurance - to be told that *"the baby will be OK" "for the staff to be supportive of the mother"* - and assistance in coping - to know "*how to prepare oneself psychologically and physically to face the stress, fear, etc."*
- 9/34 (27%) women indicated that it was most important for them to know how a premature birth could be prevented
- 6/34 (18%) women indicated that they wanted ongoing information on the condition of their baby as their pregnancy progressed.
- 3/34 (9%) women indicated that they wanted information on how to care for a preterm baby

What concerns do you have about being considered at high risk for preterm birth?

31/34 (91%) women indicated concern regarding the baby's survival chances, possible complications or permanent disabilities associated with prematurity and fetal development, especially lung maturation

Additional concerns:

- future care of the baby, how long the baby might be in hospital, whether it would be possible to breastfeed a premature baby, the uncertainty of the situation - *"so many unknowns, so many 'ifs' cause fear"*

Are there things that mothers at high risk of preterm birth do not need to know or should be taught?

All those responding to this question expressed a desire to be told "*everything"*

"I like to know exactly what is going on and get all the facts straight, so I can prepare myself both physically and psychologically", "The more knowledge that I have the more positive I feel. Not knowing the possibilities is frightening", "...if you are prepared for the worst and it doesn't happen, it feels great. If it does, I think that being totally unprepared could cause serious problems - both personally and in your family"

3/34 (9%) women indicated the need for honesty

"*Up front honesty is the best way to go. This is enough of a surprise; you don't need any more surprises because you weren't told something", "I prefer to know as much as possible and appreciate honesty in my doctors, coupled with human compassion"*

Several women included the need for advice for those who communicate information to women at high risk of preterm birth:

"*Give information gradually so mother has time to absorb and accept at her own pace", "Don't tell them something they may have done or not done has increased the risk. It adds to the guilt", "The use of alarming-sounding medical terms that when defined aren't life-threatening [is frightening] - not taking down to a mother but make sure she's familiar with the phases and terminology you're using - don't assume someone else has already explained - don't get overly technical - quoting statistics doesn't reassure - you want to know how your baby is doing"*

What would you tell someone (a friend or relative) to help them cope with being at high risk for preterm birth?

6/34 (18%) women indicated to tell other women to rest and relax

6/34 (18%) women indicated to tell other women to trust in the health care system

"I would try to remind them how advanced medicine is and the chances for survival are high", "Reassure them that absolute care is taken when handling preterm labour - competent doctors and

What information is the most important for a mother who is at risk for preterm birth to know?
nurses, modern technology", "Make sure you know what is happening at all times. Listen closely to what you are told and obey the medical staff" 4/34 (12%) women indicated the importance of keeping informed *"Inform yourself - talk to others who have gone through it", "To seek professional help and information and not to listen to those who know little or nothing", "Ask as many questions as they can regarding effects of preterm labour on baby and mother and read articles/books on preterm births"* Advice to maintain a positive attitude was also given: *"Don't go on a guilt trip", "Keep an optimistic and positive attitude no matter what", "Hope for the best, prepare for the worst", "Positive imagery and relaxation help"*

Table 9: Aspects of hospitalised women's' stressful experience of their possible preterm labour (Gaucher & Payot 2011, N = 5)

Italics represent direct quotations of women.

Non-italics represent field-workers' reporting of women's words.

Mourning
Having faced bad news regarding several aspects of their health or pregnancy, women tried to adapt quickly from living a healthy pregnancy to preparing for the challenges of prematurity, and found this to be difficult; the roles they had been preparing to play as parents changed. Some women at risk of a hysterectomy faced the possibility of no longer being able to bear children.
Perceptions of prematurity
All women had negative views about prematurity; several of them compared it with *'horror stories'* or '*hell'*. All women wished to avoid delivering prematurely.
Isolation
Women felt isolated from their usual support systems: 4 had been transferred from another hospital and their families lived far from the institution used for the present study. They expected their hospitalisation and bed rest to become prolonged, which was perceived as another difficult challenge to overcome. Furthermore, participants believed that they had lost their intimacy or privacy during their hospitalisation experience.
Powerlessness
Women expressed a strong feeling of powerlessness and loss of control. They believed that they had to accept all treatments offered to them to obtain the best possible outcome for themselves and for their baby: *"There is nothing we can do. We're a little powerless in all this. So we let ourselves go. We let go and we let them do anything to us."* (Mother 5) They were overwhelmed by the number of events experienced in a short period of time; the uncertainty of these events added insecurity and stress: *"Uncertainty, it's like vertigo or a precipice. And there is a lot of uncertainty. We don't know when I will deliver. We don't know how I will deliver. We don't know how it will go for the baby. We don't know what awaits the baby after. And we can get surprises, good or bad, for months after that. So it's a lot of uncertainty for a long time."* (Mother 3)
Main concerns
The baby's health and outcome were the main concerns for most women. One was most worried about her own medical condition. Another had been born prematurely herself, and focused on potential attachment difficulties as a parent and on a prolonged separation from her other children. All participants expressed some concerns about organising their families' lives around a prolonged hospital stay: *"Yesterday, I was preparing my children's things, but I didn't know what to prepare. I had to give them extra everything because I didn't know when I would be back. One of my children goes to school, one goes to daycare and the third one stays at home (…) and he's having his first birthday tomorrow. Now they are staying in 2 different households. One child is at my mother's house and 2 children are at my mother-in-law's."* (Mother 2)
Consultation as a stressor

Mourning
Women were generally informed by the obstetrical team in charge of their medical care that they would meet with a neonatologist. However, 1 woman had not been told this and found out only when approached about participating in the present study; she asked to partake in the study and was, therefore, included after she met with the team responsible for her care. Similar to other participants, she perceived the consultation as an additional source of stress: "Simply knowing that we'll meet the neonatologist is a stressor in itself. It's something really big (…) *The fact that I am being offered to meet the neonatologist before anything else makes me realise that, in my case, it is highly probable that I will deliver prematurely."* (Mother 5) However, all of the participants looked forward to the consultation so that their questions would be answered; they also hoped that the neonatologist could somehow reassure them, although the information they sought was not perceived as reassuring in itself: "*I think that the more the neonatologist will tell me, the more stressed I will be. But I don't like (…) not knowing the answers*." (Mother 1) "*I am looking forward to meeting them so that they can reassure us. Well, maybe not so that they can reassure us, but so that they can tell us the truth*." (Mother 2)

Table 10: The expectations of women hospitalised for preterm labour regarding the prenatal consultation (Gaucher & Payot, 2011, N = 5)

Italics represent direct quotations of women.

Non-italics represent field-workers' reporting of women's words.

Expectations from the consultation – Reassurance
Being reassured was the most important objective of the prenatal consultation. Women realised that they might receive worrisome information about possible complications related to prematurity. They hoped that the neonatologist would find ways to reassure them. "*Being reassured and just knowing what to expect. Because right now, I don't really know what to expect. So it's those 2 aspects, I think. (…) And what I can do as a mother to make sure, really make sure, that my baby is healthy and happy. Because that's really what I want."* (Mother 4)
Expectations from the consultation - Information and content
All women expected to receive clear, precise details and statistics about short-term and long-term complications of prematurity specific to their medical condition and related to gestational age. Some anticipated themes were respiratory distress, neurological complications, sepsis, feeding difficulties and length of hospitalisation. They hoped the neonatologist would describe some of the technology in the NICU. They reported having learned about prematurity and its complications from friends working in health care, from the media or from their own physicians. Only 2 of the participants underwent active follow-up for high-risk pregnancies before their enrolment in the present study. One woman suggested that parents visit the NICU before delivery, and believed that written documentation or pictures could be helpful.
Expectations from the consultation - Parental roles and responsibilities
Women expected the neonatologist to explain what their responsibilities would be and what would be expected of them. They wanted help organising their professional and family lives so they could be available for their baby. They wanted to know how they would be allowed to touch or hold their babies, and wanted to discuss breastfeeding and feeding strategies. Some wanted to know how they might participate in decision-making processes regarding their baby's treatment plans. One woman expressed concern about excessive care and had prepared questions to ask the neonatologist about her legal rights *"I'm not sure the neonatologists would make the same decisions that I would and I am worried they might impose their decisions on us."* (Mother 3)
Expectations from the consultation - Consistency of information
Women expected all of the different medical teams involved in their care to communicate among one another to hold consistent discourses about their situation. They reported inconsistency between health care providers' messages as an added source of stress.

Table 11: The expectations of women hospitalised for preterm labour regarding the neonatologist (Gaucher & Payot, 2011, N = 5)

Italics represent direct quotations of women.

Non-italics represent field-workers' reporting of women's words.

Expectations from the neonatologist - Structure of the consultation
Women who were interviewed believed that the best time to meet the neonatology team was before labour and delivery. They hoped their spouse would be present. They believed that the neonatologists should explain their role first, and then volunteer information about prematurity and its possible complications. One woman suggested that they sit down during the consultation. They all expected the neonatologists to be open to listening to their concerns and to provide time to answer their questions: *"Sometimes, I find it goes fast, that we don't have time to ask our questions. (…) It would only take the doctor an extra minute or 2, but it would save us from being anxious and having unanswered questions."* (Mother 3)
Expectations from the neonatologist – Trust
It was very important that the neonatologist instils a feeling of trust. Women wanted to know that they were in the best place for their baby and themselves to receive optimal care: *"We are handing over our lives and our baby's life into the hands of people we've never met before. So, if there's no trust, it's impossible."* (Mother 3)
Expectations from the neonatologist - Support and strategies
Most women expected the neonatologist to offer support and help them develop strategies to cope with their situation: *"It's very important to have a good doctor who can answer your questions and reassure you. (…) I mean, at least they're there to answer your questions and be supportive*." (Mother 4) Some also thought that neonatologists should refer them to other members of the health care team to explore various aspects of the problem. One woman, who had undergone in vitro fertilisation and fetal reduction, would have preferred to be referred to her own obstetrician for additional information and support.

Table 12: Parents' views of a prenatal tour of a NICU unit during a high-risk pregnancy (Griffin 1997, N = 13)

Italics represent direct quotations of women.

Non-italics represent field-workers' reporting of women's words

Benefits of the tour
Parents described benefits of the tour, including that it • decreased their fears • inspired hope for their baby's prognosis • provided reassurance about care in the NICU • prepared them for their baby's NICU hospitalisation All parents described at least one of these benefits, including 5 mothers who said the tour was overwhelming or difficult because of the appearance of newborns. *'Well, it's just hard when you see something like that. They were so young and so precious and fighting for their lives…. But you are more put at ease by seeing the care that they do receive and the attention that you get. But it's still frightening to see babies that small'*
Decreased their fears
Parents reported that because the tour was informative, it decreased their fears about the NICU and the type of care that their newborn might require.

Benefits of the tour

'Because it's so difficult to handle when you don't know. I know it's scary at times and I think the more education that you can receive about it, the better prepared you are to handle it should it happen'

Parents stated that just knowing that the NICU existed was helpful.

'Just to know that it was there. And I think it put my wife more relaxed and at ease the fact that they had a facility there that was nearby. We didn't have to worry about going to another hospital because they didn't have a special care nursery. Just the fact that it was there, we could see it, we know that it looked like and so if we were faced with that problem we were at least familiar with it.'

The tour gave mothers information about the NICU they needed to share with other family members. One mother indicated that she had gained an understanding of the unit and was better prepared to talk to her child about the NICU. Three of 4 mothers who were not who were not accompanied on the tour by the fathers reported that they had shared information about the NICU with the fathers, which was comforting to them. One of these mothers described her husband's reaction to their infant's admission to the NICU

'My husband was calm because I had already told him what to expect'

Inspired hope for newborn's prognosis

For several mothers, the tour inspired hope for their newborn's prognosis, especially when the mothers saw very premature infants who were said to be progressing well. One mother said

'The tour gave me hope that he was going to be fine. Seeing babies younger than him thrive……and then seeing the babies approximately his age survive thriving and doing well'

Another mother said

'It showed me that there is a lot more hope, and I thought about a few years ago or even 10 years ago, babies like this wouldn't have made it'

One mother said that after the tour, she was determined to take better care of herself and adhere to her prescription for bed rest to decrease the chance that her infant would be born prematurely.

Provided reassurance about care in the NICU

Parents reported that the tour was comforting and reassuring because it gave them an opportunity to observe the type and quality of care that the infants received. One mother said

'I was a lot more comfortable now seeing how they are giving the care and just seeing the environment they are in'

Parents felt encouraged when they observed the way that nurses cared for the infants. One mother said

'I saw the love, compassion and empathy that they showed for each of the babies there. So I knew he was going to be treated well'

Another mother commented

'Knowing they do care about them and they do realise that they are human and not machines…. You could feel that they really cared and worried'

It was especially helpful for the parents to see so many nurses and physicians in the NICU; hearing specific information about primary nursing also helped some mothers to feel more comfortable. Those mothers explained that it was reassuring to know that their questions could be answered because the primary nurse would know their infant.

Prepared parents for their newborn's NICU hospitalisation

All parents whose infants subsequently were cared for in the NICU reported that the tour prepared them for the experience. These parents explained that it helped to acquaint them with the NICU before delivery. One father said

'…..we didn't have to worry and wonder. It (the tour) made us understand how it all worked so that we were familiar with it when we did go there. And we didn't worry about what was going to be done because they explained everything beforehand. So, we pretty much knew exactly what their procedures were and how everything was dealt with instead of finding out as they did it…. The tour pretty much prepared us for what we were going to see when we went up there.'

One mother speculated on how her reaction to her infant's hospitalisation in the NICU would have been had she not toured the NICU while she was pregnant. She said

'I think it would have been a much more negative experience had I not toured and when there and saw the tubes in my baby's throat and the tape and everything. I don't know if I would have been able to take that….'

Benefits of the tour

For one mother, the tour's importance became evident after her infant was born

'Well I didn't really think much of it until she was born. I thought, well this is an interesting place and all that, but after she was actually born and brought here I kept thinking to myself, I'm glad I came and saw the place before she was born. It kind of helped ease knowing where she was going to be. It made it a lot easier'

Finally, a mother who initially was overwhelmed after the tour expressed how it prepared her for her newborn's admission to the NICU. She said

'I knew what to expect once I was there. So, I relaxed, and it wasn't overwhelming after I had him and he went to the (NICU)'

Evaluation of arrangement and conducting of the tour

Parents evaluated and provided suggestions on the way the tour was arranged and conducted and offered advice to other parents. In general, all parents recommended that parents in similar circumstances should be offered a prenatal tour of the NICU. One father said,

'I think you should go to the hospital and should try to get a tour of it…. You shouldn't be intimidated by the hospital and all the goings on in a nursery…. you have to get over the fear and ask the right questions and be familiar with that'

Parents advised that more health care providers suggest tours to parents diagnosed with a high-risk pregnancy. Two mothers also recommended that other perinatal health care providers should tour the NICU so that they can be supportive to parents. One mother perceived that her need to tour the NICU was not supported by the staff on the antepartum unit. She said

'So, I think some of them should be a little bit more realistic and help the patient prepare for their early delivery much more, rather than saying "Oh, I don't think they should have taken her there" or "it's too much for her"…. If they can just empathise with the patient and be a little more positive, I think the whole stay there would be a lot better as a result'

Parents also evaluated and gave specific advice in a number of areas, including

- tour arrangements
- type of information provided on the tour
- the behaviours and knowledge of the tour conductor

Evaluation of the tour - Arrangement of the tour

Parent's recommendations for timing of the tour varied. However, several recommended that parents tour the NICU soon after their pregnancies are identified as high-risk. One mother recommended that to minimise anxiety, parents take the tour soon after deciding to do so.

Parents who toured with their partners commented that having each other as a support person was helpful. They recommended that the tour be scheduled so that the partner or other support person could accompany the parent. One mother said

'Now that's the part I wish I could have changed. I wished my husband or somebody had been with me. But nobody was with me at the time.'

One couple also recommended that the tour should be scheduled around other appointments to avoid an additional trip to the hospital'

Evaluation of the tour - Type of information given on the tour

Parents reported that it was important to receive detailed information on the following

- newborns who had a diagnosis or gestational age similar to what was anticipated for their newborn
- a description of equipment for their newborns
- roles of staff members
- a description of the parental role in the NICU, including the visitation policy

A mother said

'Just by introducing me to people and explaining the various ages of and their survival and the babies that make it there. That was very comforting'

A parent suggested that parents meet with the neonatologist before the tour. It was important for parents to hear about the parental role. One mother said,

'They said if your baby was there, you could come up at any time, if you were the parent….. you could come in and they do encourage bonding with the baby, you can feed the baby, that type of thing. That did put me at ease.'

Benefits of the tour
However, not all parents perceived that they received adequate information on the parental role. A mother said, *'The parental role during the tour could have been more explicit because I was sure of my role during the tour, what would be expected of me or what I could do as far as caring for my baby.'* The need for more specific information became apparent to parents after their infants were cared for in the NICU. These parents indicted that they wanted more information on expectations for their role in the NICU, breastfeeding, sibling visitation, and the potential for the baby to be transferred from the NICU to another unit before discharge. Two parents suggested that hand-outs would supplement or reinforce information that was given during the tour and assist parents to inform family and friends about the NICU. Parents reported that the tour should be individualised to meet the specific needs of parents. Parents perceived the tour as individualised when they went as a couple or an individual rather than in a group, had an opportunity to ask questions, and saw newborns who had a diagnosis or gestational age similar to that expected for their newborn. Therefore, it was critical for the nurse conducting the tour to know the parents' maternal-fetal diagnosis. Several parents made additional suggestions, such as having an opportunity to go on a second tour or changing the order in which the NICU patient care areas are shown; these demonstrate the parents' individual needs
Evaluation of the tour - Behaviour/knowledge of the tour conductor
Most parents reported that the nurses who conducted the tours were knowledgeable and comforting. These nurses were describe as compassionate, concerned, helpful, and considerate of the time parents needed to understand the information and ask questions. One mother said *'She was a warm lady… putting her hand on my arm, and just somebody touching me made me feel like (I was) relaxed…'* One father stated that the nurse who conducted the tour *'knew what was going on and knew the staff, and the staff apparently thought a lot of her…'*

Table 13: Information and support needs of women during the birth of their preterm baby (Sawyer 2013, N = 39)

Italics represent direct quotations of women.

Non-italics represent field-workers' reporting of women's words.

Overall satisfaction with care
'Overall, how satisfied would you say you were with the care that you received during the birth?' Extremely satisfied with care and nothing could be improved = 31/39 (80%) parents Generally satisfied with care but certain things could have been improved (e.g. provision of information) = 7/39 (18%) Dissatisfied with her care = 1/39 (2%)
Staff professionalism - Information and explanation
33/39 parents (39 mothers, 4 fathers and 6 mothers in a couple) mentioned this theme Provision of information was really important and was mentioned by 33 participants (85%). They wanted to be told what would happen during the birth (particularly if they were having a caesarean section), what type of anaesthetic would be administered, and what was going to happen to their baby when he or she was born. The anaesthetist was someone who stood out in participants' minds in terms of providing detailed information and explanations. "so we actually go down into the operating theatre and again the anaesthetist was there and talking to [us] as she said 'I will stay with you the whole time'....and she talked us through everything that was happening and for both of us that was just outstanding, absolutely" (1 Mother, C/S). It was perceived that someone taking the time to explain what was happening helped them cope with the situation and made the experience less 'traumatic' *"it was a traumatic experience. I think, if it hadn't been explained to us exactly step by step it would have been more traumatic...It was just so much easier, because they did go out of their way and they explained absolutely everything to you*" (2 Father, C/S).

Overall satisfaction with care

Participants also wanted information to be explained in a way they could easily understand.

"They told you everything that was going on, what was happening. They make sure you understood, make sure he [father] understood what was going on" (7 Mother, C/S).

One mother wanted more information than she was given during the birth. She had some medical knowledge, and would have liked to know about what was happening throughout her operation in more detail ,

"So you feel prodding, and I wasn't told much. I felt I wasn't told much when I was actually in there and hadn't, I didn't know when they'd started to open me up, cut me open...So I didn't know what they were doing, water's, broken my waters....None of that was ever communicated to me." (8 Mother, C/S).

Six participants (15%) commented that the different members of staff introduced themselves and told them what they would be doing. This helped them feel less like they were in a room with people they did not know.

"I mean they were all very, I remember there being people in the room and they were all introducing themselves and what they did" (6 Mother, C/S).

Staff professionalism - Staff calm in a crisis

11/39 parents (7 mothers, 1 father and 3 mothers in a couple) mentioned this theme. Nineteen participants (49%) described feeling frightened of what was going to happen during the birth and for the outcome of their baby. However, the calm attitude of the staff helped them feel more comfortable and at ease.

"You're not as frightened. It's daunting going in a room when you've never been in. All your bits are going to be on show. And you're worried about your children. Are they gonna survive? Are they gonna be born stillborn? You know....they were so relaxed, they made me feel so comfortable" (4 Mother, C/S).

"I think it was them staying relaxed. Even though it was a rush, it was a stressful time, you could see that, but they were very good at staying calm. But I suppose that's their job in a way, but they were actually very good at it" (19 Mother, C/S).

Staff professionalism - Confident and in control

8/39 parents (8 mothers, no couples) mentioned this theme. The confidence displayed by staff was important to participants as it demonstrated capability and control. One woman described that the surgeon in charge of her operation portrayed total confidence.

"And the way he mastered the team, I got the absolute… he had an air of confidence and in control of the entire team. He knew what every person was doing. And he was very commanding as well" (5 Mother, V).

Having confidence in the staff seemed to make it easier to hand over control to them. One woman described that she did not feel that she needed to be in control. She trusted the staff and was happy for them to take control of the situation.

"Absolute confidence in the staff. I didn't feel like I needed to know every step of the way. I was able to just step back, realise that control was not mine. The control was where it should be, with professionals, and they would take good care of them [the babies]" (5 Mother, V).

Four mothers (10%) described the doctors as being firm with them, but said this was exactly what they needed. They wanted the staff to take control of the situation and tell them what to do

"it was very very quick, very shouty: 'you have to do this, you have to do this now'. It was made very clear to me if I didn't push he wouldn't survive. Erm, which was absolutely fantastic, which was what needed to be done" (3 Mother, V).

Staff professionalism - Staff not listening to the woman

8/39 parents (6 mothers, 1 father and 1 mother in a couple) mentioned this theme. This area contributed to a negative experience of care for participants. Seven mothers (18%) expressed disappointment that the staff did not always listen to what they had to say. These women described telling staff that they felt they were in labour and close to giving birth, and often the staff did not believe or trust what they were saying, which left women feeling ignored and frustrated.

"And then when I started to get pains, I started to tell the midwives, or the nurses that were there. And felt that they didn't actually believe me, because they put me on monitors. And where my waters had gone, the monitors don't pick up the contractions as well. So they were just saying 'no, no, no, the contractions are not real.....basically [you] can't be feeling this amount of pain" (19 Mother, C/S).

Overall satisfaction with care

One woman described how she tried to tell the midwife that she was about to have her baby, but was not listened to, and as a result no staff were present at the birth (23, Mother, V; Table S1).

"*The only kind of downside to it, was I kept saying to her, all my family have very quick labours..... I kept saying to her I need to push I need to push and she said I've only checked you half an hour ago, you're only 3cm ..... and she went I'm just popping out the room.....and at that point I just pushed and her head popped out, and no one was in the room apart from me and my partner*"(23 Mother, V).

Staff empathy

21/39 parents (15 mothers, 1 father and 5 mothers in a couple) mentioned this theme. Participants' experiences of their care during the birth were also influenced by the interpersonal interactions with care providers, in particular by caring and emotional support, and encouragement and reassurance.

Staff empathy - Caring and emotional support

Twenty-one participants (54%) spoke about the 'warm and friendly' attitude of the staff. In terms of satisfaction with their experience, it was important that they were treated in a pleasant manner. Two very different quotes illustrate the importance of the staff treating them as an individual and receiving personalised care.

"I just found our experience very good, it was very I suppose personal in a sense. I wasn't, I didn't feel like a piece of meat. I felt like a human.....and people were caring" (3 Mother, V).

"*But the midwives that should have shown me compassion in the beginning didn't. They were just not bothered"* (30 Mother, V).

Mothers spoke about the importance of a member of staff always being with them, and this generally referred to the presence of a midwife.

"one of the nurses just steps out the way, holds your hand, and talks to you.....So it's just nice to have someone there, talking to you and holding your hand and sort of walking you through everything instead of everyone buzzing around" (2 Mother, C/S).

One mother whose baby was born with many complications and died less than 24 hours after the birth described how the caring and supportive attitude of one midwife made her experience of the birth less traumatic than it could have been.

"the midwives were incredible, so during the birth,...we had this amazingly lovely kind of West African um midwife who was, oh just love, like lovely, so nice so, supportive and caring and empathetic and everything that you could possibly want and just really supportive and, so the birth process itself actually, in the scheme of things was relatively easy thing then to go to because I felt very supportive... and she was so lovely" (32 Mother, V).

Staff empathy - Encouragement and reassurance

23/39 parents (16 mothers, 3 fathers and 4 mothers in a couple) mentioned this theme. Twenty-3 participants (59%) mentioned wanting encouragement and reassurance from the staff. They understood that the staff have to be realistic about the situation and the prognosis for their baby, but found it really helpful and encouraging if the staff were able to reassure them in some way.

"Obviously so they can't lie... but just kind of being positive I think really really helps um 'cause you know, it's it's quite terrifying not having had an operation before and um you know you don't quite know what to expect and things so just people you know just reassuring you, saying nice things" (14, C/S Mother).

"And that's what you want is reassurance, that time, and so yeah, it was very good" (1 Father, C/S).

Encouragement from the staff also influenced their experience with care at birth. One woman who was feeling scared and tired described how a midwife encouraged her to continue.

"Yeah we were whisked upstairs and at that point I couldn't feel the hand moving so I really freaked out. One of the midwives was there and she could feel a pulse, calm down, gave me cuddles, really calmed me down and said 'you're ok, you've got to do this, you'll get through it.' Really sort of geed me up and gave me that extra bit of strength really" (3 Mother, V).

Another mother described how praise from a midwife contributed positively to her experience.

"You know she was constantly praising "you, you're doing really well, just breathe through it", you know and things like that whereas you get some midwives who just aren't the nicest, so um, the fact that she was as nice as she was" (23 Mother, V).

Involvement of the father

Overall satisfaction with care
16/39 parents (7 mothers, 5 fathers and 4 mothers in a couple) mentioned this theme. It was important to the mothers that the baby's father was involved in the birth, and the extent to which staff involved them contributed to a positive or negative experience with care. For example, 2 women (5%) described how the staff tried to delay the caesarean section so the father could get there for the birth. Three women (8%) also discussed that they had planned their partner's involvement in the birth, and therefore appreciated any effort the staff made to make them feel more involved. "He got there really quick. But they involved him, once they brought him [to the operating theatre], they told him everything while he was getting changed, what to expect." (2 Mother, C/S). "I found it reassuring that they were very happy with [husband] to be sort of looking over their shoulders and sticking his nose in and whatever, so there was no "stand over there dad" (12 Mother, C/S). Four women (10%) talked of regret that the baby's father was not able to participate more and was not encouraged to feel more involved in the birth by the staff. *"Erm he found it very awkward...When they were being born he just sat out there, wasn't really able to participate...So he felt like a spare part.....when we were rushed to the surgical unit… there were so many people in the room, he felt he didn't know where to stand. He didn't want to get in the way. He knew he needed to get there.. let everyone get on with their job. But he felt in the way"* (5 Mother, V). *"I don't think anyone even really spoke to [the father], I mean I I'm reflecting on it now, I don't think anyone did, how was he involved, he wasn't involved at all, so yeah how are you feeling, is there anything I can do, yeah*" (31 Mother, V). It was also important to fathers that they were encouraged to feel involved in the birth. One of the fathers interviewed described how fathers are not normally made to feel involved in the birth, but that this time he was involved from the start. "Because normally they don't talk to you. To a woman, they say 'right we've got to do this, got to do that' so the lady knows exactly what's happening to her and why. For the bloke.....'Stay down the pub and we'll give you a ring when it's all done and you can come up when it's all nice and clean, in a blanket.' But with [name of hospital], it was completely different" (2 Father, C/S).
Birth environment
17/39 parents (11 mothers, 3 fathers and 3 mothers in a couple) mentioned this theme. Participants discussed features of the delivery suite and operating theatre that contributed to their positive experience at the birth. Five participants (13%) described that the radio was playing during the birth, which made the environment seem less frightening. *"You know they didn't make it scary in any way at all, they were all quite happy, I think the radio was playing, which was good, you know things like that. The environment didn't seem scary*" (1 Mother, C/S). Three women (8%) also commented on the views from the windows of the operating theatre. It helped them feel 'connected' with the outside world and help take their mind off things. *"it can take your mind off it a bit rather than just sort of grey walls um so yea so I mean that's very much what we remember actually and often sort of comment on it you know to people*" (14 Mother, C/S).

Table 14: Pre-delivery counselling experiences and information and support needs of parents with babies born between 23-26 gestational weeks (Young 2012, N = 10)

Italics represent direct quotations of women.

Non-italics represent field-workers' reporting of women's words

Content - Theme: Knowledge
None of the families had any previous knowledge regarding preterm birth. (Family 1 did have 2 children who were extremely preterm births at 24 and 26 weeks' gestational age, but they responded in reference to their first child.) Before being counselled, most parents had assumed that with extreme preterm labour, there was no chance of survival.

Content - Theme: Knowledge

*[He] told me all the issues…I didn't even think that… it was an option to even have a [baby at] 26 weeks…. We were, in all honesty and bluntness, prepared to have a burial for this child. We didn't know what to expect, or severe abnormalities, and we talked about it…through the nigh*t. (Family 3)

All parents wanted information that was clearly stated regarding the likelihood of survival and what to expect at delivery. All parents desired to be fully informed of the immediate risks for their child.

…what we needed would to be told that [they] would administer steroids, his best chances are that you last another 48 hours there could be complications if he doesn't, um, vis-à-vis, breathing… moment by moment until his birth happens and then [they'll] let you know what you have to face. (Family 4)

One set of parents recounted the experience of having multiple members of the neonatal team counsel them about various aspects of the NICU including ongoing research projects. They believed that this manner of counselling lacked compassion and would have preferred fewer counsellors focusing on information of immediate relevance such as survival and prognosis.

…it would almost be a bit more compassionate to tell people we'll deal with it once the baby comes then, you know, we'll see what problems arise, there could be some, but going into the great detail before added a lot of stress to the fact that we were early and all of those things just kept going through our head. (Family 4)

Content - Theme: Resuscitation wishes

Most families did not recall explicitly being asked about their resuscitation wishes.

We want to focus on just the baby and then if that happens, then we'll deal with it at that time. But we never had that opportunity, other than just between ourselves…they should bring it up and they should discuss it with the parents and then the parents have that opportunity to say, "no, we don't want to talk about it"… (Family 8)

In retrospect, 3 couples (Families 3, 5 and 9) may not have chosen resuscitation, had they known all of the potential complications of prematurity. The parents who lost one twin (Family 9) believed the other twin suffered to such an extent while in the NICU that they would not have proceeded with resuscitation had they known "what was in store." One mother was counselled alone in the middle of the night and believed her awareness was affected by medication.

But, to be honest, if somebody would have told me that this is what my life would be like, I don't think that I would have chosen resuscitation. I might have chosen to hold (twin A) for the 7 minutes that he cried and let him die. (Family 5)

Even parents who had deferred the ultimate decision to the team indicated that parents should have clear opportunities to express their wishes.

Content - Theme: Additional resources:

All parents suggested that written information, in addition to verbal counselling, would have helped them feel informed and supported. The parents who were provided with pictures with NICU found that they enhanced their understanding (Family 1). One mother suggested having a video or a virtual tour of the NICU (Family 10) to help prepare for this experience.

Process - Theme: Timing of counselling during pregnancy:

Most of the families were seeing high-risk obstetricians during the pregnancy. They wished that they had received counselling about prematurity when the pregnancy was first deemed high-risk. Three couples believed they were falsely reassured by their doctor? Physician about the risks of preterm delivery (Families 3, 4 and 9). One mother, who finally conceived via in vitro fertilisation after having multiple miscarriages due to an incompetent cervix, recalled:

They were just saying don't worry about it though, so I said OK. But I knew when I got pregnant it was pretty iffy all the way. (Family 4)

One couple (Family 1) did suggest that early information regarding prematurity would cause needless worry; this couple was one of 2 who did not need to see a high-risk specialist before delivery. Two couples (Families 3 and 6) commented that while the risks for conditions such as Down syndrome are discussed antenatally, there is no information routinely given about prematurity even though it is common. They suggested that written pamphlets be available at obstetricians' or family physicians' offices.

Process - Theme: Timing of counselling during maternal hospitalisation:

Seven families waited in hospital more than 24 hours, and even couples requiring emergent management waited a few hours before birth occurred. One mother (Family 5) recalled being admitted twice with spotting at 24 and 25 weeks before going into labour at 26 weeks. She was not

Content - Theme: Knowledge
counselled until the third admission in the middle of the night. By then she was anaemic and on medications that affected her awareness, and fell asleep during the conversation.
Process - Theme: Ongoing counselling:
After the initial emergency counselling, parents wanted the opportunity to hear the news again, together, if there was time (i.e., if delivery was not imminent). The mother who was admitted for weeks after the initial counselling, due to an incompetent cervix, and her partner did not see the team until after the birth. *…if they'd have come in even one or 2 at a time instead of 6 at a time, and spaced it out and then revisit a day later, just to even pop their head in to say hi, how are you doing. Oh, I'm OK….that would have made the just before the birth thing a whole lot easier…* (Family 4) Although parents acknowledged that physicians are busy and cannot always cater to parents' schedules, they believed that a follow-up visit after parents have had a chance to digest information and formulate questions would improve the communication process.
Process - Theme: Impact of counsellors' attitude:
Parents indicated that counsellors' messages regarding the survival and prognosis of their extremely premature neonate should be performed in a compassionate manner and that hope should be conveyed after the decision to resuscitate had been made. *I don't know what the legalities are, but my feeling at the time was that oh, we needed a lot of positive reinforcement at that moment and what we got was the exact opposite*. (Family 4) Parents believed that some counsellors were unnecessarily negative. One mother recalls a physician who simply stated that the team would not proceed with resuscitation. *He said to me, OK, if the baby is born today, what we are going to do is just wrap it up, we won't do any heroics, we'll just wrap him up you can hold him for a little bit and then he'll probably just go*. (Family 1) This mother recalled being devastated by this mental imagery and described how she subsequently avoided this particular physician throughout the child's course in the NICU.

The second part of this evidence review, which aimed to test the effectiveness of interventions or packages of care for women at risk for preterm labour, included the results of 2 RCTs and a qualitative study of a convenient sample to evaluate parents' views of a prenatal tour of a NICU. The qualitative study is considered as indirect evidence for this part of the evidence review given the non-comparative nature of its study design and the limitation on testing the role of intervention. However, it gives insight of parents' care experience, which was considered as complimentary on the results of 2 RCTs.

Table 15: GRADE findings for the comparison of antenatal information/support intervention with routine care in women with a high risk of preterm birth

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Routine care	Relative (95% CI)	Absolute (95% CI)	Quality
Postnatal depression											
1 study (Oakley 1990)	Randomised trial	Serious 1	No serious inconsistency	No serious indirectness	Serious2	None	92/230	107/228	RR 0.85 (0.69 to 1.05)	70 fewer per 1000 (from 145 fewer to 23 more)	Low
Less than very satisfied with antenatal care											
1 study (Villar 1992)	Randomised trial	No serious risk of bias	No serious inconsistency	No serious indirectness	Serious2	none	51/945	45/942	RR 1.13 (0.76 to 1.67)	6 more per 1000 (from 11 fewer to 32 more)	Moderate

CI confidence interval, RR relative risk
1 Oakley 1990: Blinding of participants, clinicians and outcome assessors was unclear
2 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID.

3.1.3 Evidence statements

Section 1; themes of additional information given prior to birth

One descriptive study (n=34) found that women hospitalised for being at high risk of preterm birth rated all 18 teaching topics on a questionnaire as being important to know. Teaching topics relating to the consequences of prematurity for the baby and problems for the baby associated with preterm birth were the most important for women and their families/partners. The responses to open ended questions confirmed that the women's/families overriding concerns were for the wellbeing of their unborn babies.

One qualitative study (n=5) of women who had a baby born before 32 gestational weeks reported that the majority of women were very satisfied with their care. Four main themes emerged as important determinants of positive or negative experiences of care during preterm birth: staff professionalism, staff empathy, involvement of fathers and the birth environment.

One qualitative study (n=39) found that women admitted to hospital for preterm labour reported that their experience was stressful and that they felt a sense of isolation and powerlessness. Although the prenatal consultation with a neonatologist was also regarded as stressful, they expected that it would empower them by providing reassurance, clear information regarding the consequences of prematurity for their baby and regarding their parental role. Consistency of information from medical teams was important. They expected to have trust in their neonatologist who would provide support and strategies to help cope with their situation.

One qualitative study (n=10) of pre-delivery counselling for extremely premature babies identified the content and process of the counselling as the 2 main themes concerning parents. Parents wanted clearly stated information regarding likelihood of survival, what to expect at delivery and wanted to be informed of the immediate risks to their baby. Parents reported that additional written information would be helpful. Only 40% recalled explicit questions regarding their resuscitation wishes but believed that parents should have the opportunity to state explicitly their resuscitation wishes. Parents would have liked to receive counselling when the pregnancy was diagnosed as being at high-risk of preterm delivery. The timing of counselling when mothers were admitted was not always optimal and that ongoing counselling would be helpful, as well as counselling for fathers. Counselling should be performed in a compassionate manner.

One qualitative study of a prenatal NICU tour during a high-risk pregnancy (n=13) identified that the tour was of benefit to parents in that it decreased their fears, inspired hope for their baby's prognosis, provided reassurance about NICU care and prepared parents for their baby's NICU stay. Parents commented on the timing of the tour and their wish to have their partner present. They believed it was important to have detailed information regarding babies with a diagnosis or gestational age similar to that anticipated for their baby, a description of equipment for their babies, the roles of staff members and a description of the parental role in the NICU, including the visitation policy and feeding policy.

Section 2; interventions or packages of care designed to provide additional information prior to or during birth compared to usual care

Two RCTs of 458 and 1887 women reported that there were no significant differences in postnatal depression (low quality) or satisfaction with care (moderate quality) when women with a high risk of preterm birth who received additional antenatal support were compared to those who received routine care.

3.1.4 Health economics profile

No search for health economic evidence was undertaken for the question on what additional information and support should be given to women (antenatally or during labour) and their families where the woman is at increased risk of preterm labour, or is suspected or diagnosed to be in preterm labour, or has a planned preterm birth. It was anticipated that the opportunity cost of any recommendations arising from this review would be negligible and, reflecting that this is usually not an issue with important cost implications, it is very unusual to find published health economic evidence assessing the provision of information and support in this area. Furthermore, the Committee were keen that existing NICE guidelines on patient experience in adult NHS services be followed.

Therefore this question was not identified as a priority for health economic analysis.

3.1.5 Evidence to recommendations

3.1.5.1 Relative value placed on the outcomes considered

3.1.5.2 Consideration of clinical benefits and harms

As a first principle, the Committee agreed that kindness, compassion and empathy were crucial principles in all interactions between clinicians and the pregnant woman and her partner because otherwise the content of any information imparted becomes redundant.

The effectiveness review found evidence that the additional care provided in the packages examined did not improve the outcomes of postnatal depression or satisfaction with care. However, the Committee supported the recommendations on the themes revealed by the qualitative review for the delivery of information and support to women at high risk of preterm delivery and to their partners and families.

The Committee noted that women might prioritise information about the risks and consequences of preterm birth for their babies over information regarding their own outcomes of preterm birth. However, they acknowledged that, due to time pressures during preterm delivery, midwives and antenatal staff often have to fine-tune information given before the preterm birth to provide the most important details regarding the delivery. They therefore recognised the value of providing information and support as early as possible in the antenatal period, taking into account the level and nature of the risk(s) and the imminence of delivery for women with suspected, diagnosed or established preterm labour.

The Committee recognised that not all women would be aware of the symptoms and signs of preterm labour or have prior knowledge of the care that would be offered to them and their babies if they deliver preterm and considered this to be an important component of information provision. The Committee considered that women should receive information from healthcare professionals regarding the immediate and long-term consequences to the baby of preterm birth.

Reassurance and trust in health professionals were recurring themes in the qualitative evidence. The Committee has interpreted this as requiring the provision of honest and realistic information about a woman's individual situation. Honesty about the level of certainty regarding possible outcomes was considered valuable among members of the Committee although it was recognised that healthcare professionals would need to achieve a balance between preventing further anxiety whilst not withholding information. Although the Committee recognised that not all women would want statistical predictions of future events, if statistics regarding risk were to be provided, then this should be done in line with the NICE Patient Experience Guideline. It was recommended that natural frequencies might be more acceptable in providing information to women and their partners or families.

The Committee agreed that verbal information should be supplemented with written information and recognised that it could be helpful to provide guidance on where to access further information including information in other formats/media and support organisations.

Information provision should be revisited over time, for example because of a changing clinical situation, or because information provided verbally may not be absorbed at the time it is given, and questions may come to mind subsequently.

Consistency of information was also considered a key consideration. Inconsistencies in information provided to the woman by different members of the healthcare team as well as differences in the information given to each parent were acknowledged a source of anxiety that led to a reduction in trust. The Committee believed that improved communication between the healthcare team could help mitigate against this.

Other themes of information provision that the Committee considered particularly relevant related to mother's attachment with their newborn, the stress of hospitalisation necessitating separation from older children and disruption to daily life and the importance of joint decision making with healthcare professionals. The role of neonatal staff in listening to women was considered critical, and women should be offered the opportunity to speak with a neonatologist within 24 hours if they wish to do so.

It was noted that providing women at high risk of preterm labour with a tour of NICU as soon as possible after the risk is identified might increase their confidence, knowledge and reduce anxiety. Meeting staff and seeing equipment seems to reassure parents that if their baby was born prematurely there were mechanisms and people in place to help their baby. The Committee considered it important that these tours were individualised (e.g. depending on level of risk, the tour might include some rooms and not others) and that partners were encouraged to participate in this tour where possible. It was felt that the tour needs to be 'real' not virtual because part of the reassurance will be derived from contact with the healthcare professionals during the tour. In cases where women are not clinically able to visit the unit, a virtual tour and discussion with staff from the neonatal unit should be facilitated.

Finally, the Committee recognised that information regarding resuscitation and withdrawal of care was important to inform discussions that women at risk of preterm or very preterm labour may wish to have antenatally. It was acknowledged that the woman's or parents' wishes might change over time, following reflection on their present situation or as situations changed and hence on-going opportunities for dialogue regarding resuscitation and treatment should be made available.

3.1.5.3 Consideration of health benefits and resource uses

The recommendations made by the Committee have negligible resource implications and aim to follow NICE guidelines on patient experience in adult NHS services. They reflect best current practice where information and support to women will always be an important component of the package of care that women receive. The recommendations stress key elements of information and support that women should be made aware but do not reflect a decision about competing information and support alternatives.

3.1.5.4 Quality of evidence

All of the qualitative studies used appropriate research methods and were well reported. However, only 2 were considered to have results that were directly transferable to women receiving NHS care in England. The Committee considered the likelihood of recall bias in one study which decreased their confidence in its findings, although similar themes were expressed in a second study of a similar population of very preterm babies.).

3.1.5.5 Other considerations

The recommendations in this section were based on both the interpretation of qualitative evidence reviewed and on Committee's expert opinion. The Committee was aware of several reports that aligned with the findings of this review and with their own experience, for example, the report from Bliss. Their POPPY report recommended that a tour of a neonatal unit be provided for parents at high risk of preterm labour to provide information and prepare them about what to expect. This is reassuring in terms of guidance for current practice already including delivery of information and support and which implements available evidence. Furthermore the NICE Patient experience guideline (CG138 http://www.nice.org.uk/guidance/cg138/chapter/guidance) makes recommendations, the principles of which should be used for women who are at high risk of or who are in preterm labour.

This section was identified as a priority area for equality issues in the scoping stage of guideline's development. The Committee followed the suggestions from the stakeholders in the scoping stage and recommendations on additional information and support were supported for women at increased risk of preterm labour with suspected, diagnosed or established preterm labour or having a planned preterm birth. The recommendations supported the provision of information in different formats (for example both oral and written information) and following the principles in the NICE guideline on patient experience in adult NHS services to address any equality issues.

3.1.6 Recommendations

1. **When giving information and support to women at increased risk of preterm labour, with suspected, diagnosed or established preterm labour, or having a planned preterm birth (and their family members or carers as appropriate):**
    - give this information and support as early as possible, taking into account the likelihood of preterm birth and the status of labour
    - follow the principles in the NICE guideline on patient experience in adult NHS services
    - give both oral and written information
    - describe the symptoms and signs of preterm labour
    - explain to the woman about the care she might be offered.

2. **For women who are having a planned preterm birth or are offered treatment for preterm labour in line with recommendations 30 – 43 (and their family members or carers as appropriate), provide information and support that includes:**
    - information about the likelihood of the baby surviving and other outcomes (including long-term outcomes) and risks for the baby, giving values as natural frequencies (for example, 1 in 100)
    - explaining about the neonatal care of preterm babies, including location of care
    - explaining about the immediate problems that can arise when a baby is born preterm
    - explaining about the possible long-term consequences of prematurity for the baby (how premature babies grow and develop)
    - ongoing opportunities to talk about and state their wishes about resuscitation of the baby
    - an opportunity to tour the neonatal unit
    - an opportunity to speak to a neonatologist or paediatrician.

4 Prevention of preterm birth

4.1 Introduction

Prevention of preterm birth would reduce the number of babies suffering from prematurity and its complications and would reduce the burden of disability that results. Effective prevention depends upon identifying the woman at high risk of preterm delivery, either because they are at risk of spontaneous preterm labour or have medical complications which make them likely to require planned preterm birth. This chapter addresses care for women with recognised risk factors for spontaneous preterm labour, that is, a history of spontaneous preterm birth or mid-trimester loss, a history of P-PROM in previous pregnancy or a history of cervical surgery (for example knife cone biopsy).

Prevention also depends upon availability of effective interventions which can be offered to high-risk women. This chapter reviews the evidence for the use of prophylactic vaginal progesterone and prophylactic cervical cerclage.

The search strategies for this chapter can be found in Appendix E, the excluded studies for this chapter can be found in Appendix G, the evidence tables for this chapter can be found in Appendix H, and the forest plots for this chapter can found in Appendix I.

4.2 Prophylactic progesterone

4.2.1 Introduction

Two studies published in 2003 suggested that vaginal progesterone and/or intramuscular 17 hydroxyprogesterone caproate could prevent preterm birth in women at high risk. Additional studies since then have added further to the literature, but uncertainties about efficacy, long term outcomes and appropriate target groups remain. Both the American College of Obstetricians and Gynaecologists and the Canadian Society of Obstetrics and Gynaecology endorse the use of progesterone prophylaxis in selected women. The Royal College of Obstetricians in the UK has indicated that progesterone should not be used for this purpose unless as part of a randomised trial. Direct extrapolation from US guidelines are not possible, because17 hydroxyprogesterone caproate, given intramuscularly and widely used in the US, is unavailable in the UK. The available UK formulations, vaginal or oral progesterone, are the subject of the review in this chapter.

4.2.2 Review question

What is the clinical effectiveness of prophylactic progesterone (vaginal or oral) in preventing preterm labour in pregnant women considered to be at risk of preterm labour and birth?

4.2.3 Description of included studies

One Cochrane SR (Dodd 2013) and one IPD SR and meta-analysis (Romero 2012) were included for this question. We only considered studies that investigated the role of progesterone (either vaginal or oral) as prophylactic measures to prevent preterm delivery in asymptomatic women. Only 9 out of 36 RCTs included in the review by Dodd 2013 and 4 out of 5 in the IPD meta-analysis by Romero 2012 matched our protocol by reporting the outcomes of interest and are presented in this section. Systematic reviews and RCTs focussing on women with existing signs or symptoms of preterm labour were excluded.

Results are presented by different risk factors for preterm delivery and within each of these sections, results are given by comparison of interventions.

Women with a history of previous spontaneous preterm birth

7 RCTs from the included Cochrane review (Dodd 2013) were included for this section reporting results for the following comparisons:

- vaginal progesterone compared with no treatment (Akbari 2009, Majhi 2009)
- vaginal progesterone compared with placebo (Cetingoz 2011, da Fonseca 2003 and O'Brien 2007)
- oral progesterone compared with placebo (Glover 2011, Rai 2009)

The timing of assessment of cervical length and the inclusion of women with cervical cerclage varied between the included studies. In one trial (Mahji 2009) cervical length was assessed using ultrasound at 16-24 weeks and women with cerclage were explicitly excluded from the study. Another study (Glover 2011) followed a more detailed pattern of cervical length measurement; ultrasound to assess cervical length was performed at baseline prior to 20 gestational weeks. If the cervical length was >25mm then a repeat measurement was performed at 24 weeks. If cervical length at baseline was 10-25mm, women had fortnightly measurements until 24 gestational weeks. Assessment was performed every week in women whose baseline cervical length was between 5 and 10mm and cervical cerclage was offered to women with a cervical length of under 5mm prior to 24 gestational weeks. Use of cervical cerclage was an exclusion criterion for the study. In 2 RCTs (O'Brien 2007, Cetingoz 2011) in this section, ultrasound was performed between 16-20 weeks and 20-24 weeks and women with planned or current cerclage were excluded from analysis. In another RCT (da Fonseca 2003), no details were given regarding ultrasonography to assess cervical length but 3.5% of the study population had cervical cerclage prior to recruitment as this was included in the definition of a women being at high risk for preterm labour and birth. One study (Rai 2009) reported that ultrasound was performed in the second and early third trimester to assess cervical length and the need for cervical cerclage without further details on timing or whether cerclage was performed. Finally, no details regarding cervical length measurement or use of cerclage were available for one RCT (Akbari 2009). .

Two of the studies in this section included mixed populations; the study by Cetingoz 2011 included a high proportion of women with multiple pregnancy (45%) and a low proportion of women with uterine malformations (8%). However, results were presented separately for women with a history of previous spontaneous birth and only these results are included for the purposes of this review. The study by da Fonseca 2003 included a mixed population consisted of 8% of women who were defined as being at high risk because of uterine malformation or an incompetent cervix (with cerclage placement) but no subgroup analysis was available therefore the results from the whole sample were used in this analysis..

Women with a short cervix in the current pregnancy

4 RCTs from the IPD review (Romero 2012) contributed to this section. All 4 RCTs (Fonseca 2007, Hassan 2011, O'Brien 2007, Cetingoz 2011) compared vaginal progesterone with placebo. Three out of 4 RCTs were multicentre studies conducted in UK, or different countries around the world (USA, Belarus, Chile, Czech Republic, India, Israel, Italy, Russia, South Africa and Ukraine). One study was conducted in Turkey (Cetingoz 2011).

The timing of cervical length measurement was different across the included studies; the inclusion criteria for one study (Fonseca 2007) were restricted to women who underwent routine ultrasonography at 20^{+0} to 25^{+0} weeks of gestation and cervical length of 15mm or less. In another study (Hassan 2011), women had ultrasonographic screening between 19^{+0} and 23^{+6} gestational weeks and were eligible to participate if cervical length was between 10 and 20mm.

The 2 other trials (O'Brien 2007; Cetingoz 2011) included women with cervical length ≤25 mm who had at least one previous history of spontaneous preterm delivery. In one of these (O'Brien 2007), the gestational age of included women ranged between 16^{+0} and 22^{+6} gestational weeks.

All included studies in this section reported the use of cerclage as an exclusion criterion. However, three of them reported that some women would have received rescue cerclage following randomisation but the proportion of women who received this intervention was not reported (Hassan 2011; O'Brien 2007; Cetingoz 2011 Turkey).

Subgroup analysis was also reported based on cervical length for the outcome of preterm birth of less than 33 weeks.

Women with history of preterm pre-labour rupture of membranes, history of mid-trimester loss, history of cervical trauma (including surgery), positive fetal fibronectin test

No studies were identified that looked at the effectiveness of prophylactic progesterone in these stratified risk groups.

Dosages used in included studies

Various doses of progesterone were administered across all included RCTs from 90 mg daily (O'Brien 2007); 100mg daily (Akbari 2009; Cetingoz 2011; da Fonseca 2003; Majhi 2009); 200mg daily (Rai 2009); to 400mg daily (Glover 2011) and to 200mg weekly (Rai 2009).

The use of progesterone commenced from, or prior to, 20 weeks' gestation in 2 trials (Glover 2011; O'Brien 2007), and continued up to a gestation of 28 weeks (da Fonseca 2003); 34 weeks (Akbari 2009; Cetingoz 2011; Glover 2011); 36 weeks of pregnancy (Rai 2009); and to 37 weeks (O'Brien 2007) gestation.

4.2.4 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The following GRADE profiles are presented:

- Table 16: GRADE profile for comparison of vaginal progesterone versus no treatment in women with a previous history spontaneous preterm birth
- Table 17: GRADE profile for comparison of vaginal progesterone versus placebo in women with a previous history spontaneous preterm birth
- Table 18: GRADE profile for comparison of oral progesterone versus placebo in women with a previous history spontaneous preterm birth
- Table 19: GRADE profile for comparison of vaginal progesterone versus placebo in women with ultrasound identified short cervix

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

Table 16: GRADE profile for comparison of vaginal progesterone versus no treatment in women with a previous history spontaneous preterm birth

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vaginal progesterone	No treatment	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal mortality											
1 study (Dodd 2013)	randomised trial	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	none	3/69 (4.3%)	10/72 (13.9%)	RR 0.31 (0.09 to 1.09)	96 fewer per 1000 (from 126 fewer to 13 more)	Low
Neonatal death											
1 study (Dodd 2013)	randomised trials	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	none	3/69 (4.3%)	10/72 (13.9%)	RR 0.31 (0.09 to 1.09)	96 fewer per 1000 (from 126 fewer to 13 more)	Low
Preterm birth less than 34 weeks											
1 meta-analysis of 2 studies (Dodd 2013)	randomised trials	Serious[1]	Serious[3]	No serious indirectness	Very serious[4]	none	4/119 (3.4%)	19/122 (15.6%)	RR 0.27 (0.05 to 1.38)	114 fewer per 1000 (from 148 fewer to 59 more)	Very low
Preterm birth less than 37 weeks											
1 meta-analysis of 2 studies (Dodd 2013)	randomised trials	Serious [1]	No serious inconsistency	No serious indirectness	No serious imprecision	none	14/119 (11.8%)	42/122 (34.4%)	RR 0.34 (0.2 to 0.59)	227 fewer per 1000 (from 141 fewer to 275 fewer)	Moderate
Neonatal sepsis											
1 meta-analysis of 2 studies (Dodd 2013)	randomised trials	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	none	0/119 (0%)	7/122 (5.7%)	RR 0.13 (0.02 to 1.01)	50 fewer per 1000 (from 56 fewer to [1] more)	Low

CI confidence interval, RR relative risk
1 Akbari 2009: Method of randomisation, allocation concealment, completeness of outcome data and blinding of outcome assessors were unclear
2 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default
3 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%)
4 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs

Table 17: GRADE profile for comparison of vaginal progesterone versus placebo in women with a previous history spontaneous preterm birth

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vaginal progesterone	Placebo	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal mortality											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	11/309 (3.6%)	11/302 (3.6%)	RR 0.98 (0.43 to 2.22)	1 fewer per 1000 (from 21 fewer to 44 more)	Low
Intrauterine fetal death											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	5/309 (1.6%)	4/302 (1.3%)	RR 1.22 (0.33 to 4.51)	3 more per 1000 (from 9 fewer to 46 more)	Low
Neonatal death											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	6/309 (1.9%)	7/302 (2.3%)	RR 0.84 (0.28 to 2.46)	4 fewer per 1000 (from 17 fewer to 34 more)	Low
Preterm birth less than 34 weeks											
1 meta-analysis of 2 studies (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	none	4/109 (3.7%)	22/104 (21.2%)	RR 0.17 (0.06 to 0.48)	176 fewer per 1000 (from 108 fewer to 199 fewer)	High
Preterm birth less than 37 weeks											
1 meta-analysis of 3 studies (Dodd 2013)	randomised trials	No serious risk of bias	Very serious[2]	No serious indirectness	Serious[3]	none	148/418 (35.4%)	160/406 (39.4%)	RR 0.67 (0.37 to 1.21)	130 fewer per 1000 (from 248 fewer to 83 more)	Very low

CI confidence interval, RR relative risk, NC not calculable

1 Evidence was downgraded by 2 due to very serious imprecision as 95%CI crossed 2 default MIDs

2 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and no plausible explanation was found with sensitivity or subgroup analysis

3 Evidence was downgraded by 1 due to serious imprecision as 95%CI crossed one default MID

Table 18: GRADE profile for comparison of oral progesterone versus placebo in women with a previous history spontaneous preterm birth

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oral progesterone	Placebo	Relative (95% CI)	Absolute (95% CI)	Quality
Adverse drug reaction											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	5/74 (6.8%)	7/74 (9.5%)	RR 0.71 (0.24 to 2.15)	27 fewer per 1000 (from 72 fewer to 109 more)	Low
Perinatal mortality											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	3/74 (4.1%)	7/74 (9.5%)	RR 0.43 (0.12 to 1.59)	54 fewer per 1000 (from 83 fewer to 56 more)	Low
Neonatal death											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	4/119 (3.4%)	19/122 (15.6%)	RR 0.27 (0.05 to 1.38)	114 fewer per 1000 (from 148 fewer to 59 more)	Low
Preterm birth less than 34 weeks											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Serious[2]	none	22/74 (29.7%)	37/74 (50%)	RR 0.59 (0.39 to 0.9)	205 fewer per 1000 (from 50 fewer to 305 fewer)	Moderate
Preterm birth less than 37 weeks											
Preterm birth less than 37 weeks (therapy started after 20 weeks)											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Serious2	none	5/19 (26.3%)	8/14 (57.1%)	RR 0.46 (0.19 to 1.11)	309 fewer per 1000 (from 463 fewer to 63 more)	Moderate
Pregnancy prolongation (weeks) (Better indicated by higher values)											
1 study (Dodd 2013)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	none	74	74	-	MD 4.47 higher (2.15 to 6.79 higher)	High

CI confidence interval, RR relative risk

1 Evidence was downgraded by 2 due to very serious imprecision as 95%CI crossed 2 default MIDs

2 Evidence was downgraded by 1 due to serious imprecision as 95%CI crossed one default MID

Table 19: GRADE profile for comparison of vaginal progesterone versus placebo in women with ultrasound identified short cervix

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vaginal progesterone	Placebo	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal death											
1 IPD meta-analysis of 4 studies (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	12/365 (3.0%)	18/358 (5.0 %)	RR 0.64 (0.31 to 1.31)	5 fewer per 1000 (from 10 fewer to 4 more)	Low
Intrauterine fetal death											
1 IPD meta-analysis of 4 studies (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	6/365 (1.6%)	7/358 (1.9%)	RR 0.82 (0.28 to 2.42)	3 fewer per 1000 (from 10 fewer to 20 more)	Low
Neonatal death											
1 IPD meta-analysis of 2 studies (Dodd 2012)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	6/365 (1.6%)	11/358 (3.0 %)	RR 0.53 (0.20 to 1.39)	7 fewer per 1000 (from 11 fewer to 5 more)	Low
Preterm birth < 28 weeks											
1 IPD meta-analysis of 4 studies (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	Serious[3]	Serious[2]	none	20/365 (5.5%)	39/358 (10.8%)	RR 0.51 (0.31 to 0.85)	7 fewer per 1000 (from 2 fewer to 10 fewer)	Low
Preterm birth < 33 weeks											
1 IPD meta-analysis of 1 study (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	Serious[3]	Serious[2]	none	44/365 (12.0%)	72/358 (20.1%)	RR 0.56 (0.40 to 0.80)	6 fewer per 1000 (from 6 fewer to 8 fewer)	Low
Cervical length < 10 mm (subgroup analysis)											
1 IPD meta-analysis of 1 study (Romero 2012)	randomised trials	Serious[4]	No serious inconsistency	serious[3]	Very serious[1]	none	NR	NR	RR 0.83 (0.49 to 1.41)	2 fewer per 1000 (from 7 fewer to 6 more)	Very low
Cervical length 10-20 mm (subgroup analysis)											
1 IPD meta-analysis of 1 study (Romero 2012)	randomised trials	Serious[4]	No serious inconsistency	No serious indirectness	Serious[2]	none	NR	NR	RR 0.52 (0.35 to 0.67)	7 fewer per 1000 (from 5 fewer to 9 fewer)	Low
Cervical length 21-25 mm (subgroup analysis)											

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vaginal progesterone	Placebo	Relative (95% CI)	Absolute (95% CI)	Quality
1 study (Romero 2012)	randomised trials	Serious[3]	No serious inconsistency	No serious indirectness	Very serious[1]	none	NR	NR	RR 0.50 (0.10 to 2.41)	7 fewer per 1000 (from 13 fewer to 20 more)	Very low
Preterm birth < 35 weeks											
1 IPD meta-analysis of 4 studies (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	Serious[3]	Serious[2]	none	67/365 (18.3%)	100/358 (27.9%)	RR 0.67 (0.51 to 0.87)	5 fewer per 1000 (from 2 fewer to 7 fewer)	Low
Preterm birth < 37 weeks											
1 IPD meta-analysis of 4 studies (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	Serious[3]	No serious imprecision	none	127/365 (34.8%)	141/358 (39.4%)	RR 0.91 (0.75 to 1.10)	1 fewer per 1000 (from 3 fewer to 1 more)	Moderate
Neonatal sepsis											
1 IPD meta-analysis of 4 studies (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	11/365 (3.0 %)	14/358 (3.9 %)	RR 0.80 (0.37 to 1.74)	3 fewer per 1000 (from 9 fewer to 10 more)	Low
Bronchopulmonary dysplasia											
1 IPD meta-analysis of 2 studies (Romero 2012)	randomised trials	No serious risk of bias	No serious inconsistency	No serious indirectness	Very serious[1]	none	5/249 (2.0%)	14/231 (6.0%)	RR 0.76 (0.21 to 2.79)	3 fewer per 1000 (from 11 fewer to 25 more)	Low

CI confidence interval, RR relative risk, NC not calculable, NR not reported
1 Evidence was downgraded by 2 due to very serious imprecision as 95%CI crossed 2 default MIDs
2 Evidence was downgraded by 1 due to serious imprecision as 95%CI crossed one default MID
3 One study included a twins pregnancies
4 Subgroup analysis

4.2.5 Evidence statements

4.2.5.1 Women with a history of spontaneous preterm birth

Low quality evidence from a single RCT with a hundred and forty women found that although there was a lower risk of perinatal mortality and neonatal death for those who had received prophylactic vaginal progesterone compared to no treatment groups this difference was not significant.

Two RCTs of almost 250 women showed that the risk of preterm birth before 34 weeks in pregnancy was not significantly different between those women in vaginal progesterone and no treatment groups. This evidence was of very low quality.

Moderate quality evidence from 2 RCTs of almost 250 women showed a significantly lower risk of giving preterm birth less than 37 weeks for women who had received vaginal progesterone as a prophylactic measure compared to women who received no treatment. This evidence also showed a similar direction of protective effect for the vaginal progesterone group which was found for the outcome of neonatal sepsis compared to no treatment group, although this effect did not reach significance marginally (low quality evidence).

Low quality evidence from one study of over 600 women which compared the role of vaginal progesterone compared to placebo to prevent preterm delivery found no significant difference in any of the neonatal outcomes (perinatal mortality, intrauterine fetal death, neonatal death and preterm birth less than 37 weeks). However, high quality evidence from 2 studies of over 200 women showed that there is a significantly lower risk of preterm births under 34 weeks in women who received vaginal progesterone compared to those who received placebo.

Moderate to high quality evidence from one RCT of one hundred and fifty women showed that there is a significantly lower risk of preterm birth under 34 weeks and longer prolongation of pregnancy in weeks for women who received oral progesterone compared to those who received placebo. However, moderate and low quality evidence from individual studies (with sample sizes between 25 to over 200 women) found no significant differences between the treatment groups for adverse drug reactions, perinatal mortality, neonatal death and preterm birth under 37 weeks.

Women with ultrasound identified short cervix

There was moderate quality evidence from an IPD meta-analysis of 4 RCTs with a total sample size of over 700 women with short cervix that showed that women who received vaginal progesterone have significantly fewer preterm births under 28, 33 and, 35 weeks compared to those who received placebo. There was no strong indication that some subgroups based on the cervical length are benefited more or less by the intervention of progesterone. However, there was some indication that there may be a significantly lower risk of preterm birth in women with cervical length of 10-20mm who received progesterone compared with those who received placebo. There was no evidence of benefit for the subgroups of <10mm or between 21 and 25mm. This IPD meta-analysis also showed no significant differences between the 2 treatment groups for perinatal mortality, intrauterine fetal death, neonatal death, preterm birth under 37 weeks, bronchopulmonary dysplasia or neonatal sepsis (moderate to low quality evidence from 2 to 4 RCTs with a couple of hundred women involved).

Women with a history of preterm pre-labour rupture of membranes, a history of mid-trimester loss, a history of cervical trauma (including surgery) and women with a positive fetal fibronectin test in the current pregnancy

No studies identified.

There were no data available for the outcomes of maternal mortality or emotional/psychological impact nor for the neonatal outcomes of congenital abnormalities or neurodevelopmental disability for any of the groups of women at risk of preterm delivery.

4.2.6 Health economics profile

This was not prioritised for economic analysis as it was not thought likely that recommendations would have a high cost impact as progesterone is cheap.

A search was undertaken for health economic evidence on prophylactic progesterone to prevent preterm labour in women considered to be at risk of preterm labour and birth. A total of 149 studies were identified by the search. After reviewing titles and abstracts, 5 papers were obtained. Two of these studies were included in the literature review (Cahill 2010; Pizzi 2014) and are reported in more detail in Appendices. Both studies suggested that vaginal progesterone could be considered cost-effective relative to no treatment although the reporting on one of the studies (Cahill 2010) was of particularly poor quality.

4.2.7 Evidence to recommendations

4.2.7.1 Relative value placed on the outcomes considered

The Guideline Committee decided that all the outcomes (both neonatal and maternal) set out in the protocol were important for this review question. However, no evidence was found for either the neonatal outcomes of congenital abnormality or neurodevelopmental disability or the maternal outcomes including maternal mortality, side-effects or emotional/psychological impact for any of the groups of women at risk of preterm delivery.

The outcome of preterm delivery or the ability to prolong pregnancy was considered by the Committee as a proxy for neonatal benefit when comparing prophylactic progesterone with no treatment groups. However it was discussed that this beneficial effect can be compromised if there is an associated maternal infection given the known adverse effect of intrauterine infection on neonatal outcomes. It was thought that the harms consequent on exposure to an adverse in utero environment could ‘off-set’ the benefit of delaying the birth.

4.2.7.2 Consideration of clinical benefits and harms

The synthesis of evidence showed that significantly fewer women with a history of spontaneous preterm delivery who had received vaginal progesterone as a prophylactic measure delivered before 37 weeks of gestation compared to women who received no treatment. There was also some protective effect on the outcome of neonatal sepsis seen for babies born by mothers received vaginal progesterone compared to those who did not receive any prophylactic treatment. For the same group of women (with a history of spontaneous preterm delivery), oral progesterone was found to significantly reduce the number of preterm births before 34 weeks and to prolong the duration of pregnancy when compared to placebo. The lowest gestational age of included women in the reviewed studies was 16 weeks, therefore the Committee decided to adopt this gestational age as the lowest threshold for the recommendation. For women identified with short cervix on ultrasound scan, there was evidence of benefit on reducing preterm births under 28, 33, and 35 weeks with vaginal progesterone treatment when compared to those who received placebo. However, vaginal progesterone did not provide a significant benefit for the other neonatal

outcomes (perinatal mortality, intrauterine fetal death, neonatal death, preterm birth under 37 weeks, bronchopulmonary dysplasia or neonatal sepsis).

The Committee concluded that vaginal and oral progesterone are beneficial prophylactic measures for those women at risk for preterm birth, and noted that the evidence was limited for the oral compared to vaginal route of administration of progesterone.

No evidence was found on neonatal adverse events or long-term outcomes. However, the Committee assumed that any delay in preterm delivery tends to bring better long-term outcomes for the babies unless there is an unexpected and as yet unknown short or long-term side-effect profile specific to the intervention, or unless in utero infection is present, which can be harmful to the baby if exposure is prolonged.

Lastly, the Committee highlighted the importance of discussing the benefits and risk of prophylactic progesterone to women and taking into account their preferences.

4.2.7.3 Consideration of health benefits and resource uses

The cost of progesterone is cheap and therefore would not have a major cost impact on the NHS although there may be additional costs in identifying the women who are most likely to benefit from treatment, which is women with a short cervix. However, there is evidence that prophylactic progesterone can play a role in delaying preterm in such women and this could plausibly generate considerable savings by averting some of the expensive adverse outcomes of prematurity.

4.2.7.4 Quality of evidence

The majority of evidence for women with a previous history of preterm birth was of moderate to low quality. The main domain of quality assessment affected was imprecision and that was influenced by the small sample size of studies which reported some of the outcomes. Although the included studies have used different inclusion criteria on women's selection and used different dosages of progesterone and different protocols for treatment when women went into labour, there was no inconsistency on the direction of effects. One of the main differences in the baseline characteristics of the studies included in this review is that in the O'Brien study all women had a previous history of preterm birth in immediately previous pregnancy. Drug dosages and methods of administration did vary but not in a way that makes a logical sense as explanation for the difference in outcomes.

Evidence on women with documented short cervix (through an ultrasound) came from an IPD meta-analysis. This type of meta-analysis is considered a gold standard method and synthesis of evidence as it allows the use of previously unreported data, improves assessment of the primary study quality, standardization of outcome measures, undertaking an intent-to treat analysis, and use of optimal analytical methods. Subgroup and multivariable analyses would not have been possible without the availability of IPD. Studies that contributed to the analysis of short cervix in current pregnancy were all of high quality with low risk of bias. The Committee noted that meta-analysis of data at this section performed by IPD which gave more precision and credibility to the interpretation of the result.

4.2.7.5 Other considerations

The Committee also discussed the impact of timing of cervical length measurement on neonatal and maternal outcomes and the role of intravenous progesterone as an alternative option although no randomised evidence was found for this type of progesterone

The Committee was aware of SR and RCT evidence on 17 hydroxyprogesterone caproate. Although this treatment not a licensed product in the UK and therefore not included in the reviews, the benefits are thought to be similar to those of progesterone, supporting the Committee's recommendations.

The Committee also considered the evidence on prophylactic cerclage when drafting recommendations for this section and discussed the overlap of these interventions in delaying preterm delivery.

4.2.8 Key conclusions

The Committee concluded that for women with documented short cervix, progesterone can play a prophylactic role on the delay of preterm delivery for different gestations and discussed that more research is needed to explore the long term neonatal effects of this intervention.

The recommendations are based on both clinical evidence research and the committee's expert opinion.

4.2.9 Recommendations

The recommendations for prophylactic progesterone are in section 4.4.

4.3 Prophylactic cervical cerclage

4.3.1 Introduction

Cervical cerclage, also known as a cervical stitch, is a treatment used to prevent the cervix opening too early during a pregnancy causing either a late miscarriage or preterm birth. Indications for prophylactic cerclage include a history that increases the risk of spontaneous second-trimester loss or preterm delivery, and/or cervical shortening seen on ultrasound. Preventing or delaying preterm birth may have significant benefit in terms of reducing the severity of diseases of prematurity and associated complications. However, there is uncertainty about which women are most likely to benefit from this intervention.

4.3.2 Review question

What is the clinical effectiveness of prophylactic cervical cerclage in preventing preterm labour in women considered to be at risk of preterm labour and birth?

The following groups were considered at risk of preterm labour and birth:

- a history of spontaneous preterm birth
- a history of preterm pre-labour rupture of membranes
- a history of mid-trimester loss
- a history of cervical trauma (including surgery)
- a short cervix that has been identified on scan

4.3.3 Description of included studies

Two SR and meta-analyses (Alfirevic 2012, Berghella 2011) were included to answer this review question, one of which is an IPD meta-analysis (Berghella 2011) and the other a Cochrane SR and meta-analysis (Alfire 2012).

The population included in both SRs was women at risk of preterm labour namely those with previous spontaneous preterm birth or midtrimester loss, singleton gestation and short cervix (less than 25 mm).

Only RCTs were included in both SRs; IPD meta-analysis (Berghella 2011) of 5 RCTs (sample size of 504) looked only at the role of cerclage to prevent preterm birth and perinatal mortality and morbidity when compared to no treatment. The SR and meta-analysis of 12

RCTs by Alfirevic 2012 (total sample size of 3328) expanded its focus to include comparisons between prophylactic cerclage and alternatives (either no treatment or any alternative intervention e.g. progesterone), or included comparisons between different cerclage protocols (history, ultrasound, physical exam-indicated cerclage). Another difference between the two meta-analyses was that Alfirevic 2012 included RCTs (9) which compared cervical cerclage with no treatment for women at risk for preterm delivery as diagnosed with either previous history alone (4) or cervical length ultrasound examination (single or serial) (5) whereas Berghella 2011 restricted their inclusion criteria to those trials (5) including women who were assessed as high risk by having both a history and ultrasound testing. More specifically, inclusion criteria for the IPD meta-analysis was restricted to short cervical length (<25 mm in 4 trials and ≤15 mm in one trial) identified on either single or serial ultrasound scans before 24 weeks' gestation (4 trials) or before 27^{+0} weeks' gestation (one trial). All the studies reported by Berghella 2011 were also included in Alfirevic 2012.

For the 2 trials included in the meta-analysis by Alfirevic 2012 that compared a policy of history-indicated cerclage with a policy of ultrasound-indicated cerclage, all women had a history of spontaneous preterm birth (or midtrimester loss) suggesting a high risk of preterm birth in the current pregnancy. In both trials women allocated to the ultrasound arms had a cerclage when cervical length was found to be less than or equal to 20 mm (54% and 33% respectively). In one trial, all women in the history-only indicated arm had a cerclage, however in the second trial, for women randomised to the history-only indicated cerclage arm, cerclage was performed only if the obstetrician considered it necessary (20% cerclage).

Multiple pregnancies were excluded or results were presented separately from single pregnancies in both SRs.

The included RCTs came from a variety of locations (USA [4 trials], UK [1 trial], The Netherlands [1 trial], Iran [1 trial], Nigeria [1 trial], France [1 trial], South Africa [1 trial], multicentre [2 trials]). IPD from some of the trials were used by the authors of the SR.

Although the gestational age of women in the included studies was well reported and ranged between 10^{+0} weeks to 33^{+6} weeks at study's recruitment, the gestational age when the cerclage was performed was not reported. There was a wide variation on the type of suture used across the included studies; in majority of them (8) the McDonald suture was used, whereas one study used a suture "similar" to the McDonald suture and another one included the Shirodkar suture. Two studies did not provide any information on the type of suture used.

4.3.4 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The following GRADE profiles are presented and subgroups analyses were performed by the different ways of assessment of high risk status for preterm delivery (whether only based on history taking or ultrasound (single or serial) or based on both methods).

- Table 20: GRADE profile for comparison of prophylactic cervical cerclage versus no cerclage
- Table 21: GRADE profile for comparison of prophylactic cervical cerclage versus progesterone (17OHP-C)
- Table 22: GRADE profile for comparison of policy of prophylactic history-indicated cerclage versus policy of cerclage indicated by serial ultrasound-scanning in women with a previous preterm birth

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

Serious neonatal morbidity was defined in the Cochrane review by Alfirevic 2012 as any of the following: respiratory distress syndrome, intraventricular haemorrhage, necrotising entercolitis or sepsis, mechanical ventilation, major adverse outcome before hospital discharge bronchopulmonary dysplasia, retinopathy of prematurity, positive fetal blood culture, or other life-threatening morbidity.

1 **Table 20: GRADE profile for comparison of prophylactic cervical cerclage versus no cerclage**

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal death (women considered at risk of preterm birth by any indication)											
1 meta-analysis of 8 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	100/1196 (8.4%)	128/1195 (10.7%)	RR 0.78 (0.61 to 1)	24 fewer per 1000 (from 42 fewer to 0 more)	Moderate
Perinatal death (subgroup analysis only for those women considered at high risk of preterm labour due to their previous history alone)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	None	62/770 (8.1%)	77/769 (10%)	RR 0.8 (0.58 to 1.1)	20 fewer per 1000 (from 42 fewer to 10 more)	Low
Perinatal death - (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by one-off ultrasound scan)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	None	2/26 (7.7%)	3/30 (10%)	RR 0.77 (0.14 to 4.25)	23 fewer per 1000 (from 86 fewer to 325 more)	Low
Perinatal death (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by serial ultrasound scan)											
1 meta-analysis of 4 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	serious[1]	None	24/253 (9.5%)	37/256 (14.5%)	RR 0.66 (0.41 to 1.06)	49 fewer per 1000 (from 85 fewer to 9 more)	Moderate
Perinatal death (subgroup analysis only for those women considered at high risk of preterm labour due to their history of previous preterm birth and identification of a short cervix in the current pregnancy by either one-off or serial ultrasound scan)											
1 meta-analysis of 5 studies (Berghella 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	22/250 (8.8%)	35/254 (13.8%)	RR 0.65 (0.40 to 1.07)	48 fewer per 1000 (from 83 fewer to 10 more)	Moderate
Perinatal death (subgroup analysis only for those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	very serious[3]	None	12/147 (8.2%)	11/140 (7.9%)	RR 1.01 (0.46 to 2.22)	1 more per 1000 (from 42 fewer to 96 more)	Low
Serious neonatal morbidity (all women considered at risk of preterm birth for any indication)											
1 meta-analysis of 4 studies	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	very serious[3]	none	39/407 (9.6%)	42/411 (10.2%)	RR 0.95 (0.63 to 1.43)	5 fewer per 1000	Low

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	Quality
(Alfirevic 2012)										(from 38 fewer to 44 more)	
Serious neonatal morbidity (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by one-off ultrasound scan)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	None	2/26 (7.7%)	3/30 (10%)	RR 0.77 (0.14 to 4.25)	23 fewer per 1000 (from 86 fewer to 325 more)	Low
Serious neonatal morbidity (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by serial ultrasound scan)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	very serious[3]	None	25/234 (10.7%)	30/241 (12.4%)	RR 0.84 (0.51 to 1.37)	20 fewer per 1000 (from 61 fewer to 46 more)	Low
Serious neonatal morbidity - (subgroup analysis only for those women considered at high risk of preterm labour due to their history of previous preterm birth and identification of a short cervix in the current pregnancy by either one-off or serial ultrasound scan)											
1 meta-analysis of 4 studies (Berghella 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	17/207 (8.2%)	28/196 (14.3%)	RR 0.60 (0.34 to 1.06)	57 fewer per 1000 (from 94 fewer to 9 more)	Moderate
Serious neonatal morbidity-(subgroup analysis only for those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	very serious[3]	None	12/147 (8.2%)	9/140 (6.4%)	RR 1.4 (0.61 to 3.23)	26 more per 1000 (from 25 fewer to 143 more)	Low
Preterm birth before 37^{+0} weeks –(all women considered at risk of preterm birth for any indication)											
1 meta-analysis of 9 studies (Alfirevic 2012)	randomised trials	serious[5]	serious[6]	no serious indirectness[7]	serious[1]	None	389/1464 (26.6%)	480/1434 (33.5%)	RR 0.8 (0.69 to 0.95)	67 fewer per 1000 (from 17 fewer to 104 fewer)	Very Low
Preterm birth before 37^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to their previous history alone)											
1 meta-analysis of 4 studies (Alfirevic 2012)	randomised trials	serious[8]	serious[6]	no serious indirectness[9]	serious[3]	None	215/1038 (20.7%)	249/1007 (24.7%)	RR 0.86 (0.59 to 1.27)	35 fewer per 1000 (from 101 fewer to 67 more)	Very Low
Preterm birth before 37^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix together in the current pregnancy by one-off ultrasound scan)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	9/26 (34.6%)	19/30 (63.3%)	RR 0.55 (0.3 to 0.99)	285 fewer per 1000	Moderate

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	Quality
										(from 443 fewer to 6 fewer)	
Preterm birth before 37^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by serial ultrasound scan)											
1 meta-analysis of 4 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	serious[6]	no serious indirectness[4]	serious[1]	None	110/253 (43.5%)	144/257 (56%)	RR 0.78 (0.6 to 1.02)	123 fewer per 1000 (from 224 fewer to 11 more)	Low
Preterm birth before 37^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to their history of previous preterm birth and identification of a short cervix in the current pregnancy by either one-off or serial ultrasound scan)											
1 meta-analysis of 5 studies (Berghella 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	105/250 (42%)	154/254 (60.6%)	RR 0.70 (0.58 to 0.83)	182 fewer per 1000 (from 103 fewer to 255 fewer)	Moderate
Preterm birth before 37^{+0} weeks - (subgroup analysis only for those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	serious6	no serious indirectness[4]	serious[1]	None	55/147 (37.4%)	68/140 (48.6%)	RR 0.8 (0.55 to 1.16)	97 fewer per 1000 (from 219 fewer to 78 more)	Low
Preterm birth before 35^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to their history of previous preterm birth and identification of a short cervix in the current pregnancy by either one-off or serial ultrasound scan)											
1 meta-analysis of 5 studies (Berghella 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	71/250 (28.4%)	105/254 (41.3%)	RR 0.70 (0.55 to 0.89)	124 fewer per 1000 (from 45 fewer to 186 fewer)	Moderate
Preterm birth before 34^{+0} weeks –(in all women considered at risk of preterm birth for any indication)											
1 meta-analysis of 8 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	serious[1]	None	210/1196(17.6%)	277/1196 (23.2%)	RR 0.79 (0.68 to 0.93)	49 fewer per 1000 (from 74 fewer to 16 fewer)	Moderate
Preterm birth before 34^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to their previous history alone)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	serious[2]	serious[6]	no serious indirectness	very serious[3]	None	106/770 (13.8%)	138/769 (17.9%)	RR 0.76 (0.4 to 1.46)	43 fewer per 1000 (from 108 fewer to 83 more)	Very Low
Preterm birth before 34^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by one-off ultrasound scan)											

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	Quality
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	None	6/26 (23.1%)	11/30 (36.7%)	RR 0.63 (0.27 to 1.46)	136 fewer per 1000 (from 268 fewer to 169 more)	Low
Preterm birth before 34^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by serial ultrasound scan)											
1 meta-analysis of 4 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	serious[1]	None	65/253 (25.7%)	90/257 (35%)	RR 0.77 (0.55 to 1.1)	81 fewer per 1000 (from 158 fewer to 35 more)	Moderate
Preterm birth before 34^{+0} weeks - (subgroup analysis only for those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	serious[1]	None	33/147 (22.4%)	38/140 (27.1%)	RR 0.82 (0.55 to 1.22)	49 fewer per 1000 (from 122 fewer to 60 more)	Moderate
Preterm birth before 32^{+0} weeks – subgroup analysis in women considered at high risk of preterm labour due to their history of previous preterm birth and identification of a short cervix in the current pregnancy by either one-off or serial ultrasound scan											
1 meta-analysis of 5 studies (Berghella 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	48/250 (19.2%)	75/254 (29.5%)	RR 0.66 (0.48 to 0.91)	100 fewer per 1000 (from 27 fewer to 154 fewer)	Moderate
Preterm birth before 28^{+0} weeks –(all women considered at risk of preterm birth for any indication)											
1 meta-analysis of 8 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	serious[1]	None	118/1196 (9.9%)	148/1196 (12.4%)	RR 0.8 (0.64 to 1)	25 fewer per 1000 (from 45 fewer to 0 more)	Moderate
Preterm birth before 28^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to their previous history alone)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	None	60/770 (7.8%)	73/769 (9.5%)	RR 0.82 (0.59 to 1.13)	17 fewer per 1000 (from 39 fewer to 12 more)	Low
Preterm birth before 28^{+0} weeks (subgroup analysis only for those considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by one-off ultrasound scan)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	None	3/26 (11.5%)	5/30 (16.7%)	RR 0.69 (0.18 to 2.62)	52 fewer per 1000 (from 137 fewer to 270 more)	Low

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	Quality
Preterm birth before 28^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history)and identification of a short cervix in the current pregnancy by serial ultrasound scan)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness[4]	serious[1]	None	36/253 (14.2%)	52/257 (20.2%)	RR 0.71 (0.48 to 1.04)	59 fewer per 1000 (from 105 fewer to 8 more)	Moderate
Preterm birth before 28^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to their history of previous preterm birth and identification of a short cervix in the current pregnancy by either one-off or serial ultrasound scan)											
1 meta-analysis of 5 studies (Berghella 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	32/250 (12.8%)	51/254 (20.1%)	RR 0.64 (0.43 to 0.96)	72 fewer per 1000 (from 8 fewer to 114 fewer)	Moderate
Preterm birth before 28^{+0} weeks - (subgroup analysis only for those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	None	19/147 (12.9%)	18/140 (12.9%)	RR 1.01 (0.55 to 1.83)	1 more per 1000 (from 58 fewer to 107 more)	Low
Preterm birth before 24^{+0} weeks - (subgroup analysis only for those women considered at high risk of preterm labour due to their history of previous preterm birth and identification of a short cervix in the current pregnancy by either one-off or serial ultrasound scan)											
1 meta-analysis of 5 studies (Berghella 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	None	13/250 (5.2%)	28/254 (11%)	RR 0.48 (0.26 to 0.90)	57 fewer per 1000 (from 11 fewer to 82 fewer)	Moderate
Maternal side effects (vaginal discharge, bleeding, pyrexia not requiring antibiotics) –(in all women considered at risk of preterm birth for any indication)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	serious[10]	serious[11]	no serious indirectness	serious[1]	None	83/491 (16.9%)	49/462 (10.6%)	RR 2.25 (0.89 to 5.69)	133 more per 1000 (from 12 fewer to 497 more)	Very Low
Maternal side effects (vaginal discharge, bleeding, pyrexia not requiring antibiotics) - (subgroup analysis only for those women considered at high risk of preterm labour due to their previous history alone)											
1 meta-analysis of 2 studies (Alfirevic 2012)	randomised trials	serious[10]	serious[11]	serious[9]	serious[1]	None	71/364 (19.5%)	47/336 (14%)	RR 1.57 (0.76 to 3.24)	80 more per 1000 (from 34 fewer to 313 more)	Very Low
Maternal side effects (vaginal discharge, bleeding, pyrexia not requiring antibiotics) - (subgroup analysis only for those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	12/127 (9.4%)	2/126 (1.6%)	RR 5.95 (1.36 to 26.06)	79 more per 1000	High

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	Quality
										(from 6 more to 398 more)	
Pyrexia (fever of 38°C or more) –in all women considered at risk of preterm birth (any indication)											
1 meta-analysis of 3 studies (Alfirevic 2012)	randomised trials	serious[12]	no serious inconsistency	no serious indirectness	no serious imprecision	None	38/630 (6%)	15/615 (2.4%)	RR 2.39 (1.35 to 4.23)	34 more per 1000 (from 9 more to 79 more)	Moderate
Pyrexia (fever of 38°C or more) - (subgroup analysis only for those women considered at high risk of preterm labour due to their previous history alone)											
1 meta-analysis of 2 studies (Alfirevic 2012)	randomised trials	serious[12]	no serious inconsistency	no serious indirectness	serious[1]	None	34/503 (6.8%)	15/489 (3.1%)	RR 2.22 (1.22 to 4.01)	37 more per 1000 (from 7 more to 92 more)	Low
Pyrexia (fever of 38°C or more)- (subgroup analysis only for those women considered at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by one-off ultrasound scan)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	None	1/26 (3.8%)	0/30 (0%)	RR 3.44 (0.15 to 81.09)	NC	Low
Pyrexia (fever of 38°C or more) - (subgroup analysis only for those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	None	3/101 (3%)	0/96 (0%)	RR 6.66 (0.35 to 127.2)	NC	Low

1 *CI confidence interval, RR relative risk, NC not calculable, PTL preterm labour*
2 *1 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID2 Method of randomisation not clearly reported in 3 trials;*
3 *allocation concealment not clearly reported in 2 trials*
4 *3 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs*
5 *4 One study did not exclude women with advanced cervical dilatation or exposed fetal membranes (numbers not reported). 7% of women in the control arm of that study*
6 *received cerclage and it is not clear whether intention-to-treat analysis was performed*
7 *5 Method of randomisation not clearly reported in 6/9 trials; allocation concealment not clearly reported in 4/9 trials*
8 *6 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and no plausible explanation was found*
9 *with subgroup analysis*
10 *7 In 2 studies women in the control arm received cerclage and it is unclear whether intention-to-treat analysis performed in those trials*
11 *8 Method of randomisation not clearly reported in 4 trials; allocation concealment not clearly reported in 3 trials*
12 *9 11% of women in the control arm of one study received cerclage and it is unclear whether intention-to-treat analysis was performed*
13 *10 Unclear method of randomisation and allocation concealment in 2 trials*
14 *11 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and no plausible explanation was found*
15 *with subgroup analysis*
16 *12 Method of randomisation not clearly reported in 2 trials; allocation concealment not clearly reported in 1 trial*

Table 21: GRADE profile for comparison of prophylactic cervical cerclage versus progesterone (17OHP-C)

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cerclage	Progesterone	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal death											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	None	14/42 (33.3%)	11/37 (29.7%)	RR 1.12 (0.58 to 2.16)	36 more per 1000 (from 125 fewer to 345 more)	Low
Serious neonatal morbidity ("respiratory distress syndrome requiring mechanical ventilation > 24h, intraventricular haemorrhage, neonatal sepsis or necrotising enterocolitis)											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	None	9/42 (21.4%)	7/37 (18.9%)	RR 1.13 (0.47 to 2.74)	25 more per 1000 (from 100 fewer to 329 more)	Low
Preterm birth before 37 completed weeks											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	None	22/42 (52.4%)	22/37 (59.5%)	RR 0.88 (0.6 to 1.3)	71 fewer per 1000 (from 238 fewer to 178 more)	Low
Preterm birth before 28 completed weeks											
1 study (Alfirevic 2012)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	None	10/42 (23.8%)	7/37 (18.9%)	RR 1.26 (0.53 to 2.97)	49 more per 1000 (from 89 fewer to 373 more)	Low

17-OHP-C, 17 α-hydroxyprogesterone caproate, CI confidence interval, RR relative risk

1 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs

Table 22: GRADE profile for comparison of policy of prophylactic history-indicated cerclage versus policy of cerclage indicated by serial ultrasound-scanning in women with a previous preterm birth

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	History-indicated cerclage	Serial scanning	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal death											
1 study (Alfirevic 2012)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	None	14/125 (11.2%)	10/122 (8.2%)	RR 1.37 0.63 to 2.96)	30 more per 1000 (from 30 fewer to 161 more)	Very Low
Serious neonatal morbidity (Composite measure of morbidity not adequately described)											
1 study (Alfirevic 2012)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	None	7/125 (5.6%)	4/122 (3.3%)	RR 1.71 (0.51 to 5.69)	23 more per 1000 (from 16 fewer to 154 more)	Very Low
Preterm birth before 37 completed weeks											

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	History-indicated cerclage	Serial scanning	Relative (95% CI)	Absolute (95% CI)	Quality
1 study (Alfirevic 2012)	randomised trials	serious[3]	no serious inconsistency	serious4	very serious[2]	None	5/45 (11.1%)	8/52 (15.4%)	RR 0.72 (0.25 to 2.05)	43 fewer per 1000 (from 115 fewer to 162 more)	Very Low
Preterm birth before 34 completed weeks											
1 study (Alfirevic 2012)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	None	19/125 (15.2%)	18/122 (14.8%)	RR 1.03 (0.57 to 1.87)	4 more per 1000 (from 63 fewer to 128 more)	Very Low
Maternal infection requiring intervention (antibiotics or delivery)											
1 study (Alfirevic 2012)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	None	0/125 (0%)	1/122 (0.82%)	RR 0.33 (0.01 to 7.91)	5 fewer per 1000 (from 8 fewer to 57 more)	Very Low
Maternal side effects (vaginal discharge, bleeding, pyrexia not requiring antibiotics)											
1 study (Alfirevic 2012)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	None	6/122 (4.9%)	11/121 (9.1%)	RR 0.54 (0.21 to 1.42)	42 fewer per 1000 (from 72 fewer to 38 more)	Very Low

1 *CI confidence interval, RR relative risk*
2 *1 Women in the serial scanning group received significantly more progesterone than women in the history-indicated cerclage group (39% vs. 25%)*
3 *2 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs 3 Unclear method of randomisation and allocation*
4 *concealment*
5 *4 54% of women in the control arm received cerclage and it is unclear whether intention-to-treat analysis performed*

6

7

8

4.3.5 Evidence statements

4.3.5.1 Prophylactic cerclage compared with no cerclage

Moderate to very low quality evidence from 2 meta-analyses (one was IPD) of 500 to over 3000 women found that, although fewer perinatal deaths and preterm births and neonatal morbidity occurred in women who had cerclage compared with women who did not have cerclage, this difference was only significant for the outcomes of preterm birth before 37^{+0} weeks, 34^{+0} weeks and 28^{+0} weeks. Further sub analysis based on the way the high risk status for preterm birth was assessed, showed that this significant difference only remained for the subgroup of women assessed by the combination of risk due to previous history (preterm labour or midtrimester loss) and short cervix by serial or one-off ultrasound testing.

High to very low quality evidence from the aggregated meta-analysis (from several hundred women) showed that although a higher proportion of women in cerclage group experienced maternal adverse events and pyrexia compared to no- treatment group, this difference was significant only for the outcome of pyrexia when all women who were considered at risk of preterm birth were included (independently of indication). Subgroup analyses of high quality evidence showed that a significantly higher proportion of women who were considered at low or uncertain risk of preterm delivery based on their previous history but with a short cervix in current pregnancy identified by one-off ultrasound scan experienced maternal adverse events in cerclage group compared to no treatment group.

Further subgroup analyses for the outcome of pyrexia showed that the difference in the proportion of women in the cerclage arm who experienced pyrexia compared to the no-treatment arm remained significantly high only for those women who were considered at high risk based on their previous history alone. There was no significant difference in risk of pyrexia between subgroups classified as at high risk of preterm labour due to both their previous history and identification of a short cervix in the current pregnancy by one-off ultrasound scan or those women considered at low or unspecified risk of preterm labour due to their previous history but with a short cervix in the current pregnancy identified by one-off ultrasound scan (low quality evidence)

4.3.5.2 Prophylactic cerclage compared with prophylactic progesterone

Low quality evidence from the single trial (of 79 women) which reported this comparison showed no significant differences in the outcomes of perinatal deaths, serious neonatal morbidity or preterm births before 37^{+0} weeks or before 28^{+0} weeks between women who received prophylactic cerclage and women who received prophylactic progesterone.

4.3.5.3 Policy of prophylactic history-indicated cerclage compared with policy of cerclage indicated by serial ultrasound scanning

Very low quality evidence from a single trial (over 200) comparing a policy of prophylactic history-indicated cerclage compared to a policy indicated by serial ultrasound scanning found no significant difference in perinatal deaths, serious neonatal morbidity, preterm births before 37^{+0} weeks or 34^{+0} weeks, maternal infection requiring intervention and maternal side effects (vaginal discharge, bleeding, and pyrexia not requiring antibiotics), although only 20% of women in the “history indicated” group were actually treated with cerclage. Results should be interpreted with caution as the obstetrician in this trial decided that women in the group of policy of prophylactic history-indicated cerclage did not need one (there were no strict criteria for this decision) when 100% should have had it in this group if it were truly a RCT comparing prophylactic history-indicated cerclage

4.3.6 Health economics profile

A single search was undertaken for health economic evidence on prophylactic cervical cerclage to prevent preterm labour in women considered to be at risk of preterm labour and birth and rescue cervical cerclage in preventing preterm birth in women in suspected preterm labour. A total of 60 articles were identified by the search. After reviewing titles and abstracts, 3 papers were obtained. These studies were all excluded because they were not economic evaluations or were published conference abstracts. Therefore, no relevant economic evidence was identified for this question.

This question was identified as a priority for health economic analysis as current practice is varied and there is a lack of consensus on many aspects of care. Therefore the Committee thought it would be important for recommendations to be supported by cost effectiveness evidence especially as there are potentially large cost savings from preventing preterm birth. However, no new economic analysis was undertaken due to a lack of evidence of difference on "functional outcomes" (such as perintal death and serious neonatal morbidity) in the clinical review, with most studies not powered to detect any differences, making it difficult to assess treatment effectiveness.

4.3.7 Evidence to recommendations

4.3.7.1 Relative value placed on the outcomes considered

The Committee considered that reductions in serious neonatal outcomes and longer term morbidity were the most important outcomes for this review question. They considered a reduction in incidence of preterm birth to be a useful proxy measure for these outcomes so this was prioritised as well among the other neonatal outcomes. The Committee agreed that it would be also informative to report any long-term infant neurodevelopmental outcomes or neurodevelopmental disability as a single outcome for comparing the effectiveness of prophylactic cerclage with the other options.

In terms of maternal outcomes, the Committee prioritised mortality and adverse effects, including maternal infection requiring further intervention and cervical trauma that can require future repair because prophylactic cerclage is an invasive procedure. The Committee also discussed the importance of the emotional or psychological impact on women undergoing this type of prophylactic intervention.

The Committee set out at the protocol stage that outcomes will be assessed according to the way the risk of preterm labour is assessed in the studies; whether there is a previous experience of preterm delivery (or mid trimester loss) or there are risk factors associated with current pregnancy such as a short cervix.

4.3.7.2 Consideration of clinical benefits and harms

The evidence from the 2 SRs and meta-analyses (one aggregated and one IPD) showed that among the neonatal outcomes reported in the studies (perinatal death, serious neonatal morbidity, preterm delivery), only delivering preterm before 37^{+0} , 34^{+0} and 28^{+0} weeks was significantly different between the cerclage and no treatment groups, with the cerclage group favoured for this outcome.

The evidence base also suggested that there may be an increase in maternal adverse effects in women who received prophylactic cerclage compared to those who did not. The Committee did however note that it was not possible to distinguish the nature of the individual adverse effect and thus it was hard to determine the clinical significance of this result. However, they discussed in depth the associated risks for the pregnancy from this technique such as uterine contractions, bleeding or infection which may lead to miscarriage

or preterm labour. These risks were balanced against the benefit from mechanical support to the cervix.

The only available data on specific adverse events was for pyrexia which was analysed separately. The results did show a significant increase in the risk of experiencing pyrexia in the group that received prophylactic cerclage compared to no treatment. However, there was still some uncertainty as to the clinical significance of this result given that none of the trials specified whether the women who had pyrexia had also received antibiotics.

The IPD meta-analysis reported outcomes specific to women with a history of previous preterm birth and a short cervix in the current pregnancy identified by ultrasound scan. No further evidence was identified that provided information about women with other historical indications, for example a history of cervical trauma (including surgery). However, enough information was available to perform subgroup analysis between those women assessed as at high risk for preterm labour only from history taking or from investigating the cervical length (with serial or one-off ultrasound testing). Sub-group analyses were performed to look at the outcomes according to the different risk factors for preterm birth that had been used as indicators for the use of prophylactic cerclage in the trials. Analyses were conducted for sub-groups of women who had been identified as being at high risk due to their history alone, or at high risk due to their history and the presence of a short cervix in the current pregnancy identified by one-off or serial ultrasound scan. No differences were found between groups for the outcomes of perinatal death or serious neonatal morbidity. The Committee felt the contrast between the findings for perinatal death in these sub-groups and the findings in the overall analysis could potentially be attributed to the smaller sample sizes included in the sub-groups and were therefore reluctant to draw any firm conclusions from this.

For women with a history of previous preterm birth who were also found to have a short cervix on either a single ultrasound scan or serial ultrasound scans, there was evidence of a significant reduction in preterm birth before $37^{+0,}$ 35^{+0}, 32^{+0}, 28^{+0} and 24^{+0} weeks gestation for women who had received prophylactic cerclage compared to those who did not. This conclusion was in line with the Committee's clinical experience.

Prophylactic cerclage was also compared against progesterone and no difference in rates of preterm birth was found between the 2 interventions for any of the neonatal outcomes. Therefore a recommendation with a choice of either of these 2 prophylactic interventions was drafted.

The Committee also noted that in women with a previous preterm birth, the comparison of a policy of prophylactic cerclage on the basis of clinical history versus a policy of cerclage indicated by serial ultrasound-scanning was not very informative because the estimates of effects between the 2 groups were biased by the design limitations of the trial. For these reasons, the Committee did not place confidence in these results.

In summary, the results of the review were that the benefits of prophylactic cerclage, in terms of reduction in preterm birth, were more likely to be seen in the sub-group of women who had had both a previous preterm birth and a short cervix in the current pregnancy. This reflected the Committee's clinical experience. Moreover, as women in the overall analysis also included a proportion of women with this particular combination of risk factors, they felt it was plausible that the benefits seen in the overarching group were likely to be due to the influence of these women on the overall result. They noted that there was a paucity of evidence about the emotional and psychological impact of prophylactic cerclage and transvaginal scanning. Hence they concluded that the recommendations should be tailored to a specific group of women for whom the benefit of this intervention is most certain.

4.3.7.3 Consideration of health benefits and resource uses

The Committee felt that prophylactic cerclage was likely to be an expensive intervention due to the setting in which it is delivered and required clinical expertise of the health-care professionals providing the care.

However they also acknowledged that the management of preterm birth and the associated neonatal outcomes are extremely costly and therefore considered that the overall health benefits likely to be obtained from offering this prophylactic intervention to selected women not only justified the resource use but that the initial costs incurred would be likely to be off-set by large cost savings downstream

4.3.7.4 Quality of evidence

The Committee noted that evidence was available for most of the prioritised outcomes with the exception of cervical trauma and women's emotional/psychological impact but that the quality varied among the outcomes.

The quality of the majority evidence was moderate to low with risk of bias (due to lack of information on study design) and imprecision to be the most affected areas of quality assessment.

The Committee also concluded that although results of subgroup analyses informed their decision making, most studies were not sufficiently powered to detect a significant difference in outcomes (so the evidence was downgraded for very serious imprecision for these outcomes) between the groups therefore effects on subgroups should be interpreted with caution. This does not apply for the subgroup analyses presented by Berghella 2011 IPD meta-analysis given the superiority of the quality of this analysis. IPD meta-analysis is considered a gold standard in meta-analysis given that uses the 'raw data' of individual patients from included studies instead of the published summary results of studies in a traditional meta-analysis. Compared to subgroup analyses in a single study or in a traditional meta-analysis, an IPD offers important potential advantages, such as: (1) increased possibilities to perform more complex statistical analyses that better match the underlying data; (2) more power compared to single studies and traditional meta-analyses; (3) higher validity of subgroup analyses by avoiding ecological bias and by taking the distribution of other patient characteristics into account; (4) improved flexibility and standardization of defining subgroups across studies; and (5) opportunities to examine the consistency of subgroup effects across studies

4.3.7.5 Other considerations

These recommendations were based on both the clinical interpretation of evidence and on Committee's expert opinion. Because individual studies included women whose history of previous preterm birth varied between 16 and 36 weeks, and cervical length < 25mm in the current pregnancy, they relied on the characteristics of women in the majority of studies in both meta-analyses to inform their recommendations.

4.3.8 Key conclusions

The Committee considered that whether or not cerclage can be said to be beneficial to improve neonatal outcomes depends on whether the reduction in preterm birth (for which there is evidence) is translated into a reduction in preterm birth-related neonatal morbidity (for which there is no evidence). There is evidence of limited harm to the mother in terms of increased pyrexia, but the clinical significance of this is not clear. The Committee concluded that the evidence of benefit is not so great to recommend that all women at risk of preterm birth due to a previous history and/or a short cervix should have prophylactic cerclage, but neither is evidence of harm so great to lead to justify recommending against its use.

From their clinical expertise the Committee felt that some sub-groups of women may benefit from prophylactic cerclage, but this cannot be fully supported by the evidence. It was felt that performing cerclage in all women thought to be at risk of preterm birth due to a previous history and/or a short cervix would result in over-treatment of women who would not necessarily benefit and may lead to iatrogenic harm. Insertion of a suture is an invasive procedure and the Committee's experience was that some women will choose not to have the intervention without clear benefit. They did however feel that the evidence of reduced preterm birth across a range of gestations in women who had a history of preterm birth plus a short cervix in the current pregnancy identified on ultrasound scan suggested that the balance of benefits and harms was most likely to be optimised in this group and therefore the recommendation should be targeted towards these women. However, women at higher risk of preterm labour such as those with a history of preterm prelabour rupture of membranes (P-PROM) in previous pregnancy or a history of cervical surgery should also be considered for this prophylactic treatment option. The Committee were disappointed with the availability of appropriate data as they were uncertain of the benefit for women with certain risk criteria such as cone biopsy/ large loop excision of the transformation zone

4.4 Recommendations

3. **Offer a choice of either prophylactic vaginal progesterone or prophylactic cervical cerclage to women:**
    - with a history of spontaneous preterm birth or mid-trimester loss between 16^{+0} and 34^{+0} weeks of pregnancy and
    - in whom a transvaginal ultrasound scan has been carried out between 16^{+0} and 24^{+0} weeks of pregnancy that reveals a cervical length of less than 25 mm.

   Discuss the benefits and risks of prophylactic progesterone and cervical cerclage with the woman and take her preferences into account.

4. **Offer prophylactic vaginal progesterone to women with no history of spontaneous preterm birth or mid-trimester loss in whom a transvaginal ultrasound scan has been carried out between 16^{+0} and 24^{+0} weeks of pregnancy that reveals a cervical length of less than 25 mm.**

5. **Consider prophylactic cervical cerclage for women in whom a transvaginal ultrasound scan has been carried out between 16^{+0} and 24^{+0} weeks of pregnancy that reveals a cervical length of less than 25 mm and who have either:**
    - had preterm prelabour rupture of membranes (P-PROM) in a previous pregnancy or
    - a history of cervical trauma.

1 ## 4.5 Research recommendations

2

Research question	1. What is the clinical effectiveness of prophylactic cervical cerclage alone compared with prophylactic vaginal progesterone alone and with both strategies together for preventing preterm birth in women with a short cervix and a history of spontaneous preterm birth?
Why this is needed	
Importance to 'patients' or the population	Preterm birth causes significant neonatal morbidity and mortality, as well as long-term disability. Therefore strategies for preventing preterm birth are important. There are recognised risk factors for preterm birth, and so interventions can be offered to women with these risk factors. Both prophylactic cervical cerclage and prophylactic vaginal progesterone are effective in preventing preterm birth in women with a short cervix and a history of preterm birth, but there is limited evidence on which is more effective, and the relative risks and benefits (including costs) of each. More randomised research is needed to compare the relative effectiveness of prophylactic cervical cerclage and prophylactic vaginal progesterone in improving both neonatal and maternal outcomes. This will help women and healthcare professionals to make an informed decision about which is the most effective prophylactic option.
Relevance to NICE guidance	The research would affect one of the key recommendations in future updates of the guideline and would therefore be highly relevant
Relevance to the NHS	Preterm birth causes significant neonatal morbidity and mortality, and survivors may have life-long physical and neurological disabilities. Thus any intervention that prevents preterm birth will reduce the requirement for health care and social care resources, with considerable financial savings.
National priorities	NHS Outcomes Framework #1: Preventing people from dying prematurely
Current evidence base	Both prophylactic cervical cerclage and prophylactic vaginal progesterone were found to be effective in reducing the risk of preterm birth in women with a short cervix and a history of previous preterm birth, but there is limited evidence on which is more effective, and the relative risks and benefits (including costs) of each. More research is needed to allow women and their caregivers to make an informed decision.
Equality	This group is defined only by gestational age at delivery.

5 Diagnosing preterm prelabour rupture of membranes (P-PROM)

5.1 Introduction

P-PROM is the presenting symptom in around 20% of all women who develop spontaneous preterm labour. Whilst many women with preterm rupture of the fetal membranes go into labour fairly quickly thereafter, those who do not are at risk of infection ascending into the uterine cavity. Such infection can be seriously harmful to mother and baby, hence a diagnosis of P-PROM warrants careful clinical monitoring, to facilitate early detection and treatment of in utero infection and chorioamnionitis. Hence accurate diagnosis of P-PROM is important. The aim of this question was to determine the diagnostic accuracy of placental alpha-microglobulin-1, nitrazine (pH), insulin-like growth factor binding protein-1, fetal fibronectin and diagnostic panty-liner with polymer-embedded strip to diagnose preterm pre-labour rupture of membranes (P-PROM). These indexes were considered either individually or in combination.

5.1.1 Review question

What is the diagnostic accuracy of the following tests to identify preterm pre-labour rupture of membranes:

- **Placental alpha-microglobulin-1**
- **Nitrazine (pH)**
- **Insulin-like growth factor binding protein-1**
- **Fetal fibronectin**
- **Panty-liner with polymer-embedded strip?**

5.1.2 Description of included studies

Two prospective cohort studies were included in this review (Jain & Morris, 1998; Tagore & Kwek, 2010) which evaluated the diagnostic accuracy of specific tests for the detection of preterm pre-labour rupture of membranes (P-PROM); One study investigated the diagnostic accuracy of insulin-like growth factor binding protein-1 (IGFBP-1) for diagnosing P-PROM, whereas the other tested the use of IGFBP-1, placental alpha-microglobulin-1 (PAMG-1) and the nitrazine test to diagnose P-PROM.

One study was conducted in the United Kingdom (Jain & Morris, 1998), and the other (Tagore & Kwek, 2010) in Singapore.

The gestations of women included in one study ranged from 17 to 37 weeks (Tagore & Kwek, 2010) and from 24 to 36 weeks (Jain & Morris, 1998) in the second study. The use of tocolytics was reported only in 1 study which specified that two-thirds of women in the study received steroids with tocolysis (Tagore & Kwek, 2010).

No studies were found that examined the diagnostic accuracy of fetal fibronectin or the diagnostic panty-liner with polymer-embedded strip for preterm pre-labour rupture of membranes.

The reference test (gold standard) varied between the 2 studies. In one study pooling of the liquor in the posterior fornix in speculum examination and intact amniotic sac at birth appeared to be a reference test, although no clear definition was provided. In the second study the standard positive reference test was defined as the presence of 3 or more of the following conditions: pooling of the clear fluid seen during speculum examination,

oligohydraminous identified on ultrasound scan, signs and symptoms of chorioamnionitis, and preterm birth within a week of presentation along with a convincing history of leaking liquor.

5.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Data is reported in a modified GRADE table for reporting summary results of diagnostic studies (Table 23) for the following tests.

- placental alpha-microglobulin-1
- insulin-like growth factor binding protein-1
- nitrazine

A description of the usefulness of the positive and negative likelihood ratio (following the thresholds set up in the Methods Chapter, section 2.2.4) is given along the summary statistics for each test.

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H

Table 23: GRADE profile for predictive accuracy of diagnostic tests for identifying preterm pre-labour rupture of membranes (PPROM)

Quality assessment							Number of women	Measure of diagnostic accuracy (with 95% confidence intervals)				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Placental alpha-microglobulin-1												
1 study (Tagore & Kwek, 2010)	case series	very serious[1,2]	no serious inconsistency	serious[3]	no serious imprecision	none	100	92.07%(84 to 100)	99% (98 to 100)	547 (1.11 to >1000) Very useful	0.07 (0.02 to 0.21) Very useful	Very low
Insulin-like growth factor binding protein-1												
1 study (Jain & Morris, 1998)	case series	very serious[1,2,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	none	34	97% (82 to 100)	99% (97 to 100)	293 (0.60 to >1000) Very useful	0.02 (0.001 to 11.1) Very useful	Very low
Insulin-like growth factor binding protein-1												
1 study (Tagore & Kwek, 2010)	case series	very serious[1,2]	no serious inconsistency	serious[3]	no serious imprecision	none	94	87.5% (77 to 97)	94.4%(88 to 100)	15.75 (5.21 to 47.5) Very useful	0.13 (0.05 to 0.30) Moderately useful	Very low
Nitrazine												
1 study (Tagore & Kwek, 2010)	case series	very serious[1,2]	no serious inconsistency	serious[3]	no serious imprecision	none	98	85% (73 to 96)*	39.7% (27 to 52)	1.40 (1.10 to 1.80) Not useful	0.37 (0.16 to 0.84) Moderately useful	Very low

1 Unclear if the reference standard results interpreted without knowledge of the results of the index test
2 Unclear how women were selected for the study (a consecutive or random sample)
3 n = 6 women had twin pregnancy
4 The very wide confidence interval is due to the way this is calculated for likelihood ratios where there are very few false results and does not represent uncertainty around the point estimate, therefore the study has not been downgraded for imprecision.
5 Unclear if the same reference test was used for all participants
6 Reference test/gold standard not clearly specified. Might have used following observations: Pooling of the liquor in the posterior fornix in speculum examination intact amniotic sac at birth

5.1.4 Evidence statements

Placenta alpha-microglobulin-1

One prospective cohort study (n=100) found that placenta alpha-microglobulin-1 is a useful test in diagnosing P-PROM. Positive and negative likelihood ratios were very useful. The evidence was of very low quality.

Insulin-like growth factor binding protein-1

Two prospective cohort studies (n=128) found that insulin-like growth factor binding protein-1 had a very useful and moderately useful positive and negative likelihood ratio for diagnosing P-PROM. The evidence was of very low quality.

Nitrazine

Evidence from one prospective cohort study (n=98) found a not useful positive likelihood ratio and moderately useful negative likelihood ratio for the nitrazine test in diagnosing P-PROM. The evidence was of very low quality.

5.1.5 Health economics profile

A search was undertaken for health economic evidence on diagnostic tests to identify preterm pre-labour rupture of membranes. A total of 82 articles were identified by the search. After reviewing titles and abstracts, 14 papers were obtained but these were all excluded. Therefore, no relevant economic evidence was identified for this question.

This question was identified as a medium priority for health economic analysis but more important priorities meant that new economic analysis for this guideline was not ultimately undertaken.

5.1.6 Evidence to recommendations

5.1.6.1 Relative value placed on the outcomes considered

The Committee agreed that diagnosis of P-PROM is key to successful management and improved perinatal outcomes for affected women. They considered identification of true positive and true negative cases to be equally important for clinical decisions regarding further treatment of women. Failure to identify those women with P-PROM correctly can result in the failure to implement helpful prophylactic measures. Conversely, failure to identify women without P-PROM correctly can result in delay in discharge from hospital or inappropriate intervention such as hospitalisation or induction of labour for elective preterm birth and inappropriate use of antibiotics.

5.1.6.2 Consideration of clinical benefits and harms

The Committee recognised that the included studies showed that 2 tests (Placenta alpha-microglobulin-1, Insulin-like growth factor binding protein-1) were useful for correct identification of P-PROM, although the very low quality of evidence reduced the Committee's confidence in the results. Based on positive and negative likelihood ratio and they agreed that these 2 tests appeared to be better than nitrazine testing, that test results can be trusted to identify women who truly do have P-PROM and that women who have P-PROM are unlikely to be missed. The Committee also agreed that nitrazine should not be used as a diagnostic test for P-PROM.

The Committee noted the potential clinical harm of 'not useful' positive likelihood ratio of nitrazine concluding that the test is not useful for identifying P-PROM (identification of many false positives). The Committee were concerned that high rate of false positives may be problematic because this can unnecessarily result in a cascade of intervention such as induction of labour, elective preterm birth and use of antibiotics. The Committee were aware of evidence from randomised trials showing that administration of antibiotics to women in preterm labour with intact membranes is associated with a significant increase in the risk of cerebral palsy during childhood.

The Committee discussed amniotic pooling and concluded that this was an obvious and confirmed sign of P-PROM. Therefore they recommended that no further test for the diagnosis of P-PROM be performed when pooling of amniotic fluid is observed and that an additional diagnostic test is only required when there is uncertainty about diagnosis of P-PROM.

5.1.6.3 Consideration of health benefits and resource uses

The Committee was convinced of clinical usefulness of insulin like growth factor binding protein 1 and placental alpha macroglobulin 1 testing but noted that the cost effectiveness remains to be tested. They also noted that there is a cost associated with ongoing surveillance if a test isn't performed, which can include in-patient care.

5.1.6.4 Quality of evidence

The quality of evidence was very low and was limited to results from 2 included studies. The Committee acknowledged that many studies had been excluded but agreed that this restrictive approach was necessary to ensure that any recommendations made were based on relevant study populations.

The Committee had concerns regarding bias because of the small sample size of the included studies, as reference standards varied between the studies and as no single strategy can be used as the reference (gold) standard for diagnosis of P-PROM.

They agreed that although some useful positive and negative likelihood ratios were demonstrated, it was difficult to have confidence in the findings given the poor quality of the evidence. Hence the Committee made a recommendation that clinicians should not use the tests alone to decide what care to offer the woman and made a research recommendation for further research to examine the impact of diagnosis on management and outcome at different gestational ages.

5.1.6.5 Other considerations

The Committee acknowledged that the importance of ascertaining whether membranes are ruptured is greater in the preterm context than term because of the higher risk of complications for both the baby and the woman if a diagnosis is missed. However, they believed that prophylactic antibiotics should not be offered if diagnostic testing for P-PROM was negative and in the absence of amniotic fluid pooling, but rather that the woman should be encouraged to return if any further symptoms suggestive of P-PROM or preterm labour arose.

These recommendations were based on both the clinical interpretation of evidence and on Committee's clinical expert opinion.

5.1.7 Recommendations

6. **In a woman reporting symptoms suggestive of preterm prelabour rupture of membranes (P-PROM), offer a speculum examination to look for pooling of amniotic fluid and:**
    - if pooling of amniotic fluid is observed, do not perform any diagnostic test but offer care consistent with the woman having P-PROM (see Chapters 6, 7 and 10)
    - if pooling of amniotic fluid is not observed, consider performing an insulin-like growth factor binding protein-1 test or placental alpha-microglobulin-1 test of vaginal fluid.

7. **If the results of the insulin-like growth factor binding protein-1 or placental alpha-microglobulin-1 test are positive, do not use the test results alone to decide what care to offer the woman, but also take into account her clinical condition, her medical and pregnancy history and gestational age, and either:**
    - offer care consistent with the woman having P-PROM (see Chapters 6, 7 and 10) or
    - re-evaluate the woman's diagnostic status at a later time point.

8. **If the results of the insulin-like growth factor binding protein-1 or placental alpha-microglobulin-1 test are negative and no amniotic fluid is observed:**
    - do not offer antenatal prophylactic antibiotics
    - explain to the woman that it is unlikely that she has P-PROM, but that she should return if she has any further symptoms suggestive of P-PROM or preterm labour.

9. **Do not use nitrazine to diagnose P-PROM.**

10. **Do not perform diagnostic tests for P-PROM if labour becomes established in a woman reporting symptoms suggestive of P-PROM.**

5.1.8 Research recommendations

Research question	**2. What is the diagnostic accuracy and utility of tests (placental alpha-microglobulin-1, insulin-like growth factor binding protein-1, fetal fibronectin, panty-liner with polymer-embedded strip) for diagnosing P-PROM?**
Why this is needed	
Importance to 'patients' or the population	P-PROM is relatively common. In the absence of clear pooling of amniotic fluid in the vagina, clinical assessment cannot be conclusive about the diagnosis. There is limited evidence about the accuracy of diagnostic tests (placental alpha-microglobulin-1, insulin-like growth factor binding protein-1, fetal fibronectin, panty-liner with polymer-embedded strip), and the results of available studies are inconclusive. Making the correct diagnosis is important, because women with a true positive diagnosis or a false negative diagnosis could benefit from prophylactic antibiotics, whereas women with a false positive diagnosis (who have intact fetal membranes) could be harmed by inappropriate use of prophylactic antibiotics. More research on the diagnostic accuracy of the various tests should evaluate both the performance of the tests themselves and their impact on

Research question	2. What is the diagnostic accuracy and utility of tests (placental alpha-microglobulin-1, insulin-like growth factor binding protein-1, fetal fibronectin, panty-liner with polymer-embedded strip) for diagnosing P-PROM?
	management and outcome. Studies should include subgroup analysis broken down by different gestational ages.
Relevance to NICE guidance	The research will fill a gap in the existing evidence and assist clinicians to identify the target population with P-PROM, but will not change the overall recommendation to offer antibiotics to women with P-PROM.
Relevance to the NHS	Accurate diagnosis of P-PROM may reduce the risk of infection and its serious and costly consequences for mother and baby.
National priorities	n/a
Current evidence base	Searches for the current Guideline found limited very low quality evidence from 2 prospective cohort studies (Section 6 in the full guideline). The available evidence on the tests failed to show consistently high positive or negative likelihood ratios.
Equality	No issues; the population is defined as pregnant women presenting with suspected P-PROM.
Feasibility Are there any ethical or technical issues?	The proposed research can be carried out within a realistic timescale and the interventions are relatively low-cost diagnostic tests. No ethical issues beyond those applying to perinatal research. The proposed research should evaluate both the performance of the tests themselves and their impact on management and outcome. Studies should include subgroup analysis broken down by different gestational ages.
Other comments	A false negative diagnosis is clinically important as these women will be denied prophylactic antibiotics.

1

2

6 Antenatal prophylactic antibiotics for women with P-PROM

6.1 Introduction

P-PROM is a major risk factor for intrauterine infection/ chorioamnionitis, which itself can be the cause of maternal sepsis (the leading direct cause of maternal death), and a major contributor to neonatal morbidity (such as pneumonia) and neonatal mortality. The objective of this review question is to evaluate the effectiveness of antibiotic prophylaxis offered to pregnant women whose membranes have ruptured preterm before labour has started for the prevention of early-onset neonatal infection. The focus population is women who have a diagnosis of P-PROM with no other indication for antibiotic therapy, for example in the absence of evidence of infection. We looked at the prophylactic efficacy of antibiotics compared to no antibiotics (or placebo) for improving neonatal and maternal outcomes in general and not the performance of individual antibiotic. However, further subgroup analysis was planned at the protocol stage to present results for different antibiotic classes in order to facilitate decision-making.

6.1.1 Review question

What is the clinical effectiveness of antenatal prophylactic antibiotics given to women with diagnosed preterm pre-labour rupture of membranes to improve outcomes of preterm labour?

6.1.2 Description of included studies

Three studies were included in this review (Kenyon 2013; Kenyon 2008; Mercer 2011): 2 SRs of RCTs (Kenyon 2013; Mercer, 2011) and one follow up of an earlier trial conducted in 2001 (Kenyon 2008) which was included in the earlier SRs. 16 RCTs from the SR by Kenyon 2013 met our protocol. Mercer 2011 is a further analysis of 5 of the included studies from Kenyon 2013.

The 2 included SRs evaluated the immediate and long-term effects of administering antibiotics to women with P-PROM before 37 weeks of pregnancy on maternal and neonatal outcomes. The third included study (Kenyon 2008) is a follow up of UK children at age 7 years born to women with mean gestational age of 32 weeks who participated in an earlier clinical trial that evaluated use of antibiotics in women presenting with P-PROM.

All included studies in the SRs specified that P-PROM was confirmed either with a speculum examination alone or in combination with a positive nitrazine test and 'ferning' of amniotic fluid.

Women with a diagnosis of infection or antibiotics taken during previous 7 to 10 days were excluded from the studies.

The type, route, dose and regimen of antibiotics used varied widely between the included trials in the SRs, for further details see the GRADE profiles below and the evidence table in Appendix H.

6.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The findings of the review are presented in the following GRADE profiles:

- Table 24: GRADE profile for comparison of antibiotic therapy versus placebo

Table 25: GRADE profile for comparison of antibiotic therapy versus either placebo or no antibiotic therapy

- Table 26: GRADE profile for comparison of antibiotic therapy versus no antibiotic therapy (childhood outcomes at 7-year follow up)

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

1 **Table 24: GRADE profile for comparison of antibiotic therapy versus placebo**

Quality assessment							No. of women / babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Gestation (weeks)	Antibiotic	Placebo	Relative (95% CI)	Absolute	Quality
Neonatal outcomes											
Perinatal death/death before discharge - Any antibiotic[a]											
1 meta-analysis of 12 studies (Kenyon 2013)	randomised trials	very serious[2,3,4,5,6,7,]	no serious inconsistency	serious[8]	serious [9]	20 to 34	276/4315 (6.4%)	138/1986 (6.9%)	RR 0.93 (0.76 to 1.14)	5 fewer per 1000 (from 17 fewer to 10 more)	Very low
Perinatal death/death before discharge - All penicillin (excluding co-amoxiclav)											
1 meta-analysis of 4 studies (Kenyon 2013)	randomised trials	serious[6]	no serious inconsistency	serious[10]	serious [9]	20 to 34	7/165 (4.2%)	10/167 (6%)	RR 0.78 (0.31 to 1.97)	13 fewer per 1000 (from 41 fewer to 58 more)	Very low
Perinatal death/death before discharge - Beta lactam (including co-amoxiclav)[b]											
1 meta-analysis of 2 studies (Kenyon 2013)	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious [9]	24 to 36	80/1236 (6.5%)	46/644 (7.1%)	RR 0.62 (0.15 to 2.56)	27 fewer per 1000 (from 61 fewer to 111 more)	Low
Perinatal death/death before discharge - Macrolide (including erythromycin) [c]											
1 meta-analysis of 4 studies (Kenyon 2013)	randomised trials	serious[5]	no serious inconsistency	serious[10]	serious 9	20 to 36	84/1354 (6.2%)	56/784 (7.1%)	RR 0.83 (0.43 to 1.60)	12 fewer per 1000 (from 41 fewer to 43 more)	Very low
Perinatal death/death before discharge - Other antibiotic [d]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	serious[7, 12]	no serious inconsistency	serious[10]	serious [9]	24 to 36	28/371 (7.5%)	26/391 (6.6%)	RR 1.13 (0.68 to 1.88)	9 more per 1000 (from 21 fewer to 59 more)	Very low
Neonatal encephalopathy - Any antibiotic[e]											
1 study (Kenyon 2013)[f]	randomised trials	serious13	no serious inconsistency	no serious indirectness	no serious imprecision	< 36	0/30 (0%)	0/30 (0%)	NC	NC	Moderate
Neonatal necrotising enterocolitis - Any antibiotic[f]											
1 meta-analysis of 11 studies	randomised trials	Serious [4,5,7,12,13]	no serious inconsistency	serious[14]	serious [9]	20 to 36	100/4273 (2.3%)	58/1956 (3%)	RR 1.09 (0.65 to 1.83)	3 more per 1000 (from 10 fewer to 25 more)	Very low

Quality assessment							No. of women / babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Gestation (weeks)	Antibiotic	Placebo	Relative (95% CI)	Absolute	Quality
(Kenyon 2013)											
Neonatal necrotising enterocolitis - All penicillin (excluding co-amoxiclav)[g]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	serious[13],	no serious inconsistency	no serious indirectness	serious [9]	20 to 36.	5/124 (4%)	6/138 (4.3%)	RR 0.85 (0.25 to 2.97)	7 fewer per 1000 (from 33 fewer to 86 more)	Low
Neonatal necrotising enterocolitis - Beta lactam (including co-amoxiclav)[h]											
1 meta-analysis of 2 studies (Kenyon 2013)	randomised trials	serious[4]	no serious inconsistency	serious [10]	no serious imprecision	20 to 36.	29/1236 (2.3%)	3/644 (0.47%)	RR 4.72 (1.57 to 14.23)	17 more per 1000 (from 3 more to 62 more)	Low
Neonatal necrotising enterocolitis - Macrolide (including erythromycin) [i]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	serious[13]	no serious inconsistency	no serious indirectness	serious [9]	20 to 36.	21/1322 (1.6%)	19/754 (2.5%)	RR 0.88 (0.45 to 1.69)	13 fewer per 1000 (from 58 fewer to 72 more)	Low
Neonatal necrotising enterocolitis - Other antibiotic [i]											
1 meta-analysis of 4 studies (Kenyon 2013)	randomised trials	serious[7,12]	no serious inconsistency	no serious indirectness	serious [9]	20 to 36	25/402 (6.2%)	30/421 (7.1%)	RR 0.89 (0.54 to 1.47)	8 fewer per 1000 (from 33 fewer to 33 more)	Low
Neonatal infection including pneumonia - Any antibiotic [j]											
1 meta-analysis of 12 studies (Kenyon 2013)	randomised trials	very serious [2, 4,11,6,5,12]	no serious inconsistency	serious8	no serious imprecision	20 to 36.	85/823 (10.3%)	141/857 (16.4%)	RR 0.67 (0.52 to 0.85)	47 fewer per 1000 (from 21 fewer to 68 fewer)	Very low
Neonatal infection including pneumonia - All penicillin (excluding co-amoxiclav) [k]											
1 meta-analysis of 5 studies (Kenyon 2013)	randomised trials	very serious [6,13]	no serious inconsistency	serious[10]	no serious imprecision	20 to 36.	6/258 (2.3%)	25/263 (9.5%)	RR 0.3 (0.13 to 0.68)	67 fewer per 1000 (from 30 fewer to 83 fewer)	Very low
Neonatal infection including pneumonia - Beta lactam (including co-amoxiclav) [l]											
1 study (Kenyon 2013)	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	serious [9]	24 to 29	0/31 (0%)	1/31 (3.2%)	RR 0.33 (0.01 to 7.88)	22 fewer per 1000 (from 32 fewer to 222 more)	Low

Quality assessment							No. of women / babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Gestation (weeks)	Antibiotic	Placebo	Relative (95% CI)	Absolute	Quality
Neonatal infection including pneumonia - Macrolide (including erythromycin) [m]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	serious1[3,5]	no serious inconsistency	no serious indirectness	serious [9]	20 to 36.	19/163 (11.7%)	25/171 (14.6%)	RR 0.79 (0.45 to 1.37)	31 fewer per 1000 (from 80 fewer to 54 more)	Very low
Neonatal infection including pneumonia - Other antibiotic [n]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	serious[7,12]	no serious inconsistency	serious[10]	no serious imprecision	20 to 36.	60/371 (16.2%)	90/392 (23%)	RR 0.71 (0.53 to 0.95)	67 fewer per 1000 (from 11 fewer to 108 fewer)	Very low
Birth before 37 weeks' gestation - Any antibiotic [o]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	serious[12,13]	no serious inconsistency	no serious indirectness	serious [9]	20 to 36.	3104/3642 (85.2%)	1102/1289 (85.5%)	RR 1.00 (0.98 to 1.03)	0 fewer per 1000 (from 17 fewer to 26 more)	Low
Birth within 7 days of randomisation - Any antibiotic [p]											
1 meta-analysis of 7 studies (Kenyon 2013)	randomised trials	serious[10,11]	no serious inconsistency	serious[8]	no serious imprecision	20 to 36	2388/4145 (57.6%)	1221/1820 (67.1%)	RR 0.79 (0.71 to 0.89)	141 fewer per 1000 (from 74 fewer to 195 fewer)	Low
Serious childhood disability at approximately 7 years - Any antibiotic [q]											
1 study (Kenyon 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious [10]	< 37	938/2375 (39.5%)	311/796 (39.1%)	RR 1.01 (0.91 to 1.12)	4 more per 1000 (from 35 fewer to 47 more)	Moderate
Maternal outcomes											
Maternal death - Any antibiotic [r]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	no serious risk of bias	no serious inconsistency	serious [1]	no serious imprecision	20 to 34	0/369 (0%)	0/394 (0%)	NC	NC	Moderate
Maternal death - All penicillin (excluding co-amoxiclav)											
1 study (Kenyon 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	20 to 34	0/40 (0%)	0/45 (0%)	NC	NC	High
Maternal death - Other antibiotic (not penicillin, beta-lactam or macrolide) [s]											

Quality assessment							No. of women / babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Gestation (weeks)	Antibiotic	Placebo	Relative (95% CI)	Absolute	Quality
1 meta-analysis of 2 studies (Kenyon 2013)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	20 to 36	0/329 (0%)	0/349 (0%)	NC	NC	Moderate
Maternal infection after birth prior to discharge - Any antibiotic[t]											
1 meta-analysis of 4 studies (Kenyon 2013)	randomised trials	serious[12,13]	no serious inconsistency	no serious indirectness	serious [14]	20 to 36	729/3943 (18.5%)	306/1604 (19.1%)	RR 0.91 (0.80 to 1.02)	17 fewer per 1000 (from 38 fewer to 4 more)	Low
Chorioamnionitis - Any antibiotic[u]											
1 meta-analysis of 11 studies (Kenyon 2013)	randomised trials	very serious[1,2,5,6,7,11,12]	no serious inconsistency	no serious indirectness	serious [14]	20 to 36	126/767 (16.4%)	196/792 (24.7%)	RR 0.66 (0.46 to 0.96)	84 fewer per 1000 (from 10 fewer to 134 fewer)	very low
Major maternal adverse drug reaction - Any antibiotic [v]											
1 meta-analysis of 3 studies (Kenyon 2013)	randomised trials	serious[12,13]	no serious inconsistency	no serious indirectness	serious [14]	20 to 36	0/3913 (0%)	0/1574 (0%)	NC	NC	Low

RR risk ratio, CI confidence interval, NC not calculable
a Interventions in the included studies: Mezlocillin followed by ampicillin or placebo, ampicillin and erythromycin, then amoxacillin and erythromycin or placebo, ampicillin, pivampicillin, and metronidazole or identical placebo
b Interventions in the included studies: Mezlocillin followed by ampicillin or matched placebo, ampicillin or matched placebo, penicillin or matched placebo, piperacillin or placebo
c Interventions in the included studies: Erythromycin or matched placebo, co-amoxiclav and erythromycin or matched placebo, erythromycin or matched placebo, eythromycin or matched placebo
d Interventions in the included studies: Ampicillin and erythromycin then amoxicillin and erythromycin or placebo indamycin and gentamycin or matched placebo, ampicillin, pivampicillin and metronidazole or identical placebo
e Interventions in the included study: Ampicillin, pivampicillin and metronidazole or matched placebo.
f Interventions in the included studies: Benzylpenicillin and penicillin or matched placebo, mezlocillin and ampicillin or matched placebo
ampicillin or matched placebo, mezlocillin and ampicillin or matched placebo, co-amoxiclav and erythromycin or matched placebo, ampicillin or matched placebos, penicillin or matched placebo, piperacillin or placebo, erythromycin or placebo, erythromycin or matched placebo, clindamycin and gentamycin or matching placebo
g Interventions in the included studies: Mezlocillin and ampicillin or matched placebo, mezlocillin and ampicillin or matched placebo, ampicillin or matched placebos
h Interventions in the included studies: benzylpenicillin and penicillin or matched placebo, co-amoxiclav and erythromycin or matched placebo
Interventions in the included studies: co-amoxiclav and erythromycin or matched placebo, penicillin or matched placebo, piperacillin or placebo
i Interventions in the included studies: Ampicillin or matched placebo, erythromycin or matched placebo, erythromycin or matched placebo, clindamycin and gentamycin or matching placebo

j Interventions in the included studies: benzylpenicillin and penicillin or matched placebo, metzlocillin or placebo, mezlocillin and ampicillin or matched placebo, ampicillin, Pivampicillin and metronidazole or identical placebo, co-amoxiclav or matched placebo, eythromycin or matched placebo, ampicillin or matched placebo, penicillin or matched placebo, piperacillin or placebo, erythromycin or placebo, erythromycin or matched placebo, clindamycin and gentamycin or matching placebo
k Interventions in the included studies: Metzlocillin or placebo, mezlocillin and ampicillin or matched placebo, co-amoxiclav or matched placebo, erythromycin or matched placebo, ampicillin or matched placebo
l Interventions in the included study: -Benzylpenicillin and penicillin or matched placebo
m Interventions in the included studies: Ampicillin, pivampicillin and metronidazole or identical placebo, penicillin or matched placebo, piperacillin or placebo
n Interventions in the included studies: Erythromycin or placebo, erythromycin or matched placebo, clindamycin and gentamycin or matched placebo
o Interventions in the included studies: Co-amoxiclav and erythromycin or matched placebo, penicillin or matched placebo, clindamycin and gentamycin or matching placebo
p Interventions in the included studies: Mezlocillin and ampicillin or matched placebo, ampicillin or matched placebo, mezlocillin and ampicillin or matched placebo, co-amoxiclav and erythromycin or matched placebo, ampicillin or matched placebo, piperacillin or placebo, erythromycin or placebo
q Interventions in the included study: Co-amoxiclav and erythromycin or matched placebo
r Interventions in the included studies: mezlocillin and ampicillin or placebo, ampicillin and erythromycin or placebo, ampicillin, pivampicillin and metronidazole or identical placebo
s Interventions in the included studies: ampicillin and erythromycin or placebo, ampicillin and metronidazole or identical placebo
t Interventions in the included studies: ampicillin, oral pivampicillin and metronidazole or identical placebo, co-amoxiclav and erythromycin or matched placebo, erythromycin or placebo, clindamycin and gentamycin or matched placebo
u Interventions in the included studies: IV metzlocillin or placebo, ampicillin, pivampicillin and metronidazole or identical placebo, ampicillin and ampicillin or matched placebo, co-amoxiclav or matched placebo, erythromycin or matched placebo, ampicillin and ampicillin or matched placebo, penicillin or matched placebo, piperacillin or placebo, erythromycin or placebo, clindamycin and gentamycin or matched placebo
v Interventions in the included studies: Co-amoxiclav and erythromycin or matched placebo, erythromycin or placebo, clindamycin and gentamycin or matching placebo
1 118/614 women were Group B Strep positive
2 Unclear method of randomisation in 1 study
3. Unclear allocation concealment in 4 studies
4 Data collected from an abstract in 1 study
5 15% of loss to follow up in one study
6 One study specified that 101 women were randomised but results for 115 women are reported
7 One study specified that trial stopped after intermediate evaluation showed treatment group had better outcome
8 Twin pregnancy included in 3 studies
9 Confidence intervals crossed one default MIDs
10 Twin pregnancy included in one study
11 Unclear allocation concealment in 3 studies
12 Data from one study extracted from a PhD thesis
13 Unclear allocation concealment in one study
14 Twin pregnancy included in 2 studies

Table 25: GRADE profile for comparison of antibiotic therapy versus either placebo or no antibiotic therapy

Quality assessment							No. of women / babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Gestation (weeks)	Antibiotic	No antibiotic	Relative (95% CI)	Absolute	Quality
Neonatal outcomes											
Perinatal death/death before discharge[a]											

Quality assessment							No. of women / babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Gestation (weeks)	Antibiotic	No antibiotic	Relative (95% CI)	Absolute	Quality
1 meta-analysis of 18 studies (Kenyon 2013)	randomised trials	very serious[1,2,3,4,5,6,7,8]	no serious inconsistency	serious[9]	no serious imprecision	20 to 36	299/4604 (6.5%)	172/2268 (7.6%)	RR 0.89 (0.74 to 1.08)	8 fewer per 1000 (from 20 fewer to 6 more)	Very low
Intraventricular haemorrhage[b]											
1 meta-analysis of 7 studies (Mercer, 2011)	randomised trials	very serious[12,13]	no serious inconsistency	serious [14]	no serious imprecision	20 to 36	74/572 (12.9%)	105/590 (17.8%)	RR 0.73 (0.56 to 0.95)	48 fewer per 1000 (from 9 fewer to 78 fewer)	Very low
Neonatal sepsis[c]											
1 meta-analysis of 5 studies (Mercer, 2011)	randomised trials	serious[13]	no serious inconsistency	serious [14]	no serious imprecision	20 to 36	53/485 (10.9%)	82/489 (16.8%)	RR 0.67 (0.49 to 0.91)	55 fewer per 1000 (from 15 fewer to 86 fewer)	Low
Birth delayed ≥ 7 days after randomisation[d]											
1 meta-analysis of 6 studies (Mercer, 2011)	randomised trials	very serious[12,13]	no serious inconsistency	serious [14]	no serious imprecision	20 to 36	237/515 (46%)	139/537 (25.9%)	RR 1.8 (1.52 to 2.13)	207 more per 1000 (from 135 more to 292 more)	Very low

RR risk ratio, CI confidence interval, NC not calculable.
a Interventions in the included studies: Ampicillin, ampicillin, gentamycin and amoxicillin and clavulanic acid, co-amoxiclav or matched placebo, erythromycin or matched placebo, ampicillin or matched placebos, metzlocillin and ampicillin or matched placebo, co-amoxiclav and erythromycin or matched placebo, penicillin or matched placebo, piperacillin or placebo, co-amoxiclav or placebo, erythromycin or placebo, erythromycin or matched placebo, a-mpicillin, erythromycin and amoxacillin or matched placebo, ampicillin, clindamycin and gentamycin or matched placebo, ampicillin or matched placebo, ampicillin, pivampicillin and metronidazole or identical placebo
b Interventions in the included studies: Ampicillin, ampicillin, gentamycin and amoxicillin and clavulanic acid, metzlocillin or placebo, IV Metzlocillin and ampicillin or matched placebo, piperacillin or placebo, erythromycin or matched placebo
- Cindamycin and gentamycin or matched placebo
c Interventions in the included studies: Ampicillin, ampicillin, gentamycin, amoxicillin and clavulanic acid, piperacillin or placebo, ampicillin, erythromycin and amoxicillin or matched placebo, ampicillin
d Interventions in the included studies: Ampicillin, ampicillin, gentamycin, amoxicillin and clavulanic acid or placebo, metzlocillin or placebo, metzlocillin and ampicillin or matched placebo, piperacillin or placebo, ampicillin, erythromycin and amoxicillin or matched placebo
1 No blinding in 6 studies
2 Unclear method of randomisation in 3 studies
3 Unclear allocation concealment in 8 studies.
4 Data from extracted from a PhD thesis one study
5 15% of loss to follow up in one study.
6 One study specified that 101 women were randomised but results for 115 are reported

7 One study specified that trial stopped after intermediate evaluation showed treatment group had better outcome
8 Data from extracted from an abstract in one study
9 Twin pregnancy included in 3 studies
10 No blinding in 5 studies
11 unclear method of randomisation in one study
12 Unclear allocation concealment in 2 studies
13 No blinding in 2 studies
14 Multiple pregnancy included in one study

Table 26: GRADE profile for comparison of antibiotic therapy versus no antibiotic therapy (childhood outcomes at 7-year follow up)

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Any antibiotic versus no antibiotics	Control	Relative (95% CI)	Absolute	Quality
Total death/stillbirths, death in first year, death after first year - Any erythromycin versus no erythromycin											
1 study (Kenyon 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	156/2323 (6.7%)	172/2389 (7.2%)	RR 0.93 (0.74 to 1.16)	5 fewer per 1000 (from 19 fewer to 11 more)	Moderate
Total death/death after first year - Any erythromycin versus no erythromycin											
1 study (Kenyon 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	7/2323 (0.3%)	4/2389 (0.2%)	RR 1.79 (0.52 to 6.12)	1 more per 1000 (from 1 fewer to 9 more)	Moderate
Total death/stillbirths, death in first year, death after first year - Any co-amoxiclav versus no co-amoxiclav											
1 study (Kenyon 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	None	163/2336 (7.0%)	165/2376 (6.9%)	RR 1.01 (0.80 to 1.26)	1 more per 1000 (from 14 fewer to 18 more)	Low
Total death/death after first year - Any co-amoxiclav versus no co-amoxiclav											
1 study (Kenyon 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	5/2336 (0.2%)	6/2376 (0.3%)	RR 0.85 (0.26 to 2.78)	0 fewer per 1000 (from 2 fewer to 4 more)	Moderate
Cerebral palsy- Any erythromycin versus no erythromycin											
1 study (Kenyon 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	46/1590 (2.9%)	41/1671 (2.5%)	RR 1.18 (0.77 to 1.81)	4 more per 1000 (from 6 fewer to 20 more)	Moderate
Cerebral palsy - Any co-amoxiclav versus no co-amoxiclav											
1 study (Kenyon 2008)	randomised trials	no serious	no serious inconsistency	serious[1]	no serious imprecision	None	39/1632 (2.4%)	48/1629 (2.9%)	RR 0.81 (0.53 to 1.24)	24 fewer per 1000 (from 14	Moderate

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Any antibiotic versus no antibiotics	Control	Relative (95% CI)	Absolute	Quality
		risk of bias								fewer to 7 more)	
Any functional impairment at age 7 - Any erythromycin versus no erythromycin											
1 study (Kenyon 2001)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	None	594/1551 (38.3%)	655/1620 (40.4%)	RR 0.91 (0.79 to 1.05)	36 fewer per 1000 (from 85 fewer to 20 more)	Low
Any functional impairment at age 7 - Any co-amoxiclav versus no co-amoxiclav											
1 study (Kenyon 2001)[a]	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Very serious[3]	None	645/1587 (40.6%)	604/1584 (38.1%)	RR 1.11 (0.96 to 1.25)	42 more per 1000 (from 15 fewer to 95 more)	Very low

RR risk ratio, CI confidence interval, NC not calculable.
1 Multiple pregnancy included
2 Confidence intervals crossed one default MIDs
3 Confidence intervals crossed 2 default MIDs

6.1.4 Evidence statements

Any antibiotic therapy compared with placebo

Neonatal outcomes

Very low quality evidence from meta-analysis of 12 studies with 1600 women with suspected or diagnosed preterm pre-labour rupture of membranes (P-PROM) found that neonatal infection including pneumonia is significantly lower in the group of women who received prophylactic antibiotics compared with those allocated to receive placebo.

A meta-analysis of 7 studies (low quality evidence) from almost 6000 women with P-PROM found that significantly fewer of the women who received prophylactic antibiotics gave birth within 7 days of randomisation compared with those allocated to receive placebo.

Low quality evidence from one meta-analysis of 2 RCTs of over 1800 women with suspected or diagnosed P-PROM showed a higher incidence of necrotising enterocolitis in babies born to women who received prophylactic beta lactam antibiotics including co-amoxiclav compared with those allocated to receive a placebo. No significant difference was found for the other outcomes (perinatal death, neonatal encephalopathy, positive neonatal blood culture, birth before 37 weeks of gestation, serious childhood disability at approximately 7 years) from moderate to very low quality evidence from meta-analysis of 2 to 12 RCTs (number of participants ranged from 61 to 6301).

Maternal outcomes

High to very low quality evidence from meta-analysis of 2 to 11 RCTs (number of participants ranged from 85 to 1559) found no significant difference in maternal death, maternal infection after birth prior to discharge, and major maternal adverse drug reaction in women with preterm pre-labour rupture of membranes allocated to receive prophylactic antibiotics compared with those allocated to receive placebo treatment.

Very low quality evidence from meta-analysis of 11 RCTs (n=1559) indicated that incidence of chorioamnionitis was lower in women with P-PROM allocated to receive prophylactic antibiotics compared with those allocated to receive placebo treatment.

Antibiotic therapy compared with either placebo or no antibiotic therapy

Neonatal outcomes

Very low quality evidence from meta-analysis of 18 RCTs (n=6872) showed no difference in the incidence of perinatal death/death before discharge in neonates whose mother with P-PROM was allocated to receive prophylactic antibiotics compared with those allocated to receive no antibiotic treatment.

Low and very low quality evidence from 5 to 7 RCTs (number of participants ranged from 974 to 1162) found that the incidence of each of intraventricular haemorrhage and sepsis were significantly lower and the number of births delayed by at least 7 days after randomisation was significantly higher in neonates whose mother with P-PROM was allocated to receive prophylactic antibiotics compared with those allocated to receive no antibiotic treatment.

Maternal outcomes relevant to this section were not reported.

Antibiotic therapy compared with no antibiotic therapy (childhood outcomes at 7-year follow up)

Moderate to very low quality evidence from one RCT (n=4712) showed no significant difference in the incidence of total death, death after first year, cerebral palsy and functional impairment at age 7 years in neonates whose mother was diagnosed with P-PROM and allocated to receive prophylactic antibiotics (either erythromycin or co-amoxiclav) compared with those allocated to receive no antibiotic treatment..

6.1.5 Health economics profile

A search was undertaken for health economic evidence on antenatal prophylactic antibiotics given to women with diagnosed preterm pre-labour rupture of membranes to improve outcomes of preterm labour. A total of 73 articles were identified by the search. After reviewing titles and abstracts, 2 full papers were obtained and one was included for review.

A UK study (Colbourn 2007) considered a range of prenatal strategies for preventing group B streptococcus'and other serious bacterial infections in early infancy. Their analysis looked at 12 different populations or risk groups but includes a preterm group with membrane rupture more than 2 hours before labour onset and suggested that treatment with intravenous antibiotics would be cost-effective in such a population.

This question was not prioritised for health economic analysis as it was thought by the Committee to be a cheap intervention and something that was part of current clinical practice. Furthermore, there was a strong a priori expectation that the guideline would not change current practice.

6.1.6 Evidence to recommendations

6.1.6.1 Relative value placed on the outcomes considered

The Committee prioritised the following clinical outcomes:

- in terms of maternal outcomes: mortality, maternal infections (such as chorioamnionitis) and major adverse events
- for neonatal outcomes: neonatal or perinatal mortality, number of babies born preterm, brain injury including intraventricular haemorrhage, periventricular leucomalacia (PVL)/white matter injury, necrotising enterocolitis, or any neonatal infection (including neonatal sepsis). The protocol also included any long-term outcomes in childhood (particularly functional impairments, behavioural difficulties, cerebral palsy, seizures and wheezing) by taking into account that the long term impact may be affected by other influences (not necessarily the administration of antibiotics before delivery) and for that reason long term neurological outcomes were not included.

Evidence on neonatal encephalopathy was sought but not found, however the Committee did not consider this as a critical outcome for drafting these recommendations in the context of long-term follow-up.

When considering the relative value on each outcome, the Committee assumed that outcomes relating to infection in the baby would pertain to early onset neonatal infection. Although antibiotics given to women with P-PROM appeared to reduce the rate of positive neonatal blood cultures, the Committee placed little additional weight on this outcome, over and above the other beneficial effects for the baby.

6.1.6.2 Consideration of clinical benefits and harms

The evidence from the included SRs and meta-analyses showed that prophylactic antibiotics for women with P-PROM may reduce the incidence of chorioamnionitis but this effect was not found for other indices of maternal infection.

Low quality evidence was found which showed that prophylactic antibiotics for women with P-PROM might delay birth for more than 7 days. The Committee was uncertain whether birth in the reviewed studies followed spontaneous (preterm) labour or was due to other factors. In terms of the neonatal outcomes, there was some clear evidence that neonatal infections including pneumonia and sepsis were reduced by antibiotic prophylaxis in women with P-PROM.

Although intraventricular haemorrhage appeared to be reduced with the use of prophylactic antibiotics compared to no antibiotics for women with P-PROM, the majority of evidence was from babies who were born in the late 1980s. The Committee questioned the relevance of the data to current practice, and they inferred that it is most likely that these babies may have already compromised health status so that in general, infection would have just been one more added problem rather than the primary cause of haemorrhage. Hence the Committee gave little weight to the apparent benefit of antibiotics for P-PROM in preventing haemorrhage.

In addition, the Committee discussed the challenge in interpreting the results on infection, given that there was no clear indication of the nature of infections in the included evidence and the term “pneumonia” might be too heterogeneous to draw clinically relevant results.

In summary, although antibiotics given to mothers with P-PROM seem to have little effect on the long-term health outcomes of children, the short-term advantages (reducing neonatal infection, and delaying birth) are such that the Committee decided that antibiotics should be offered routinely to all women with P-PROM.

Although the evidence base for this section was not robust, the Committee concluded that this recommendation should be strong. Giving antibiotics to women with P-PROM is currently standard clinical practice in the UK and the review of evidence in this question showed no reason to change this practice. More specifically, the evidence of no harm for the baby in terms of cerebral palsy or for the mother in terms of major maternal adverse drug reaction further confirmed the direction and the strength of the recommendation. The Committee discussed that the absence of any major maternal drug side effects can be explained because the history taking in women with P-PROM to identify allergies can determine the appropriate class of antibiotics to be offered.

Choice of antibiotic

The Committee noted that, although antibiotics overall had no effect on necrotising enterocolitis, the beta lactam antibiotics (including co-amoxiclav) significantly increased the risk of necrotising enterocolitis, although beta lactams (including co-amoxiclav) reduced neonatal infection including pneumonia. Hence the Committee decided that beta-lactam antibiotics should not be selected to improve neonatal outcomes in women with P-PROM.

Regarding other available antibiotics, the Committee considered that, in addition to the benefits of erythromycin shown in the evidence summary above, there are additional potential benefits of erythromycin as the choice of antibiotic in women with P-PROM. Firstly, erythromycin is not reported to increase the risk of necrotising enterocolitis. Secondly, it can be administered orally to target group B streptococcus, other streptococcal and staphylococcal infections, bacteria relevant to early-onset sepsis, and other micro-organisms affecting the woman and baby before labour. Thirdly, erythromycin also offers a theoretical advantage (for the woman rather than the baby) in that it can counteract mycoplasma infection that is implicated in the early stages of chorioamnionitis (this effect is not seen with

penicillins). Finally, the absorption of erythromycin across the gastrointestinal tract and the placenta is limited, which suggests a potential benefit of minimising the baby's exposure to antibiotics. The Committee noted that although there was little evidence of benefit to the baby, there was no evidence of harm. Although few antibiotics are licensed for use in pregnancy or in preterm babies there was a strong consensus within the Committee that healthcare professionals should consider antibiotic prophylaxis using erythromycin for women with P-PROM.

The Committee also specified that the recommended dosage of erythromycin in this section should be 250 mg 4 times per day, as this was the dosage used in the largest study included in the evidence basis of this recommendation.

6.1.6.3 Consideration of health benefits and resource uses

Antibiotics are cheap and infection could result in longer stay in hospital. Furthermore, infection may lead to poor health outcomes the cost of which can be very high. However, there are also concerns that the overuse of antibiotics can promote antibiotic resistance which has potentially large implications for future health benefits and costs. There was some evidence that antibiotics could delay birth which could have some resource implications, although the Committee noted that women with P-PROM are now more likely to be managed as outpatients than was historically the case.

6.1.6.4 Quality of evidence

The majority of evidence was downgraded in this section due to high risk of bias and imprecision. Lack of blinding was the most common area for the quality of evidence to be downgraded. Specifically for the outcome of prolonging delivery for 7 or more days, the Committee discussed the importance of masking treatment allocation in trials on interpretation of results and clinicians decision-making. However, the study with the highest weight in the meta-analysis was masked, therefore the Committee placed confidence in the estimates of effect for this outcome.

The Committee discussed the importance of accurate diagnosis of P-PROM and highlighted potential for harm to the baby (for example by increased risk of cerebral palsy) from inadvertent use of prophylactic antibiotics in women with intact membranes who are incorrectly diagnosed with P-PROM.

When the evidence was examined, the Committee noted that participants in the included trials would have been giving birth more than 15 years ago, with the majority delivering even longer ago, in the 1980s. These studies would have included different profile of babies' health status, being in worse health generally resulting in an overestimation of the magnitude of any effect (positive or negative). The population in these studies would not reflect the population now being treated therefore limiting the generalisability of its results.

6.1.6.5 Other considerations

These recommendations were based on both the clinical interpretation of evidence and on Guideline Committee's clinical expert opinion.

When the Committee discussed the use of erythromycin they noted that the benefit might not be entirely due to bactericidal properties but also the reduction in associated inflammatory damage might influence outcomes.

When the Committee discussed the role of antibiotics given to women with P-PROM as prophylactic measures for improving neonatal or mortality outcomes, there was a lack of information on this role at different gestational ages.

6.1.7 Recommendations

11. Offer women with P-PROM oral erythromycin 250 mg 4 times a day[d] for a maximum of 10 days or until the woman is in established labour (whichever is sooner).

12. For women with P-PROM who cannot tolerate erythromycin or in whom erythromycin is contraindicated, consider oral penicillin for a maximum of 10 days or until the woman is in established labour (whichever is sooner).

13. Do not offer women with P-PROM co-amoxiclav as prophylaxis for intrauterine infection.

14. For guidance on the use of intrapartum antibiotics, see the NICE guideline on Antibiotics for early-onset neonatal infection.

d At the time of consultation (June 2015), erythromycin did not have a UK marketing authorisation for use in pregnancy. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Prescribing guidance: prescribing unlicensed medicines for further information. The summaries of product characteristics for oral erythromycin recommend different dosages. The evidence reviewed for the guideline supports a dosage of 250 mg 4 times a day for prophylaxis in women with P-PROM.

7 Identifying infection in women with P-PROM

7.1 Introduction

Preterm pre-labour rupture of membranes exposes fetus and mother to risk of infection. Bacterial infection can be life-threatening to the mother and to fetus in utero or the baby postnatally. Ascending infection with Group B streptococcus is particularly dangerous. Other infections of the amniotic membranes or umbilical cord also put the fetus at risk of cytokine-induced white matter damage in the brain which may result in periventricular leukomalacia and cerebral palsy. The cytokines produced by the maternal immune response to infection can also trigger premature labour. Hence, there is a need to be able to identify maternal infection early to institute appropriate treatment with antibiotics and to avoid the unnecessary use of antibiotics where there is no infection. Indiscriminate use of antibiotics potentially puts the mother and fetus at risk of side effects and may encourage antibiotic resistance.

7.1.1 Review question

What is the diagnostic value of temperature, pulse, white cell count, C-reactive protein and cardiotocography (CTG) to identify infection in women with preterm pre-labour rupture of membranes (P-PROM)?

7.1.2 Description of included studies

Thirteen studies were included in this review. One of the included studies was a RCT (Lewis 1999), which randomised women with preterm pre-labour rupture of membranes to either a CTG or biophysical profile. Only data for the CTG arm are relevant to this review. One study analysed women with preterm pre-labour rupture of membranes participating in a RCT of corticosteroids (Garite and Freeman, 1982). Nine studies were prospective case series, one of which was a consecutive case series (Kurk 1990) and in the other 8 studies it was unclear whether women were recruited consecutively (Carroll 1995; Farb 1983; Fisk 1987; Hawrylyshyn 1983; Ismail 1985; Perrone 2012; Romem and Artal 1984; Yoon 1996). Two studies were retrospective case series (Del Valle 1992; Smith 2012)

Nine of the studies (Farb 1983; Fisk 1987; Hawrylyshyn 1983; Ismail 1985; Kurki 1990; Perrone 2012; Romem and Artal 1984; Smith 2012; Yoon 1996), reported values for C-reactive protein at various thresholds as a predictor of clinical amnionitis or chorioamnionitis, or histological chorioamnionitis. Four of the studies reported values for white blood cell count at various thresholds as a predictor of clinical or histological chorioamnionitis (Garite and Freeman 1982; Hawrylyshyn 1983; Romem and Artal 1984; Yoon 1996) and one study reported values for maternal temperature as a predictor of clinical or histological chorioamnionitis (Ismail 1985). The prevalence of histological chorioamnionitis ranged from 21% to 63% (6 studies) and of clinical chorioamnionitis ranged from 14% to 29% (4 studies). None of the included studies looked at the role of maternal C-reactive protein or maternal white blood cell count in identifying neonatal sepsis.

Three studies reported values for the CTG as a predictor of neonatal infection (Carroll 1995; Del Valle 1992; Lewis 1999), one study reported values for fetal heart rate as a predictor of both clinical and histological chorioamnionitis (Ismail 1985) and one study reported values for fetal heart rate as a predictor of clinical chorioamnionitis (Garite and Freeman 1982).

The mean gestational age at rupture of membranes was reported in 6 studies and ranged from 26.7 weeks (standard deviation 0.8) to 31.8 weeks (standard deviation 2.6). The duration of preterm pre-labour rupture of membranes was reported in 5 studies and ranged

from 3.5 days (standard deviation 12.1) to 16 days (standard deviation 12) (see evidence table in Appendix H for further details of included studies).

In the majority of studies, maternal serum samples for C-reactive protein determination and/or white blood cell count were taken on a daily basis from admission until birth. In the 3 studies where the CTG was the index test, the test was performed daily. The timing of the test results selected for analysis was not clearly reported in the majority of the studies. Two studies reported diagnostic accuracy values for the last CTG performed immediately before birth (Del Valle 1992; Lewis 1999), 2 studies reported predictive values for the last recorded C-reactive protein level taken before birth (Fisk 1987; Smith 2012), one study reported results for C-reactive protein taken at admission (Romem and Artal 1984), one study looked at the role of C-reactive protein samples taken at admission and samples taken 24 to 48 hours before birth (Perrone 2012) and one study focused at on white blood cell count and fetal heart rate measured at admission (Garite and Freeman 1982).

The majority of studies (9/13) required that ruptured membranes were confirmed both by visualisation of amniotic fluid and a positive biochemical test. 5 studies reported that maternal antibiotic therapy was not given during the period before birth (Yoon 1996; Hawrylyshyn 1983; Romen and Artal 1984; Farb 1983; Fisk 1987), 3 reported women were given antibiotics on clinical diagnosis of chorioamnionitis (Smith 2012; Del Valle 1992; Garite and Freeman 1982) and 2 reported the administration of routine prophylactic antibiotics (Perrone 2012; Lewis 1999). Three studies did not give any information in relation to use of antibiotic.

Two studies included women with a multiple pregnancy (10% of women in Kurki 1990; 8% of women in Fisk 1987 but only data in singleton pregnancies is included in the review for this study), 2 studies stated only singleton pregnancies were included (Perrone 2012; Yoon 1996) and in the remaining studies it was unclear if women with a multiple pregnancy were included.

There were no studies identified on the predictive value of maternal pulse to identify infection in women with preterm pre-labour rupture of membranes.

7.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Data is reported in GRADE profiles below separately for the following tests.

- Table 27: GRADE profile for predictive accuracy of C-reactive protein for identifying infection
- Table 28: GRADE profile for predictive accuracy of maternal white blood cell count for identifying infection
- Table 29: GRADE profile for predictive accuracy of fetal heart rate for identifying infection)
-
- Table 30: GRADE profile for predictive accuracy of maternal temperature for identifying infection

The specific tests and the thresholds used (e.g. C-reactive protein >2 mg/dL) are listed in the rows of the GRADE table, and the outcomes that they predict are listed in the ‘definition of outcome’ column.

Evidence from prospective case series started at high quality for the purposes of this review question and was then downgraded if there were any issues identified that would undermine the trustworthiness of the findings (for example, if it was unclear whether consecutive women were included in the study). Retrospective case series started at moderate quality and was

then downgraded if there were any issues (for example, if it was unclear whether consecutive women were included in the study).

Findings are reported separately for each study since the timing of testing, administration of antibiotic therapy, definitions of outcome measures and thresholds used vary across studies or are not clearly reported thus making pooling of data inappropriate. In order to provide a synthesis of findings the range of all values for a particular test is given in the first row of the relevant GRADE profile.

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

1 **Table 27: GRADE profile for predictive accuracy of C-reactive protein for identifying infection**

Quality assessment								Measures of diagnostic accuracy (95% confidence interval)				
Number. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Outcome and prevalence (type of infection)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Quality
C-reactive protein – all thresholds measured at a range of time points												
Overall summary of findings from 9 studies	case series	very serious[1,2]	very serious [4]	no serious indirectness	Very serious [5]	Clinical chorioamnionitis, histological chorioamnionitis or histological funisitis 14% to 63%		Range: 37% to 94% Low to high	Range: 32% to 100% Low to high	Range: 1.13 to 23.0 Not useful to very useful	Range: 0.12 to 0.72 Not useful to moderately useful	Very low to high
C-reactive protein ≥ 0.7 mg/dL measured within 72 hours of birth												
1 study (Yoon 1996)	case series	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	Histological chorioamnionitis (56%)		54% (37.78 to 70.79) Low	86% (72.75 to 98.68) Moderate	3.8 (1.46 to 9.89) Not useful	0.53 (0.36 to 0.79) Not useful	Low
C-reactive protein > 1.2 mg/dL*												
1 study (Kurki 1990)	case series	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[7]	Clinical and histological chorioamnionitis (22%)	33/147	94% (85.8 to 100) High	50% (40.82 to 59.18) Low	1.88 (1.53 to 2.30) Not useful	0.12 (0.03 to 0.47)Moderately useful	Moderate
C-reactive protein > 1.2 mg/dL measured on admission (admission to birth interval: mean 16 days (SD 12 days)												
1 study (Perrone 2012)	case series	serious risk of bias[2]	no serious inconsistency	no serious indirectness	Serious[6]	Histological funisitis (36%)	24/66	41.7% (24.5 to 61.2) Low	83.3% (69.4 to 91.7) Moderate	2.5 (1.10 to 5.71[j] Not useful	0.70 (0.49 to 1.01) Not useful	Low
C-reactive protein > 1.2 mg/dL measured 24 to 48 hours before birth												
1 study (Perrone 2012)	case series	serious risk of bias[2]	no serious inconsistency	no serious indirectness	Serious[6]	Histological funisitis (33%)	24/66	75.0% (55.1 to 88.0) Moderate	69.0% (54.0 to 80.9) Low	2.42 (1.46 to 4.02) Not useful	0.36 (0.18 to 0.75) Moderately useful	Low
C-reactive protein > 1.25 mg/dL measured at birth or last results obtained during hospital admission if discharged undelivered												
1 study (Hawrylyshyn 1983)	case series	serious[2]	no serious inconsistency	no serious indirectness	Serious7	Histological chorioamnionitis (50%)	26/52	88% (76.18 to 100) a Moderate	96% (88.76 to 100) High	23.00 (3.35 to 157.97) Very useful	0.12 (0.04 to 0.35) Moderately useful	Low
C-reactive protein ≥ 2 mg/dL measured on admission												
1 study (Romem and Artal 1984)	case series	serious[2]	no serious inconsistency	no serious indirectness	very serious [3]	Clinical chorioamnionitis (14%)	7/51	86% (59.79 to 100) Moderate	82% (70.42 to 93.21) Moderate	4.71 (2.35 to 9.46)Not useful	0.17 (0.03 to 1.08)Moderately useful	Very low
C-reactive protein > 2 mg/dL*												

Quality assessment								Measures of diagnostic accuracy (95% confidence interval)				
Number. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Outcome and prevalence (type of infection)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Quality
1 study (Farb 1983)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious [6]	Clinical chorioamnionitis (29%)	9/31	56% (23.09 to 88.02) Low	73% (54.12 to 91.34)Low	2.04 (0.83 to 5.00) Not useful	0.61 (0.28 to 1.33) Not useful	Low
1 study (Ismail 1985)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious [6]	Clinical chorioamnionitis (18%)	18/100	82% (66.12 to 100)Moderate	55% (44.11 to 65.65)Low	1.85 (1.35 to 2.53)Not useful	0.30 (0.11 to 0.87)[b] Moderately useful	Low
1 study (Farb 1983)	case series	serious[2]	no serious inconsistency	no serious indirectness	Very serious [5]	Histological chorioamnionitis (21%)	5/24	80% (44.94 to 100)Moderate	68% (47.52 to 89.32)Low	2.53 (1.15 to 5.60)Not useful	0.29 (0.05 to 1.73) Moderately useful	Very low
1 study (Ismail 1985)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious [6]	Histological chorioamnionitis (63%)	63/100	67% (55.03 to 78.31) Low	81% (68.46 to 93.70)[a] Moderate	3.52 (1.77 to 7.02) Not useful	0.41 (0.28 to 0.60)Moderately useful	Low
C-reactive protein > 2 mg/dL measured on admission (mean 16 days (SD 12 days) from admission to birth)												
1 study (Perrone 2012)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious [6]	Histological funisitis (33%)	24/66	37.5% (21.2 to 57.3 Low	90.5% (77.9 to 96.2) High	3.94 (1.36 to 11.43) Not useful	0.69 (0.50 to 0.96) Not useful	Low
C-reactive protein > 2 mg/dL taken within 48 hours of birth												
1 study (Fisk 1987)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious [6]	Histological chorioamnionitis (59%)	30/51	50% (32.11 to 67.89)[a] Low	81% (64.16 to 97.25)a Moderate	2.63 (1.01 to 6.80)[b] Not useful	0.62 (0.41 to 0.93)[b] Not useful	Low
C-reactive protein > 2 mg/dL measured 24 to 48 hours before birth												
1 study (Perrone 2012)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious [6]	Histological funisitis (33%)	24/66	54.2% (35.1 to 72.1) [d] Low	88.1% (75.0 to 94.8) [d] Moderate	4.55 (1.85 to 11.21) [b] Not useful	0.52 (0.33 to 0.82) [b] Not useful	Low
C-reactive protein > 3 mg/dL measured within 48 hours of birth												
1 study (Fisk 1987)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious 6	Histological chorioamnionitis (59%)	30/51	47% (28.81 to 64.52)[a] Low	90% (77.92 to 100) [a] High	4.9 (1.24 to 19.33)[b] Not useful	0.59 (0.41 to 0.85)[b] Not useful	Low
C-reactive protein > 3.5 mg/dL taken within 48 hours of birth												
1 study (Fisk 1987)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious 6	Histological chorioamnionitis (59%)	30/51	40% (22.47 to 57.53) [a] Low	95% (86.13 to 100) [a] High	8.4 (1.18 to 59.77) [b] Moderately useful	0.63 (0.46 to 0.86) [b] Not useful	Low
C-reactive protein > 4 mg/dL												

Quality assessment								Measures of diagnostic accuracy (95% confidence interval)				
Number. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Outcome and prevalence (type of infection)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Quality
1 study (Kurki 1990)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious 6	Clinical chorioamnionitis (22%)	33/147	72% (57.53 to 87.92) [a] Low	77% (69.49 to 84.90)[a] Moderate	3.19 (2.14 to 4.74)[b] Not useful	0.35 (0.20 to 0.62)[b] Moderately useful	Low
C-reactive protein > 4 mg/dL taken within 48 hours of birth												
1 study (Fisk 1987)	case series	serious[2]	no serious inconsistency	no serious indirectness	serious 6	Histological chorioamnionitis 59%	30/51	37% (19.42 to 53.91)a Low	100% (100 to 100) [a] High	NC Very useful	0.63 (0.48 to 0.83) [b] Not useful	Low
C-reactive protein > 5 mg/dL(measurement closest to time of birth reported)*												
1 study (Smith 2012)	randomised trial	serious[2]	no serious inconsistency	no serious indirectness - MULTIPS	serious 6	Histological chorioamnionitis (36%)	26/73	76.9% (60.73 to 93.12)[a] Moderate	31.9% (18.59 to 45.24)[a] Low	1.13 (0.85 to 1.51)[b] Not useful	0.72 (0.32 to 1.64)[b] Not useful	Low

** Timing of measurement not reported/unclear*
NC not calculable
1 Selection bias – only women who gave birth within 72 hours of amniocentesis were analysed and unclear whether consecutive women were included in the study
2 Possible selection bias - unclear whether consecutive women were included in the study
3 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)
4 Evidence was downgraded by 2 due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%). 5 Summary of findings from studies reporting a wide range of values and confidence intervals
5 Evidence was downgraded by 2 due to 95% confidence interval for the positive likelihood ratio ranges from not useful (<5) to very useful (>10)
6 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5)
7 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from moderately useful (>0.1-0.5) to very useful (0-0.1)

Table 28: GRADE profile for predictive accuracy of maternal white blood cell count for identifying infection

Quality assessment								Measures of diagnostic accuracy				
Number. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Outcome and prevalence (type of infection)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Quality
White blood cell count - all thresholds measured at a range of time points												
Overall summary of findings from 4 studies	case series	serious[1,2]	very serious[3]	no serious indirectness	Serious4	Clinical or histological chorioamnionitis (14% to 56%)		Range:16.7% to 80% Low to moderate	Range: 62% to 97.5% Low to high	Range: 2.10 to 6.70 Not useful to moderately useful	Range: 0.31 to 0.85 Not useful	Low to moderate
White blood cell count > 12,500 cells/mm^3 measured at birth												
1 study (Hawrylyshyn 1983)	case series	serious[1]	no serious inconsistency	no serious indirectness	serious4	Histological chorioamnionitis (50%)	26/52	80% (65.62 to 95.92)[a] Moderate	62% (42.84 to 80.24) [a] Low	2.10 (1.25 to 3.54)[b] Not useful	0.31 (0.13 to 0.73)[b] Moderately useful	Low
White blood cell count ≥ 12,500 cells/mm^3*												
1 study (Romem and Artal 1984)	case series	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	Clinical chorioamnionitis (14%)	7/51	43% (6.20 to 79.52)[a] Low	82% (70.42 to 93.21) [a] Moderate	2.36 (0.82 to 6.81) [b] Not useful	0.70 (0.36 to 1.35)[b] Not useful	Low
White blood cell count ≥ 13,000 cells/mm3 measured within 72 hours of birth												
Yoon 1996	case series	Serious2	no serious inconsistency	no serious indirectness	no serious imprecision	Histological chorioamnionitis (56%)	35/63	40% (23.77 to 56.23) a Low	82% (67.96 to 96.35) a Moderate	2.24 (0.92 to 5.47) b Not useful	0.73 (0.53 to 1.01) b Not useful	Moderate
White blood cell count ≥ 16,000 cells/mm3*												
1 study (Romem and Artal 1984)	case series	serious1	no serious inconsistency	no serious indirectness	serious5	Clinical chorioamnionitis (14%)	7/51	29% (0 to 62.04) a Low	95% (89.30 to 100) a High	6.29 (1.05 to 37.66) b Moderately useful	0.75 (0.47 to 1.20) b Not useful	Low
White blood cell count >20, 000 cells/mm3 measured on admission or 24-48 hours prior to birth												
1 study (Garite and Freeman 1982)	case series	Serious6	no serious inconsistency	serious indirectness 7	no serious imprecision	Clinical chorioamnionitis (15%)	36/237	16.7% (0 to 29.04)b Low	97.5% (91 to 100)b High	6.70 (2.16 to 21.0)b Moderately useful	0.85 (0.74 to 0.99)b Not useful	Low

** Timing of measurement not reported/unclear*
1 Possible selection bias - unclear whether consecutive women were included in the study
2 Selection bias – only women who delivered within 72 hours of amniocentesis were analysed and unclear whether consecutive women were included in the study
3 Evidence was downgraded by 2 due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%).
4 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5)
5 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5)

1 *6 Possible selection bias - unclear from report how women were selected for inclusion in original trial*
2 *7 A proportion of women reported to be in labour (≤44/251)1*

3 **Table 29: GRADE profile for predictive accuracy of fetal heart rate for identifying infection**

Quality assessment								Measures of diagnostic accuracy				
Number. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Outcome and prevalence (type of infection)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Quality
Fetal heart rate abnormality – all definitions												
Overall summary of findings from 5 studies	case series	Serious[1,2]	serious [3]	no indirectness	Serious[4]	All infectious morbidity (intra-amniotic infection, neonatal sepsis and presumed neonatal sepsis, neonatal pneumonia, clinical chorioamnionitis, intrauterine infection) (7% to 63%)		Range: 8% to 60.0% Low	Range: 41.33% to 100% Low to high	Range: 0.85 to infinity Not useful to very useful	Range: 0.44 to 1.00 Not useful to moderately useful	Very low to moderate
Abnormal antenatal CTG – last test before birth												
1 study (Lewis 1999)	case series	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	Total infectious morbidity (intra-amniotic infection, neonatal sepsis and presumed neonatal sepsis) (33%)	23/69	39.1% (16.93 to 65.08)[a] Low	82.6% (69.43 to 90.57)[b] Moderate	2.25 (1.00 to 5.06)c Not useful	0.74 (0.52 to 1.00)c Not useful	High
1 study (Del Valle 1992)	case series	Serious[1]	no serious inconsistency	no serious indirectness	serious[4]	Neonatal infection (sepsis and pneumonia) (7%)	5/68	60.0% (32.53 to 84.13)[c] Low	90.5% (91.99 to 100)[c] High	6.30 (2.22 to 18.0)[c] Moderately useful	0.44 (0.15 to 1.00)[c] Moderately useful	Low
1 study (Del Valle 1992)	case series	Serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	Clinical chorioamnionitis (15%)	10/68	30.0%(1.60 to 58.40)[c] Low	89.66% (81.82 to 97.49) c Moderate	2.90 (0.86 to 9.75) [c] Not useful	0.78 (0.52 to 1.00) c Not useful	Moderate
Non-reactive CTG*												

Quality assessment								Measures of diagnostic accuracy				
Number. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Outcome and prevalence (type of infection)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Quality
1 study (Caroll 1995)	case series	Serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	Intrauterine infection (positive fetal blood culture) (18%)	14/89	50% (23.81 to 76.18) [c] Low	41.33% (30.19 to 52.48) [c] Low	0.85 (0.48 to 1.49) [c] Not useful	1.00 (0.67 to 1.00) [c] Not useful	Moderate
Fetal heart rate >160 bpm*												
1 study (Ismail 1985)	case series	Serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	Clinical chorioamnionitis (18%)	18/100	22% (3.02 to 41.43)[a] Low	97% (94.22 to 100)[a] High	9.11 (1.80 to 45.99)c Moderately useful	0.79 (0.62 to 1.00)[c] Not useful	Moderate
1 study (Ismail 1985)	case series	Serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	Histological chorioamnionitis (63%)	63/100	8% (1.26 to 14.61)[a] Low	97% (92.07 to 100)a High	2.94 (0.36 to 24.18)[c] Not useful	0.95 (0.86 to 1.00)[c] Not useful	Moderate
Fetal heart rate >170 bpm on admission												
1 study (Garite and Freeman 1982)	case series	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	Clinical chorioamnionitis (15%)	36/237	13.9% (0 to 21.04)c Low	100% (100 to 100)c High	NC/infinity Very useful	0.86 (0.75 to 0.98) Not useful	Moderate

1 **Timing of measurement not reported/unclear*
2 *1 Possible selection bias - unclear whether consecutive women were included in the study*
3 *2 Possible selection bias - unclear from report how women were selected for inclusion in original trial*
4 *3 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%)*
5 *4 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5)*

6 **Table 30: GRADE profile for predictive accuracy of maternal temperature for identifying infection**

Quality assessment								Measures of diagnostic accuracy				
Number. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Outcome and prevalence (type of infection)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Quality
Maternal temperature ≥38°C												
1 study (Ismail 1985)	case series	serious1	no serious inconsistency	no serious indirectness	no serious imprecision	Histological chorioamnionitis (63%)		17% (8.09 to 26.83) Low	97% (92.07 to 100) High	6.46 (0.87 to 1.03) Moderately useful	0.85 (0.75 to 0.96) Not useful	Moderate

7 *1 Possible selection bias - unclear whether consecutive women were included in the study*

Additional data

Serial CRP measurements

Four observational studies reported additional data for serial CRP measurements and the timing of CRP elevation in relation to infection. One observational study (n=147) reported that a constant rise of ≥ 12mg/dL/day was observed more frequently in women with chorioamnionitis than in women without chorioamnionitis both 2 days before giving birth and 12 hours before giving birth.

One observational study (n=54) found that, when measured serially, CRP levels rose before white blood cell count (no further details are given). Another observational study (n=51) also found that elevated CRP appeared to be the earliest sign of chorioamnionitis. The normal within-day percentage coefficient of variation for CRP values obtained at a 6-8 hour interval was 4.8% +/- 0.9% and the between days coefficient was 12.8% +/- 2.2% (values ranged from <0.60 to 1.75 mg/dL). An increase of over 30% in the day-to-day coefficient of variation was defined as abnormal based on data collected. A 4th observational study (n=55) also concluded that CRP elevation preceded birth or clinical infection by "several days". The study reported that 12 of 13 women with consecutive estimations > 2 mg/dL were found to have histological chorioamnionitis.

Additive effect of tests

One randomised trial (n=73) reported the predictive value for CRP level, maternal temperature and white blood cell count alone and taken together. Using a logistic regression model to develop a receiver-operator characteristics curve the area under the curve (AUC) for CRP alone (AUC = 0.566), CRP plus temperature at the onset of labour (AUC = 0.696) and for CRP plus temperature plus white blood cell count (AUC = 0.697), indicating that each is 57%, 70% and 70% predictive of chorioamnionitis respectively.

7.1.4 Evidence statements

C-reactive protein

The majority of studies (ranging from 24 to 147 women) found that C-reactive protein at cut-offs ranging from ≥0.7 mg/dL to >5 mg/dL is not a useful predictor of either clinical or histological chorioamnionitis (positive likelihood ratios judged to be not useful, and mainly low sensitivity). One observational study (n=51) found the positive likelihood ratio for C-reactive protein >3.5mg/dL to be moderately useful and another study found the positive likelihood ratio for C-reactive protein >1.25 mg/dL to be very useful.

The negative likelihood ratios were either not useful or moderately useful. Negative likelihood ratios were judged to be moderately useful at cut-offs >1.2 mg/dL (in 2 out of 3 studies n=213), >1.25 mg/dL (one study n=52), ≥ 2mg/dL (in 4 out of 8 studies n=175) and >4 mg/dL (in one of 2 studies n=147). Specificity was found to be mainly low or moderate.

The evidence was generally of low quality.

White blood cell count

The evidence from 4 observational studies with over 400 women with PPROM predominately found that white blood cell count at cut-offs ranging from >12,500 cells/mm3 to ≥13,500 cells/mm3 is not a useful predictor of either clinical or histological chorioamnionitis (positive likelihood ratios judged to be not useful, moderate or low sensitivity and specificity). One observational study (n=51) found the positive likelihood ratio for white blood cell count >16,000 to be moderately useful. The same observational study found the negative likelihood

ratio for white blood cell count >12500 cells/mm3 to be moderately useful. One observational study (n=237) found the positive likelihood ratio for white blood cell count >20,000 to be moderately useful, and the negative likelihood ratio not useful in predicting clinical chorioamnionitis

The evidence was of moderate to low quality.

Fetal heart rate

Two observational studies (n=158) found that an abnormal CTG result is not a useful predictor of neonatal infection (positive and negative likelihood ratios judged to be not useful, and low to moderate sensitivity and specificity). One observational study (n=68) found a moderately useful positive likelihood ratio and negative likelihood ratio for an abnormal CTG result in predicting neonatal infection.

One observational study (n=68) found that an abnormal CTG result is not a useful predictor of clinical chorioamnionitis (positive and negative likelihood ratios judged to be not useful, and low to moderate sensitivity and specificity).

One observational study (n=100) found that a fetal heart rate >160 bpm was not a useful predictor of histological chorioamnionitis. The same study found that a fetal heart rate >160 bpm for predicting clinical chorioamnionitis had a moderately useful positive likelihood ratio and a not useful negative likelihood ratio (fetal heart rate >160 bpm had a low sensitivity and high specificity for predicting both clinical and histological chorioamnionitis). One observational study (n=237) found that a fetal heart rate >170 bpm was not a useful predictor of clinical chorioamnionitis (positive and negative likelihood ratios not useful).

The evidence was generally of moderate quality.

Maternal temperature

Evidence from one observational study (n=100) found that a raised maternal temperature had a moderately useful likelihood ratio for predicting histological chorioamnionitis but a not useful negative likelihood ratio. The specificity of raised maternal temperature was found to be high but the sensitivity was low for predicting histological chorioamnionitis. The evidence from this study was of moderate quality.

7.1.5 Health economics profile

A search was undertaken for health economic evidence on the diagnostic value of temperature, pulse, white cell count, C-reactive protein and cardiotocography (CTG) to identify infection in women with preterm pre-labour rupture of membranes (P-PROM). A total of 34 articles were identified by the search. After reviewing titles and abstracts, 2 full papers were obtained but they were both excluded as they did not evaluate the relevant investigations. This question was not prioritised for health economic analysis as they were thought by the Committee to be low cost investigations and with an expectation that they would not be found to be particularly effective.

7.1.6 Evidence to recommendations

7.1.6.1 Relative value placed on the outcomes considered

The Guideline Committee have considered all the properties of diagnostic accuracy measurements for decision-making in this topic: sensitivity, specificity, positive and negative likelihood ratio. The Committee considered the relative importance of having a high false positive and high false negative result from identification of different types of infections for women with P-PROM. Because ascending infection from mother to the fetus potentially

causes perinatal mortality or severe neonatal morbidity, the Committee would accept a test or series of tests that had some false positives as long as the false negative rate was very low. Conversely, a high false positive rate would result in an increase in hospital admissions and use of antibiotics.

7.1.6.2 Consideration of clinical benefits and harms

The Guideline Committee discussed the benefits of using different tests (CRP, white blood cell count, fetal heart rate and maternal temperature) in isolation for identification of different types of infection for women with P-PROM, The evidence of included studies did not show that any of these tests were helpful in identification of different types of infections (clinical and historical chorioamnionitis, funisitis and neonatal infection) for this group of women at risk of preterm birth that would be relevant in clinical practice.

However, the negative likelihood ratios of these tests (for example C-reactive protein ≤ 1.21.5 or 2 mg/dL measured on admission) were found to be better in ruling out women with P-PROM with no indications of infections. The Committee considered that these tests may provide useful reassurance against a diagnosis of infection when consistent with the clinical picture, rather than helping to make a positive diagnosis. Different thresholds were used per test in each study. In majority these thresholds were fairly low compared to the ones used for decision-making in clinical practice and the single high values has not been investigated in the evidence so the Committee considered these limitations in the interpretation of results.

The Committee discussed that a CRP threshold of 2 mg/dl and white blood cell count of 12,500 cells/mm3 would be the most common thresholds used in clinical practice above which a result would be considered to be abnormal. The evidence did not support these thresholds as useful markers of infection. White blood cell counts are usually higher than normal in pregnant women. Therefore, they concluded that when using these tests they should be only combined with clinical assessment to diagnose infections for women with PPROM.

7.1.6.3 Consideration of health benefits and resource uses

Costs of CRP of £1.03 have recently been cited (http://qir.bmj.com/content/2/2/u204012.w1749.full.pdf) although this may reflect just the laboratory costs and not the costs of obtaining the sample, although the Committee suggested a cost of around £5 per test. The other investigations are also relatively inexpensive requiring a very small amount of health care professional time

7.1.6.4 Quality of evidence

The majority of evidence was moderate to low quality as some of the included studies were small and there was serious risk of bias. Measurements of sensitivity and specificity requires a clinically relevant threshold to be defined but the evidence was presented based on the selected thresholds by the authors that were usually lower than the one used in clinical practice. There was a high variability between the included studies in the selection of population, definition of diagnostic tools and the measurements reported but not unusual for this type of studies (diagnostic). The Guideline Committee discussed the generalization of results which reported timing of outcomes measured since birth were less useful because they do not reflect the real clinical scenario. The age of included studies and the lack of representativeness of entire population of women with PPROM were limitations in the study design of evidence included in this section. The common use of antibiotics across the included studies was also a factor that may have distorted the direction of results identified.

The Committee has also discussed the limitations of evidence applicability due to the way outcomes were reported as a single cut off point was used instead of change infection rates across time.

7.1.6.5 Other considerations

Sepsis was the commonest direct cause of maternal death in the most recent Confidential Enquiry on Maternal and Child Health. Bacterial infection with Group B Streptococcal infection or gram negative bacteria can spread very rapidly to the fetus and can cause severe disease in mother and fetus/baby which may not be amenable to antibiotic therapy. It may not be clinically appropriate to observe and wait for test results and urgent antibiotics and delivery may be the more appropriate action.

These recommendations are based on both the clinical interpretation of evidence and on Guideline's clinical expert opinion.

7.1.7 Key conclusions.

No consistent findings were found for the different tests of identifying different types of infections for women with P-PROM.

7.1.8 Recommendations

15. **Use a combination of clinical assessment and biomedical tests to diagnose intrauterine infection in women with P-PROM.**

16. **Do not use any of the following in isolation to confirm or exclude intrauterine infection in women with P-PROM:**
    - a single test of C-reactive protein
    - white blood cell count
    - cardiotocography.

17. **If the results of the clinical assessment or any of the biomedical tests are not consistent with each other, continue to observe the woman and consider repeating the tests.**

7.1.9 Research recommendations

Research question	3. What is the diagnostic accuracy of serial C-reactive protein testing to identify chorioamnionitis in women with P-PROM?
Why this is needed	
Importance to 'patients' or the population	Identifying infection in women with P-PROM is needed to allow appropriate management. Early diagnosis of infection allows consideration of therapeutic strategies (including antibiotics and/or early birth). Effective treatment of infection is particularly important given that sepsis is a common direct cause of maternal death. There is currently limited evidence that serial C-reactive protein testing might be useful, but the Committee is aware that this strategy is in common practice. Evidence from diagnostic studies is needed about the accuracy of serial C-reactive protein testing for identifying chorioamnionitis, which is one of the most common and serious infective complications of P-PROM.
Relevance to NICE guidance	Medium; the research would inform future updates of the guideline and addresses a commonly-encountered clinical scenario of clinical importance.
Relevance to the NHS	The Guideline Committee was aware that serial measurement of C-reactive protein is in common practice for monitoring women with P-PROM. Whilst not an expensive test, if shown to be unhelpful there would be cost savings.

Research question	3. What is the diagnostic accuracy of serial C-reactive protein testing to identify chorioamnionitis in women with P-PROM?
National priorities	NHS Outcomes Framework #1: Preventing people from dying prematurely
Current evidence base	Although limited evidence showed that serial C-reactive protein testing might be useful, the Guideline Committee was aware that this strategy is in common practice. Evidence is needed on its effectiveness in identifying chorioamnionitis, one of the most common and serious infective complications of P-PROM.
Equality	The population is defined by gestational age.
Feasibility	The research is feasible and the intervention is low-cost. There are no ethical issues other than those usually pertaining to perinatal research.
Other comments	None

8 'Rescue' cervical cerclage

8.1 Introduction

Cervical cerclage, also known as a cervical stitch, is a treatment used to prevent the cervix opening too early causing either a late miscarriage or preterm birth. Cerclage may be performed as a prophylactic measure where there is a history that increases the risk of spontaneous second-trimester loss or preterm delivery and/or cervical shortening seen on ultrasound; this type of cerclage is considered in Chapter 4 . It can also be performed as a salvage measure ('rescue' or non-prophylactic cerclage) when women have presented with premature cervical dilation, often with exposed fetal membranes, and in some cases where the membranes have prolapsed into the vagina. 'Rescue' cerclage is not a common procedure, and is most often performed in the mid-trimester around the time of viability. Preventing or delaying preterm birth at this gestation might have significant benefit in terms of reducing mid-trimester loss and avoiding the consequences of extreme prematurity. However, the procedure carries risks, and there is uncertainty about which women are most likely to benefit. This chapter considers only 'rescue' or non-prophylactic cervical cerclage.

8.1.1 Review question

What is the clinical effectiveness of non-prophylactic 'rescue' cervical cerclage in preventing preterm birth for women in suspected preterm labour?

8.1.2 Description of included studies

Six studies are included in the review for the role of 'rescue' cervical cerclage of this review question (Althuisius 2003; Curti 2012; Daskalakis 2006; Olatunbosun 1995; Stupin 2008; Aoki 2013). As we only identified one RCT to match our protocol (Althuisius 2003), we also considered comparative cohort studies, either prospective (Daskalakis 2006; Olatunbosun 1995) or retrospective (Curti 2012; Stupin 2008; Aoki 2013).

All women included in the studies were at risk of preterm birth. Women who did not have 'rescue' cervical cerclage were confined to bed rest. The additional use of supportive treatment with tocolysis, antibiotics, corticosteroids and low molecular weight heparins varied between the studies.

The lowest gestational age of women included in the studies reviewed was 16 weeks while the highest gestation was 27 weeks. Mean gestation (SD) at the time of emergency cerclage was reported only in 3 out of the 6 included studies and was around the age of 22 weeks (22.4 (1.7) weeks, 22.4 (2.1) weeks and 22.2 (3.3) weeks).

8.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The following GRADE profile is presented:

- Table 31: GRADE profile for comparison of 'rescue' cervical cerclage versus no cerclage

For quantitative outcomes such as the pregnancy prolongation that were presented by either parametric (mean, SD) or non-parametric (median, range) measures, the GRADE profile includes information only on results from studies with parametric measures that could be used to calculate the absolute effects.

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

1 **Table 31: GRADE profile for comparison of 'rescue' cervical cerclage versus no cerclage**

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	"Rescue" Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal death (including any intrauterine death and neonatal death up to 7 days postpartum)											
1 study (Stupin 2008)	cohort study	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	None	5/89 (5.6%)	13/72 (18.1%)	RR 0.31 (0.12 to 0.83)	125 fewer per 1000 (from 31 fewer to 159 fewer)	Very Low
Neonatal survival											
1 study (Althuisius 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[3]	serious[2]	none	9/16 (56.3%)	4/14 (28.6%)	RR 1.97 (0.77 to 5.01)	277 more per 1000 (from 66 fewer to 1000 more)	Low
1 study (Curti 2012)	cohort study	serious[4]	no serious inconsistency	no serious indirectness	serious[2]	none	30/37 (81.1%)	8/15 (53.3%)	RR 1.52 (0.92 to 2.5)	277 more per 1000 (from 43 fewer to 800 more)	Very Low
1 study (Olatunbosun 1995)	cohort study	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	none	17/22 (77.3%)	9/15 (60%)	RR 1.29 (0.80 to 2.06)	174 more per 1000 (from 120 fewer to 636 more)	Very Low
1 study (Daskalakis 2006)	cohort study	serious[5]	no serious inconsistency	no serious indirectness	serious[2]	none	24/25 (96%)	4/7 (57.1%)	RR 1.68 (0.88 to 3.21)	389 more per 1000 (from 69 fewer to 1000 more)	Very Low
Serious neonatal morbidity (defined as admission to neonatal intensive care unit and/or neonatal deaths)											
1 study (Althuisius 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[3]	serious[2]	none	10/16 (62.5%)	14/14 (100%)	RR 0.64 (0.43 to 0.94)	360 fewer per 1000 (from 60 fewer to 570 fewer)	Low
Interval between study entry and delivery days) (better indicated by higher value)											
1 study (Althuisius 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[3]	very serious[6]	None	54 (SD47)	20 (SD 28)	-	MD 34 higher (3.11 higher to 64.89 higher)	Very Low
1 study (Daskalakis 2006)	cohort study	serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	none	62 (SD27)	22 (SD18)	-	MD 40 higher (26.97 higher to 53.03 higher)	Low
Preterm birth (not defined)											
1 study (Aoki 2013)	cohort studies	very serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	12/15 (80%)	20/20 (100%)	RR 0.8 (0.61 to 1.05)	200 fewer per 1000 (from 390	Very Low

Quality assessment							Number of women		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	"Rescue" Cerclage	No cerclage	Relative (95% CI)	Absolute (95% CI)	
										fewer to 50 more)	
Preterm birth between 22^{+0} to 27^{+6} weeks											
1 study (Aoki 2013)	cohort studies	very serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	3/15 (20%)	16/20 (80%)	RR 0.25 (0.09 to 0.7)	600 fewer per 1000 (from 240 fewer to 728 fewer)	Very Low
Preterm birth before 34 weeks											
1 study (Althuisius 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[3]	serious[2]	none	7/13 (53.8%)	10/10 (100%)	RR 0.56 (0.34 to 0.93)	440 fewer per 1000 (from 70 fewer to 660 fewer)	Low
Preterm birth before 32 weeks											
1 study (Daskalakis 2006)	cohort study	serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	None	9/29 (31.0%)	16/17 (94.1%)	RR 0.33 (0.19 to 0.57)	631 fewer per 1000 (from 405 fewer to 762 fewer)	Low
Maternal side effects (include cervical laceration and cervical dystocia due to scar tissue preventing cervical dilation)											
1 study (Daskalakis 2006)	cohort study	serious[5]	no serious inconsistency	no serious indirectness	very serious[6]	None	4/29 (8.2%)	0/17 (0%)	RR 5.4 (0.31 to 94.55)	NC	Very low

CI: confidence interval; RR: relative risk; NC: not calculable; MD: mean difference

1. Majority of evidence has only 1 indirect aspect of population: 19% of women had a multiple pregnancy (20% in cerclage group, 18% in no cerclage group)

2. Evidence was downgraded by 1 due to serious imprecision as 95% CI crossed one default MID

3. Majority of evidence has only 1 indirect aspect of population: 30% of women had a multiple pregnancy (23% in intervention group, 40% in control group)

4. Study states women were allocated to treatment but it was not clear how this allocation was made, Unclear care protocol for women in the no cerclage group

5. Likely variation in additional treatments (e.g. tocolysis, antibiotics, corticosteroids) between groups

6. Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs

7. High selection, performance, attrition and detection bias

8.1.4 Evidence statements

Low and very low quality evidence from one small RCT (n=30) and 3 cohort studies (number of participants ranged from 32 to 52) found no significant difference for the outcome of neonatal survival in babies whose mothers had 'rescue' cerclage compared with those who did not have cerclage.

Low and very low quality evidence from 3 individual cohorts and 1 RCT (number of participants ranged from 23 to 161) showed significantly fewer perinatal deaths, preterm births before 22^{+0} to 27^{+6}, 32 and 34 weeks in women who had 'rescue' cerclage compared with women who did not. Very low quality evidence from the small RCT and one cohort study showed that the interval between the study entry and birth was longer in women who had 'rescue' cerclage compared with women who did not. In addition, low quality evidence from the same RCT trial showed significantly lower risk of serious neonatal morbidity (defined as admission to neonatal intensive care unit and/or neonatal deaths) in babies whose mothers had cerclage compared with women who did not have cerclage.

The outcomes of neonatal survival and maternal side effects were not found significantly different between the 2 groups (rescue cerclage and no cerclage) based on low to very low quality evidence from a small RCT and individual cohort studies.

8.1.5 Health economics profile

A single search was undertaken for health economic evidence on prophylactic cervical cerclage to prevent preterm labour in women considered to be at risk of preterm labour and birth and rescue cervical cerclage in preventing preterm birth in women in suspected preterm labour. A total of 60 articles were identified by the search. After reviewing titles and abstracts, 3 papers were obtained. These studies were all excluded because they were not economic evaluations or were published conference abstracts. Therefore, no relevant economic evidence was identified for this question.

This question was not identified as a priority for health economic analysis as the Committee reflected that the intervention would only be relevant for a very small proportion of the patient population.

8.1.6 Evidence to recommendations

8.1.6.1 Relative value placed on the outcomes considered

In terms of neonatal outcomes, the Committee considered neonatal mortality as critical for this review concluding that any mortality up to 1 year could be reported as a single outcome. Tthe Committee included early neonatal survival because only the first week was available for analysis.

Given that the purpose of this intervention is to delay birth, preterm birth and the interval between the procedure and delivery were both prioritised as important outcomes. The Committee included neonatal sepsis, chronic lung disease and bronchopulmonary dysplasia as potential associated adverse events. The Committee also considered long-term infant neurodevelopmental outcomes such as neurodevelopmental disability because these are common adverse events associated with preterm birth and any reduction in these outcomes would be a significant indicator of the effectiveness of the intervention. However, given the absence of available data, the Committee agreed that serious neonatal morbidity could be considered as a surrogate for the neonatal adverse events outcomes selected originally.

In term of maternal outcomes, maternal mortality and maternal adverse effects, were prioritised for this review question including infection requiring intervention and cervical

trauma requiring repair because rescue cerclage is a difficult procedure and there is a potential risk of such events occurring. In addition, the Committee felt that evidence regarding maternal emotional/psychological impact should be also assessed due to the invasive nature of the procedure and considering the stressful circumstances under which it might be conducted (when urgent treatment is required).

8.1.6.2 Consideration of clinical benefits and harms

The Committee recognised that both the randomised and observational evidence supported well a benefit of 'rescue' cerclage for the outcomes of reducing serious neonatal morbidity, preterm birth below 27, 32 and 34 weeks and increasing the interval between intervention and delivery. The Committee had serious concerns regarding the lack of data for the outcome of neonatal sepsis as, based on their clinical experience, this was likely to be a significant complication associated with the procedure. They concluded that information about the risks of rescue cerclage, as well as its potential aims and benefits, should be communicated to the women and their family members or carers as appropriate.

The Committee's clinical opinion was that benefit of performing 'rescue' cerclage beyond 32 gestational weeks would be limited and may not outweigh the potential harms. They agreed that the recommendation to perform 'rescue' cerclage for women with a dilated cervix and exposed, unruptured membranes should reflect both the quality of reviewed evidence (low to very low) and the gestational age of women included in these studies (of gestational age between 16 to 27 weeks).

They were also aware that rescue cerclage can be a technically difficult procedure to perform that requires specialist skills and expertise to mitigate the risks of maternal or neonatal adverse events.

For these reasons, the Committee recommended that the decision for rescue cerclage must be made only after discussion with a consultant obstetrician. The decision should take account of the woman's gestation and her own stated wishes after a full discussion.The Committee decided that, in their clinical opinion, rescue cerclage would cause harm to women with signs of infection, active vaginal bleeding or uterine contractions and so they decided upon a strong recommendation of not offering 'rescue' cerclage to these groups of women.

8.1.6.3 Consideration of health benefits and resource uses

The Committee felt that rescue cerclage was likely to be an expensive intervention due to the setting in which it is delivered and the expertise required of the health-care professionals providing the care.

They also felt that although the evidence shows a reduction in preterm birth, perinatal death and neonatal morbidity, it was of low quality and the chance of poor outcomes is quite high despite emergency cerclage. They acknowledged that the management of preterm birth and the associated outcomes is extremely costly – both financially and in terms of parental anxiety - and therefore, if the intervention delayed birth beyond key gestational milestones, then the initial costs incurred would be likely to be offset by large cost savings downstream.

Furthermore, the Committee felt that the overall cost impact for the health services would be small as rescue cerclage would only be an appropriate intervention for a small proportion of the patient population.

8.1.6.4 Quality of evidence

The majority of the evidence included was of low to very low quality. Only a small RCT of 30 women was included in the evidence base whereas the rest of the evidence is coming from cohort studies. Indirectness and imprecision were the main areas affected in studies quality

assessment. More specifically, some of the cohorts included were mixed populations in terms of use of additional tocolytics (or other adjunctive treatment) and the inclusion of twin pregnancies. Potential bias in the single RCT arose from the imbalanced allocation of women with multiple pregnancies to treatment groups which was of particular concern for the Committee. The sample size of the included studies was relatively small thus restricting the confidence in the estimates of effects.

8.1.6.5 Other considerations

The recommendations were based on both the interpretation of clinical evidence reviewed and on Guideline Committee's expert opinion.

The Committee noted that this review did not set out to compare the different types of stitches that could be used and the conclusions of the findings may be limited under these restrictions. They were also aware that there is an existing RCOG Green Top Guideline No 60 that provides more detail on how to perform this intervention.

The Committee also noted that it would be important to take into account of the effectiveness of other possible interventions (such as magnesium sulfate for neuroprotection) when considering whether to offer rescue cerclage at lower gestational ages.

8.1.7 Key conclusions

In light of all their reservations about the evidence on the effectiveness of rescue cerclage, the Committee decided that a decision should be made with caution regarding its application.

For these reasons, the Committee felt that women should be clearly informed about the potential risks and benefits of the procedure. It was noted that a specific definition of preterm labour had been used in some of the included studies (dilated cervix with exposed fetal membranes) and that this should be reflected in the recommendations in terms of indications and contraindications for the use of rescue cerclage. They also noted that the gestation at which rescue cerclage was undertaken was an important consideration as was the skill/experience of the practitioner performing the procedure. The importance of the woman and her partner having confidence and trust in the obstetrician was also highlighted. The Committee concluded that whether or not to insert a rescue cervical suture was a complex judgement that should be undertaken on an individualised basis with full involvement of the woman.

The Committee also took the view that there should be some obligation to collect data about outcomes of its use in a national registry.

8.1.8 Recommendations

18. **Consider 'rescue' cervical cerclage for women between 16^{+0} and 27^{+6} weeks of pregnancy with a dilated cervix and exposed, unruptured fetal membranes.**

19. **Do not offer 'rescue' cervical cerclage to women with signs of infection, active vaginal bleeding or uterine contractions.**

20. **When deciding whether to offer 'rescue' cervical cerclage:**
    - take into account gestational age and the extent of cervical dilatation
    - discuss with a consultant obstetrician and consultant paediatrician.

21. **Explain to women for whom 'rescue' cervical cerclage is being considered (and their family members or carers as appropriate):**

- about the risks of the procedure
- that it aims to delay the birth, and so increase the likelihood of the baby surviving and of reducing serious neonatal morbidity.

8.1.9 Research recommendations

Research question	4. What is the clinical effectiveness of 'rescue' cerclage in improving outcomes for women at risk of preterm birth?
Why this is needed	
Importance to 'patients' or the population	There is some evidence from randomised studies that 'rescue' cerclage might be effective in improving neonatal outcomes in women with a dilated cervix and exposed, unruptured fetal membranes. However, there is uncertainty about the magnitude of this effect. The full consequences of this strategy and the subgroups of women at risk of preterm labour who might particularly benefit are not known. A randomised controlled trial would best address this question, but a national registry of the most critical outcomes (neonatal mortality and morbidity, maternal morbidity) could also be considered for women who did not want to participate in a randomised trial but who opted for 'rescue' cerclage.
Relevance to NICE guidance	The importance is high because rescue cerclage is widely used yet its evidence base is of relatively low quality, and it is likely that more high quality evidence would refine or change current recommendations.
Relevance to the NHS	If rescue cerclage is effective in delaying delivery, and this benefit is found to outweigh any harms, then it will be important to define the groups of women who may benefit from this treatment. If effective it has potential to reduce morbidity and mortality, healthcare resources and costs expended on the care of very preterm babies.
National priorities	NHS Outcomes Framework 2014-5, #1: Preventing people from dying prematurely
Current evidence base	The Guideline Committee recognised that the current evidence base was of low quality and many important questions remain unanswered. There is some evidence that 'rescue' cerclage might be effective in improving outcomes in women with a dilated cervix and exposed, unruptured fetal membranes. However, there was uncertainty in the magnitude of this effect and the full consequences and the subgroups of women at risk of preterm labour who might particularly benefit are unknown.
Equality	No
Feasibility	There is no reason in principle why further trials should not be carried out to address current uncertainties. The Committee felt that a randomised trial would provide the best evidence, but that a registry collection of outcomes (neonatal mortality and morbidity, maternal morbidity) could be considered for women who did not want to participate in a randomised trial but who opted for rescue cerclage. The ethical issues are not in principle different from those affecting other perinatal trials.
Other comments	Trials would have to be carried out in centres with neonatal facilities equipped to care for very preterm babies.

9 Diagnosing preterm labour in women with intact membranes

9.1 Introduction

It can often be unclear whether symptoms of preterm labour will result in progression to established labour and birth. Symptoms described by women may be associated with labour such as painful contractions, but could be non-specific for example low back or abdominal pain. Sometimes symptoms can occur but then settle, allowing the pregnancy to continue towards term. Investigations performed clinically (such as digital vaginal examination of the cervix), biochemically, or using ultrasound may help to distinguish women in preterm labour from those who are not. Women in whom preterm labour is correctly identified may then benefit from clinical management to try to delay birth and/or improve neonatal outcomes, whereas those who are not in preterm labour can be reassured, and further intervention is not necessary.

9.1.1 Review question

What is the diagnostic accuracy of the following (alone or in combination) in women with intact membranes to identify preterm labour leading to preterm birth:

- **clinical assessment (such as symptoms expressed by women, strength and frequency of contractions, findings on vaginal examination)**
- **biochemical testing for markers for preterm labour namely cervicovaginal fetal fibronectin and IGF-BP1 insulin-like growth factor binding protein 1**
- **cervical ultrasound features (such as cervical length and funnelling)?**

9.1.2 Description of included studies

38 prospective cohort studies were included in this review (Azlin 2010, Bagga 2010, Bartnicki 1996, Benattar 1997, Botsis 2006, Brik 2010, Burwick 2011, Danti 2011, Demirci 2009, Diaz 2009, Eroglu 2004, Giles 2000, Gomez 2005, Gramellini 2007, Holst 2006, Iams 1995, Kwek 2004, LaShay 2000, Lembet 2002, Lukes 1997, Malak 1996, McKenna 1999, Palacio 2007, Sakai 2003 Schmitz 2006, Schmitz 2008, Schreyer 1989, Senden 1996, Skoll 2006, Sotiriadis 2010, Swamy 2005, Tanir 2008, Tanir 2009, Tekesin 2005, Ting 2007, Tsoi 2006, Tsoi 2005; Van Baaren 2014).

All studies included women with signs and symptoms of preterm labour who had singleton pregnancies and intact membranes except one study (McKenna 1999) that did not specify details for women with multiple pregnancies, one study (Benattar 1997) in which 13% of women had a multiple pregnancy, and one study (Lembet 2002) where reporting of multiple pregnancies is unclear and that may have included some women with PROM.

Originally the Committee decided to consider only studies with women who had not received tocolytics as part of their management plan to delay preterm labour. There were no studies where tocolytics were not used, so the Committee decided to expand the inclusion criteria and include studies where all women had had tocolytics to preserve homogeneity of interpretation of results. However, due to the limited number of such studies (8 studies - Benattar 1997, Kwek 2004, Palacio 2007, Senden 1996, Sotiriadis 2010, Swamy 2005, Tekesin 2005, Ting 2007), it was decided to include studies with a mixed population (i.e. those who had and who had not received tocolytics) and downgrade the quality of this evidence as indirect to the population of interest.

An HTA SR was published in 2009 examining different screening techniques to diagnose preterm birth. This review included RCTs, prospective and retrospective cohort studies.

Diagnostic accuracy results were reported separately for asymptomatic and symptomatic women for some of the tests that were also the focus of this review question (digital vaginal examination, pIGFBP-1, fetal fibronectin and transvaginal ultrasound). Outcomes included preterm birth and time to delivery endpoints. This HTA review was not included in our evidence review as we only considered prospective cohort studies for inclusion in our protocol. In addition outcomes in the HTA were assessed at 7 to 10 days following presentation rather than within 7 days specified in the review protocol of this question. However, the individual studies in the HTA publication were assessed for relevance to this protocol.

An update of the fetal fibronectin section of the HTA review was published recently (HTA 2013) but was excluded from this review again due to differences in outcomes. This recent HTA report also included retrospective studies and populations with multiple pregnancies; it did not report the outcomes of interest to us, namely birth within 48 hours or 7 days of presentation.

9.1.2.1 Diagnosis using clinical examination

The presentation of evidence is presented by diagnostic test. The reference standard was considered to be progression to labour, with labour defined as progressive dilation of the cervix over the subsequent few hours/days of presentation leading to birth.

Bishop score

There were 3 studies (Schmitz 2008, Schreyer 1989, Senden 1996) included in total all of which looked at the accuracy of this diagnostic test. The studies examined different thresholds of Bishop scores to diagnose birth within 48 hours and 7 days.

The population across all studies were symptomatic women who presented between 24-36 gestational weeks.

9.1.2.2 Diagnosis using biochemical tests

Phosphorylated Insulin-like Growth Factor Binding Protein-1 (pIGFBP-1)

There are 7 studies in total for this test; Azlin 2010, Brik 2010, Danti 2011, Eroglu 2004, Kwek 2004, Lembet 2002, Tanir 2009 all of which examined pIGFBP-1 to predict preterm birth within 7 days.

Three of these studies (Brik 2010, Kwek 2004 Lembet 2002) examined pIGFBP-1 to predict preterm birth within 48 hours

Subgroup analysis of women who had cervical length estimates ≤ 30mm was given in 2 studies:

- Danti 2011 (cervical length<20mm; 20-30mm; ≤30mm)
- Azlin 2010 (cervical length<25mm)

Overall, 5 studies included symptomatic women who presented during 22-35 gestational weeks although one study (Lembet 2002) included women over 20 gestational weeks and 3 studies included women up to 36^{+6} weeks (Azlin 2010, Lembet 2002, Tanir 2009).

Four studies were conducted in resource-rich countries: Azlin 2010 was carried out in Malaysia; Brik 2010 in Spain; and Kwek 2004 and Ting 2007 in Singapore. Eroglu 2004, Lembet 2002 and Tanir 2009 were conducted in Turkey.

9.1.2.3 Diagnosis using transvaginal ultrasonography to measure cervical length

Thirteen studies in total are included in this section.

Four studies (Bagga 2010, Gomez 2005, Schmitz 2008, Tsoi 2005) examined different thresholds of cervical length measured by transvaginal ultrasound to diagnose preterm birth within 48 hours.

13 studies (Bagga 2010, Azlin 2010, Botsis, Danti 2011, Demirci 2009, Eroglu 2004, Gomez 2005, Gramellini 2007, Holst 2006, Palacio 2007, Schmitz 2008, Sotiriadis 2010, Tsoi 2005) examined different thresholds of cervical length measured by transvaginal ultrasound to diagnose preterm birth within 7 days.

Three studies also investigated the diagnostic accuracy of cervical length measurement using transvaginal ultrasound in specific subgroups. One study reported the use of different cervical length thresholds to diagnose birth within 7 days in women according to gestational age above or below 32 weeks (Palacio, 2007). A second study also examined the use of change in cervical length 24 hours after admission as a diagnostic tool to predict preterm birth within 7 days (Sotiraides, 2010). A third study described different thresholds of cervical length determined by transvaginal ultrasound to diagnose preterm birth within 48 hours and 7 days in women with a Bishop score between 4 and 7 (Schmitz, 2008).

Overall, most studies included symptomatic women who presented during 24-36 gestational weeks, although 3 studies included women with lower gestational ages (Gramellini 2007 from 20 weeks, Gomez 2005 from 22 weeks and Sotiriadis 2010 from 23 weeks) and one study included women up to 37 weeks (Bagga 2010).

Most studies for diagnosis using transvaginal ultrasonography to measure cervical length were conducted in resource-rich countries.

9.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Data is reported in the GRADE profiles below as follows:

- Single tests:
  - Table 32: GRADE profile for evaluation of a Bishop score to diagnose preterm birth within 48 hours or within 7 days
  - Table 33: GRADE profile for evaluation of pIGFBP-1 to diagnose preterm birth within 48 hours or within 7 days
  - Table 34: GRADE profile for evaluation of fetal fibronectin to diagnose preterm birth within 48 hours or within 7 days
  - Table 35: GRADE profile for evaluation of fetal fibronectin > 50ng/ml before and after cervical examination to diagnose preterm birth within 7 days
  - Table 36: GRADE profile for evaluation of cervical length measured using transvaginal ultrasound to diagnose preterm birth within 48 hours or 7 days
- Combinations of tests:
  - Table 37: GRADE profile for evaluation of Bishop score plus cervical length measured using transvaginal to diagnose pre-term birth within 48 hours or 7 days
  - Table 38: GRADE profile for evaluation of a selective test based on cervical length measured using transvaginal ultrasound plus Bishop score to diagnose birth within 48 hours or 7 days
  - Table 39: GRADE profile for evaluation of fetal fibronectin > 50ng/ml plus Bishop score to diagnose pre-term birth within 7 days

- Table 40: GRADE profile for evaluation of pIGFBP-1 test and cervical length measured using transvaginal ultrasound to diagnose pre-term birth within 7 days
- Table 41: GRADE profile for evaluation of fetal fibronectin > 50ng/ml plus cervical length to diagnose pre-term birth within 48 hours or 7 days

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

Further information regarding the proportion of women with previous preterm births is given in the GRADE profiles.

Table 32: GRADE profile for evaluation of a Bishop score to diagnose preterm birth within 48 hours or within 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Bishop score to diagnose birth within 48 hours												
Bishop score of 4 to 6[a]												
1 (Schreyer 1989)	Prospective cohort	no serious	no serious	very serious[1,2]	Serious[4]	NR	70	69.2% (41.1 to 89.0)	73.7% (67.3 to 78.2)	2.63 (1.27 to 4.09) Not useful	0.42 (0.14 to 0.87) Moderately useful	Very low
Bishop score ≥ 4[b]												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias	no serious	very serious[2,3]	Very serious[5]	NR	395	94.0% (71.0 to 100.0)	43.0% (38.0 to 48.0)	1.66 (1.20 to 1.76) Not useful	0.14 (0.01 to 0.72) Moderately useful	Very low
Bishop score ≥ 8[b]												
1 (Schmitz 2008)	Prospective cohort	no serious	no serious	very serious[2,3]	serious[5]	NR	395	35.0% (14.0 to 62.0)	97.0% (94.0 to 98.0)	12.13 (4.29 to 29.42) Very useful	0.67 (0.44 to 0.87) Not useful	Very low
Bishop score to diagnose birth within 7 days												
Bishop score >2[c]												
1 (Senden 1996)	Prospective cohort	no serious	no serious	Serious[2,6]	Very serious[5]	NR	25	100%	73%	3.10 (0.86 to 3.83) Not useful	0.17 (0.000 to 1.09) Moderately useful	Very low
Bishop score of 4 to 6[a]												
1 (Schreyer 1989)	Prospective cohort	no serious	no serious	very serious[1,2]	Serious[4]	NR	70	68.8% (44.6 to 86.9)	75.9% (68.8 to 81.3)	2.85 (1.43 to 4.64) Not useful	0.41 (0.16 to 0.81) Moderately useful	Very low
Bishop score ≥ 4[b]												
1 (Schmitz 2008)	Prospective cohort	no serious	no serious	very serious[1,2]	Serious[4]	NR	395	97.0% (84.0 to 100.0)	45.0% (39.0 to 50.0)	1.76 (1.46 to 1.82) Not useful	0.07 (0.00 to 0.40) Very useful	Very low
Bishop score ≥ 8[b]												
1 (Schmitz 2008)	Prospective cohort	no serious	no serious	very serious[1,2]	no serious imprecision	NR	395	34.0% (19.0 to 53.0)	98.0% (96.0 to 99.0)	17.83 (6.87 to 47.57) Very useful	0.67 (0.55 to 0.81) Not useful	Low

a The Bishop score comprises scoring for cervical length, consistency, position, dilation and station of head.
b Bishop score was not defined.
c Bishop score defined according to Myerschough P.R.;"Induction of labour" Chap 20 in Munro Kerr's Operative Obstetrics 10th edn. 1982, pub, Bailliere Tindall
1 Tocolysis may have been used in some women: authors state that women received no medication aside from prenatal vitamins and iron however women who were readmitted in actual labour after discharge at 48 hours are stated as having received tocolytic medication.

2 Women included in the study were a mixed population where some received tocoloysis and some did not
3 The proportion of women who received tocolysis was not reported.
4Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5)
5 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)
6 7/25 (28%) women received ritodrine, 8/25 (32%) women received antibiotic therapy and 19/25 (76%) received corticosteroids. Treatment was according to the established practice of administration when considered appropriate

Table 33: GRADE profile for evaluation of pIGFBP-1 to diagnose preterm birth within 48 hours or within 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
pIGFBP-1 test to diagnose birth within 48 hours												
1 (Brik 2010)	Prospective cohort	serious[1]	no serious inconsistency[2]	very serious[3,4]	serious[5]	9.4%	276	73.7%	64.9%	2.1 (1.52 to 2.91) Not useful	0.41 (0.19 to 0.87) Moderately useful	Very low
1 (Kwek 2004)	Prospective cohort	serious[6]	no serious inconsistency[2]	very serious[7]	very serious[8]	NR	42	66.7% (25.5 to 93.8)[a]	61.1% (54.2 to 65.6)[a]	1.71 (0.56 to 2.73)a Not useful	0.54 (0.09 to 1.37) a Not useful	Very low
1 (Lembet 2002)	Prospective cohort	no serious risk of bias[9,]	no serious inconsistency[2]	very serious[3,,10,11]	serious[12]	16%	36	93.3% (72.3 to 99.6) [a]	81.0% (65.9 to 85.5) [a]	4.90 (2.12 to 6.85) a Not useful	0.08 (0.004 to 0.42) [a] Very useful	Very low
1 (Ting 2007)	Prospective cohort	no serious risk of bias[9],	no serious inconsistency[2]	very serious[7]	no serious imprecision	NR	94	100%	74%	3.85 Not useful	NC	Low
pIGFBP-1 test to diagnose birth within 7 days												
1 (Azlin 2010)	Prospective cohort	serious7	no serious inconsistency[1]	serious[3, 14]	very serious [15]	NR	51	80.0% (32.9 to 98.9)	93.5% (88.4 to 95.5)	12.27 (2.83 to 22.16) Very useful	0.21 (0.01 to 0.76) Moderately useful	Very low
1 (Brik 2010)	Prospective cohort	serious[1]	no serious inconsistency[2]	very serious[3,4]	serious[5]	NR	276	73.1%	66.2%	2.16 (1.60 to 2.92) Not useful	0.41 (0.21 to 0.78) Moderately useful	Very low
1 (Eroglu 2007)	Prospective cohort	no serious risk of bias[9]	no serious inconsistency[2]	serious[3,16]	very serious[17]	3.9%	51	83.3% (39.2 to 99.1)	84.4% (78.6 to 86.5) [a]	5.38 (1.83 to 7.37) [a] Moderately useful	0.20 (0.01 to 0.77) [a] Moderately useful	Very low
1 (Kwek 2004)	Prospective cohort	serious[7]	no serious inconsistency[2]	very serious8	very serious[18]	NR	42	83.3% (55.6 to 96.9) [a]	73.3% (62.2 to 78.8) [a]	3.12 (1.48 to 4.56) [a] Not useful	0.23 (0.04 to 0.71) [a] Moderately useful	Very low

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
1 (Lembet 2002)	Prospective cohort	no serious risk of bias[9],	no serious inconsistency[2]	very serious[3,,10,11]	serious12	16%	36	93.8% (74.3 to 99.7) a	85.0% (69.4 to 87.9) a	6.25 (2.43 to 9.71) [a] Moderately useful	0.07 (0.004 to 0.37) [a] Very useful	Very low
1 (Tanir 2009)	Prospective cohort	no serious risk of bias[9]	no serious inconsistency[2]	serious[3,19]	serious[12]	NR	68	93.3% (69.6 to 99.6) [a]	79.2% (72.5 to 81.0) [a]	4.50 (2.53 to 5.25) [a] Not useful	0.08 (0.004 to 0.42 [a] Very useful	Very low
1 (Ting 2007)	Prospective cohort	no serious risk of bias[9],	no serious inconsistency[2]	very serious[7]	no serious imprecision	NR	94	69%	78%	3.13[a]	0,40 [a]	Very low

1 It is unclear whether clinicians were blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics where decided by the attending clinician, may have been influenced by index test results and have affected when birth (the reference standard) occurred
2 Single study analysis.
3 Women included in the study were a mixed population where some received tocoloysis and some did not.
4 Tocolysis (with nifedipine or atosiban) was used in all women who were in established preterm labour but the definition of preterm labour and the proportion of women who received tocolysis are not reported. Steroids (betamethasone) were administered as appropriate.
5 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5)
6 Clinicians were not blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics may have been influenced by index test results and have affected when birth (the reference standard) occurred
7 All women received tocolysis and corticosteroids according to existing clinical protocols
8 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)
9 The primary clinician was blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics was not influenced by index test results.
10 Women with ruptured membranes were not specified as being excluded from the study and the proportion of women with ruptured membranes is not specified
11 8/18 (44%) women who tested positive for pIGFBP-1 and 13/18 (72.2%) of women who tested negative for pIGFBP-1 received IV tocolysis (1st line treatment with magnesium sulfate). This was according to an existing protocol where women with progressive cervical change and regular contractions, despite bed rest and hydration with 500ml Ringer's lactate solution, received tocolysis
12 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from moderately useful (>0.1-0.5) to very useful (0-0.1)
13 Confidence intervals were not calculable
14 12/51 (23.53%) women received tocolysis at the discretion of the attending clinician (further details of the tocolytic used or whether corticosteroids were administered were not reported)
15 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)16 8/14 (66.7%) of women who tested positive for pIGFBP-1 and 8/37 (20.5%) of women who tested negative for pIGFBP-1 received tocolysis (first line treatment with calcium channel blockers) according to an existing protocol where women with progressive cervical change and persistent regular contractions, despite 2 hours bed rest and hydration with 500ml Ringer's lactate solution, received tocolysis. Maternal corticosteroids were given. No tocolytics or maternal steroids were used after 34 weeks gestation. The mean gestational age at enrolment was 29.5 ± 2.6.
17 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)
18 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)

19 Decisions regarding tocolytic and steroid use were made by clinicians. 23/25 (92%) women who tested positive for pIGFBP-1 and 40/43 (93 %) of women who tested negative for pIGFBP-1 received tocolysis. Symptomatic treatment included IV ritodrine hydrochloride or magnesium sulfate. Betamethasone was given twice daily to enhance fetal lung maturation where indicated.

Table 34: GRADE profile for evaluation of fetal fibronectin to diagnose preterm birth within 48 hours or within 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Fetal fibronectin test to diagnose birth within 48 hours												
1 study (Gomez 2005)	Prospective cohort	very serious[1,2]	no serious inconsistency	serious[3]	serious[4]	13%	215	58.8% (34.4 to 80.0) [a]	78.8% (76.7 to 80.6) [a]	2.77 (1.48 to 4.13) [a] Not useful	0.52 (0.25 to 0.86) [a] Not useful	Very low
1 study (LaShay 2000)	Prospective cohort	very serious[1,2]	no serious inconsistency	serious[3]	no serious imprecision	NR	118	75%	88%	6.25a Moderately useful	0.28 a Moderately useful	Very low
Fetal fibronectin test to diagnose birth within 7 days												
1 study (Bartnicki 1995)	Prospective cohort	serious[1]	no serious inconsistency	serious[6]	no serious imprecision	NR	112	100% (19.29 to 100) [a]	70.9% (61.4 to 79.2) [a]	3.44 (2.57 to 4.60) [a] Not useful	0.00	Low
1 study (Benattar 1996)	Prospective cohort	serious[1]	no serious inconsistency	serious[3,7]	no serious imprecision	NR	114	89% (55 to 100)	90 (55 to 100)	8.9 [a] Moderately useful	0.12 [a] Moderately useful	Low
1 study (Burwick 2011)	Prospective cohort	serious[2]	no serious inconsistency	serious[3]	no serious imprecision	28.9 %	52	66.7% (53.5 to 79.9)	78.3% (66.7 to 89.8)	3.0 [a] Not useful	0.43 [a] Moderately useful	Low
1 study (Diaz 2008)	Prospective cohort	serious[2]	no serious inconsistency	serious[3]	serious[4]	30%	170	75% (52.9 to 89.4)[a]	78.2 (70.7 to 84.2) [a]	3.44 (2.36 to 5.01) [a] Not useful	0.32 (0.16 to 0.64) [a] Moderately useful	Very low
1 (Eroglu 2007)	Prospective cohort	no serious risk of bias[9]	no serious inconsistency	serious[6,10]	very serious[11]	3.9%	51	83.3% (38.9 to 99.1) [a]	80.0% (74.1 to 82.1) [a]	4.17 (1.50 to 5.54) [a] Not useful	0.21 (0.01 to 0.82) a Moderately useful	Very low
1 study (Giles 2000)	Prospective cohort	serious[2]	no serious inconsistency	serious[3]	serious[4]	NR	151	68.7% (46.0 to 91.5)a	74.8 (67.5 to 82.1) [a]	2.73 (1.75 to 4.23) [a] Not useful	0.41 (0.20 to 0.87) [a] Moderately useful	Very low
1 study (Gomez 2005)	Prospective cohort	very serious[1,2]	no serious inconsistency	serious[3]	serious[4]	13%	215	64.3% (45.8 to 79.8) [a]	81.8 (79.1 to 84.1) [a]	3.54 (2.19 to 5.03) [a] Not useful	0.44 (0.24 to 0.69) [a] Moderately useful	Very low
1 study (Iams 1994)	Prospective cohort	serious[2]	no serious inconsistency	serious[3,10]	very serious[11]	32%	192	93.0% (66.0 to 99.0) a	82.0% (75.5 to 87.3) a	5.17 (3.66 to 7.30)a	0.09 (0.01 to 0.58)a Very useful	Very low

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
										Moderately useful		
1 study (LaShay 2000)	Prospective cohort	very serious[1,2]	no serious inconsistency	serious[3]	no serious imprecision	NR	118	67% (CI NR)	NR	NR	NR	Very low
1 study (Lukes 1996)	Prospective cohort	no serious risk of bias	no serious inconsistency	serious[6]	serious[8]	21%	763	86.3% (65.0 to 97.0) [a]	82.3% (79.3 to 85.0) [a]	4.89 (3.89 to 6.13) [a] Not useful	0.17 (0.06 to 0.47) [a] Moderately useful	low
1 study (Malak 1996)	Prospective cohort	serious[1]	no serious inconsistency	serious[6]	very serious[12]	19%	141	80% (44.4 to 96.9) [a]	90.2% (82.7 to 95.2)[a]	8.16 (4.2 to 15.9) a Moderately useful	0.22 (0.06 to 0.77) [a] Moderately useful	Very low
1 study (Sakai 2003)	Prospective cohort	no serious risk of bias	no serious inconsistency	very serious[13]	no serious imprecision	9%	185	73.8%	74.2%	2.86 [a] Not useful	0.35 a Moderately useful	low
1 study (Schmitz 2005)	Prospective cohort	serious[1]	no serious inconsistency	serious[6]	very serious[11]	57%	192	93.0% (66.0 to 99.0)[a]	79.0% (74.0 to 83.0)[a]	3.91 (2.96 to 5.17) a Moderately useful	0.22 (0.09 to 0.54) [a] Very useful	Very low
1 study (Senden 1996)	Prospective cohort	no serious risk of bias	no serious inconsistency	very serious[6,14]	very serious[11]	NR	25	100%	86%	5.75 (1.34 to 7.67) [a] Moderately useful	0.15 (0.000 to 0.89) [a] Moderately useful	Very low
1 study (Skoll 2006)	Prospective cohort	serious[1,15]	no serious inconsistency	serious[6]	very serious[11]	16%	149	80.0% (51.4 to 94.7)[a]	85.1% (77.6 to 90.4)[a]	5.36 (3.32 to 8.63) [a] Moderately useful	0.23 (0.08 to 0.64) [a] Moderately useful	Very low
1 study (Swamy 1996)	Prospective cohort	serious[1]	no serious inconsistency	serious[6]	no serious imprecision[5]	NR	404	67%	92%	8.37 [a] Moderately useful	0.35 [a] Moderately useful	low
1 study (Tanir 2007)	Prospective cohort	serious[2]	no serious inconsistency	serious6	no serious imprecision	NR	195	68.6%	84.4%	4.3 (2.1 to 9.8) Not useful	0.3 (0.2 to 0.5) Moderately useful	low
1 study (Tekesin 2005)	Prospective cohort	serious[2]	no serious inconsistency	serious[13]	no serious imprecision	26%	170	81.8% (48.2 to 97.7)	76.7% (69.4 to 83.1)	3.5 Not useful	0.24 Moderately useful	low
1 study (Ting 2007)	Prospective cohort	no serious risk of bias	no serious inconsistency	serious[13]	no serious imprecision	NR	94	56%	76%	2.33 [a] Not useful	0.73 [a] Not useful	low

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
1 study (Tsoi 2006)	Prospective cohort	serious1	no serious inconsistency	serious[3]	serious[8]	NR	195	94.7% (73.0 to 99.7)[a]	61.9% (59.6 to 62.5)[a]	2.49 (1.81 to 2.66)[a] Not useful	0.09 (0.004 to 0.45)[a] Very useful	Very low

CI confidence interval, NR not reported
a Calculated by the NCC-WCH technical team
1 No definition of symptoms of preterm labour
2 Blinding of clinicians to the index test was not reported
3 Whether women received tocolytic therapy was not reported
4 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5)
5 Confidence intervals were not calculable from the data provided
6 Women included in the study were a mixed population where some received tocolysis and some did not
7 n = 15 women with twin pregnancy included
8 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from moderately useful (>0.1-0.5) to very useful (0-0.1)9 The primary clinician was blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics was not influenced by index test results
10 8/14 (57.1%) women who tested positive for fFN and 8/37 (21.6%) women who tested negative for fFN received tocolytic therapy. Tocolytic therapy (with calcium channel blockers) was administered according to an existing protocol where women with progressive cervical change and persistent regular contractions, following 2 hours bed rest and hydration with 500ml Ringer's lactate solution, received tocolysis. Maternal corticosteroids were given. No tocolytics or maternal steroids were used after 34 weeks gestation. The mean gestational age at enrolment was 29.5 weeks ± 2.65
11 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)12 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1
13 All women received tocolysis
14 7/25 (28%) women received ritodrine, 8/25 (32%) women received antibiotic therapy and 19/25 (76%) received corticosteroids. Treatment was according to the established practice of administration when considered appropriate
15 Reason for admission for 30/130 women not reported

Table 35: GRADE profile for evaluation of fetal fibronectin > 50ng/ml before and after cervical examination to diagnose preterm birth within 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Fetal fibronectin > 50ng/ml before cervical examination[a]												
1 (McKenna 1999)	Prospective cohort	serious[1]	no serious inconsistency[2]	very serious[3,4]	very serious[5]	30%	50	100% (41.1 to 100.0)[c]	73.9% (68.8 to 73.9)[c]	3.83 (1.32 to 3.83)[c] Not useful	0.00 (0.00 to 0.85)[c] Very useful	Very low
Fetal fibronectin > 50ng/ml after cervical examinationa[b]												

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
1 (McKenna 1999)	Prospective cohort	serious1	no serious inconsistency [2]	very serious[3,4]	very serious[5]	30%	50	75.0% (22.7 to 98.7)[c]	65.2% (60.7 to 67.3)[c]	2.16 (0.58 to 3.02)c Not useful	0.38 (0.02 to 1.27)[c] Moderately useful	Very low

a A positive test result for fetal fibronectin was defined as > 50ng/ml. Cervical examinations were performed within 1 to 3 hours of the initial fetal fibronectin test.
b Results for changes in fetal fibronectin test results following cervical examination were also provided; 5/34 women who initially tested negative changed to positive after the second fetal fibronectin test, 2/16 women who initially tested positive changed to negative after the second test.
c Calculated by the NCC-WCH technical team.
1 The symptoms of preterm labour were not defined
2 Single study analysis
3 The use of tocolytic medications was not reported
4 The inclusion of women with multiple pregnancy is unclear
5 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)

Table 36: GRADE profile for evaluation of cervical length measured using transvaginal ultrasound to diagnose preterm birth within 48 hours or 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous preterm birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Cervical length measured using transvaginal ultrasound to diagnose preterm birth within 48 hours												
Cervical length ≤ 5mm												
1 (Tsoi 2005)	Multicentre observational cohort	no serious risk of bias[1]	no serious inconsistency[2]	serious[3,4]	serious[5]	NR	510	42.9% (24.2 to 61.2)a	97.8% (96.9 to 98.5)a	19.05 (7.93 to 41.84)a Very useful	0.59 (0.39 to 0.78)a Not useful	Low
Cervical length ≤ 10mm												
1 (Tsoi 2005)	Multicentre observational cohort	no serious risk of bias[1]	no serious inconsistency[2]	serious[3,4]	Serious[8]	NR	510	81.0% (59.0 to 93.6)[a]	93.7% (92.7 to 94.2)[a]	12.77 (8.10 to 16.14)[a] Very useful	0.20 (0.07 to 0.44)[a] Moderately useful	Low
Cervical length ≤ 15mm												
1 (Tsoi 2005)	Multicentre observational cohort	no serious risk of bias1	no serious inconsistency[2]	serious[3,4]	Serious[8]	NR	510	97.7% (78.8 to 100.0)a	84.8% (83.9 to 84.9)a	6.43 (4.91 to 6.62)a Moderately useful	0.03 (0.00 to 0.25)a Very useful	Low
1 (Gomez 2005)	Prospective cohort	serious6	no serious inconsistency[2]	very serious[3,7]	serious5	NR	215	64.7% (40.5 to 83.9)	90.4% (88.3 to 92.1)	6.74 (3.47 to 10.55) Moderately useful	0.39 (0.18 to 0.67) Moderately useful	Very low

Cervical length ≤ 25mm												
1 (Bagga 2010)	Prospective cohort	serious9	no serious inconsistency[2]	very serious[7,10]	serious[5]	NR	100	62.5% (44.6 to 76.6)a	89.5% (83.8 to 93.9)a	5.94 (2.75 to 12.60)a Moderately useful	0.42 (0.25 to 0.66)a Moderately useful	Very low
Cervical length < 30mm												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias1	no serious inconsistency[2]	very serious[3,7]	Very serious[11]	NR	395	88.0% (64.0 to 98.0)	40.0% (65.0 to 46.0)	1.48 (1.22 to 1.80) Not useful	0.29 (0.08 to 1.07) Moderately useful	Very low
1 (Gomez 2005)	Prospective cohort	serious[6]	no serious inconsistency[2]	very serious[3,7]	Very serious[11]	NR	215	88.2% (63.2 to 97.9)	53.0% (50.9 to 53.9)	1.88 (1.29 to 2.12) Not useful	0.22 (0.04 to 0.72) Moderately useful	Very low
Cervical length measured using transvaginal ultrasound to diagnose preterm birth within 7 days												
Cervical length ≤ 5mm												
1 (Tsoi 2005)	Multicentre observational cohort	no serious risk of bias1	no serious inconsistency2	very serious3,4	no serious imprecision	NR	510	37.2% (26.7 to 43.4)a	99.1% (98.2 to 99.7)a	43.44 (14.65 to 149.45)a Very useful	0.63 (0.57 to 0.75)a Not useful	Low
Cervical length ≤ 10mm												
1 (Tsoi 2005)	Multicentre observational cohort	no serious risk of bias1	no serious inconsistency2	very serious3,4	serious12	NR	510	65.1% (51.5 to 76.5)a	95.7% (94.5 to 96.8)a	15.21 (9.30 to 23.68)a Very useful	0.36 (0.24 to 0.51)a Moderately useful	Very low
Cervical length ≤ 15mm												
1 (Tsoi 2005)	Multicentre observational cohort	no serious risk of bias1	no serious inconsistency2	very serious3,4	serious13	NR	510	97.7% (86.9 to 99.9)a	88.7% (87.7 to 88.9)a	8.61 (7.04 to 8.96)a Moderately useful	0.03 (0.00 to 0.15)a Very useful	Very low
1 (Gomez 2005)	Prospective cohort	serious6	no serious inconsistency2	very serious3,4	serious12	NR	215	60.7% (43.6 to 75.1)	93.0% (90.5 to 95.2)	8.73 (4.58 to 15.66) Moderately useful	0.42 (0.26 to 0.62) Moderately useful	Very low
1 (Botsis 2006)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	serious[3,14]	very serious[15]	NR	62	81.8% (52.7 to 96.5)[a]	92.2% (85.9 to 95.3)[a]	10.43 (3.73 to 20.60)[a] Very useful	0.20 (0.04 to 0.55)[a] Moderately useful	Very low
1 (DeMirci 2011)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,16]	serious[8]	NR	209	78.9% (57.0 to 92.5)[a]	94.2% (92.0 to 95.6)[a]	13.64 (7.15 to 20.89)[a] Very useful	0.22 (0.08 to 0.47)[a] Moderately useful	Very low
1 (Gramellini 2007)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,17]	No serious imprecision	NR	108	26.3% (11.2 to 39.7)[a]	95.5% (92.3 to 98.4)[a]	5.86 (1.46 to 24.29)[a] Moderately useful	0.77 (0.61 to 0.96)[a] Not useful	low
1 (Holst 2006)	Prospective cohort	serious[9]	no serious inconsistency[2]	very serious[3,7]	serious[5]	NR	55	72.0% (56.0 to 63.0)	83.0% (70.0 to 93.0)	4.32 (1.88 to 11.04)[a] Not useful	0.34 (0.18 to 0.63)[a]	Very low

											Moderately useful	
1 (Palacio 2007)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	serious[18]	No serious imprecision	NR	333	28.6% (12.9 to 47.1)[a]	96.5% (95.4 to 97.7)[a]	8.10 (2.83 to 20.65)[a] Moderately useful	0.74 (0.54 to 0.91)[a] Not useful	low
1 (Sotiriadis 2010)	Prospective cohort	very serious[9,20]	no serious inconsistency[2]	serious[18]	very serious[21]	NR	122	83.3% (43.7 to 97.0)	95.8% (89.8 to 98.4)	20.00 (5.77 to 31.16)a Very useful	0.17 (0.01 to 0.65)a Moderately useful	Very low
Cervical length < 15mm in women admitted < 32 weeks' gestation												
1 (Palacio 2007)	Prospective cohort	no serious risk of bias 1	no serious inconsistency[2]	serious[18]	serious[5]	NR	116	10.0% (0.00 to 51.8)[a]	96.9% (96.5 to 98.8)[a]	3.23 (0.00 to 41.52)[a] Not useful	0.93 (0.49 to 1.04)[a] Not useful	low
Cervical length < 15mm in women admitted ≥ 32 weeks' gestation												
1 (Palacio 2007)	Prospective cohort	no serious risk of bias [1]	no serious inconsistency[2]	serious[18]	serious[5]	NR	217	35.3% (16.4 to 55.2)[a]	96.0% (94.4 to 97.7)[a]	8.82 (2.93 to 23.96)a Moderately useful	0.67 (0.46 to 0.89)[a] Not useful	low
Cervical length ≤ 20mm												
1 (Eroglu 2007)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	serious[22]	very serious[21]	3.9%	51	66.7% (27.1 to 91.3)[a]	95.6% (90.3 to 98.8)[a]	15.00 (2.79 to 78.49)[a] Moderately useful	0.35 (0.09 to 0.81)[a] Moderately useful	Very low
Cervical length ≤ 25mm												
1 (Bagga 2010)	Prospective cohort	serious[9]	no serious inconsistency[2]	very serious[7,10]	serious[5]	NR	100	60.0% (48.3 to 64.7)[a]	96.9% (91.6 to 99.5)[a]	19.50 (5.14 to 117.76)[a] Very useful	0.41 (0.36 to 0.57)[a] Moderately useful	Very low
1 (Gramellini 2007)	Prospective cohort	no serious risk of bias1	no serious inconsistency[2]	very serious[3,17]	serious[12]	NR	108	66.6% (45.7 to 83.3)[a]	79.3% (74.2 to 83.3)[a]	3.22 (1.77 to 5.00)a Not useful	0.42 (0.20 to 0.73)a Moderately useful	Very low
1 (Palacio 2007)	Prospective cohort	no serious risk of bias1	no serious inconsistency[2]	serious[18]	serious[12]	NR	333	71.4% (48.6 to 87.6)[a]	79.2% (77.6 to 80.3)[a]	3.43 (2.17 to 4.44)a Not useful	0.36 (0.16 to 0.66)a Moderately useful	Low
1 (Sotiriadis 2010)	Prospective cohort	very serious[9,20]	no serious inconsistency[2]	serious[18]	very serious[15]	NR	122	83.3% (43.7 to 97.0)	77.1% (67.7 to 84.3)	3.64 (1.46 to 16.61)a Not useful	0.22 (0.01 to 0.84)a Moderately useful	Very low
1 (Azlin 2010)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	serious[23]	Very serious[11]	NR	51	80.0% (31.3 to 98.9)	71.7% (66.4 to 73.8)	2.83 (0.93 to 3.78) Not useful	0.28 (0.01 to 1.03) Moderately useful	Very low
Cervical length < 30mm												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,7]	Very serious1[1]	NR	395	94.0% (79.0 to 99.0)	42.0% (37.0 to 47.0)	1.63 (1.43 to 1.84) Not useful	0.15 (0.04 to 0.57)	Very low

											Moderately useful	
1 (Gomez 2005)	Prospective cohort	serious[7]	no serious inconsistency[2]	very serious[3,7]	Very serious11	NR	215	89.3% (71.8 to 97.2)	55.6% (53.0 to 56.8)	2.01 (1.53 to 2.25) Not useful	0.19 (0.05 to 0.53) Moderately useful	Very low
1 (Danti 2011)	Prospective cohort	no serious risk of bias[24]	no serious inconsistency[2]	serious[25]	Very serious[11]	NR	102	90.0% (47.8 to 99.5)[a]	42.9% (40.3 to 43.4)[a]	1.58 (0.64 to 1.77)[a] Not useful	0.23 (0.00 to 1.53)[a] Moderately useful	Very low
Cervical length < 25mm												
1 (Palacio 2007)	Prospective cohort	no serious risk of bias [1]	no serious inconsistency[2]	serious18	very serious[21]	NR	116	75.0% (22.5 to 98.7)[a]	85.7% (83.8 to 86.6)[a]	5.25 (1.39 to 7.34)a Moderately useful	0.29 (0.02 to 0.92)a Moderately useful	Very low
Cervical length < 25mm												
1 (Palacio 2007)	Prospective cohort	no serious risk of bias [1]	no serious inconsistency[2]	serious[18]	Serious[5]	NR	217	70.6% (45.2 to 88.4)[a]	75.5% (73.3 to 77.0)[a]	2.88 (1.69 to 3.85)[a] Not useful	0.39 (0.15 to 0.75)[a] Moderately useful	Low
Change in cervical length 24 hours after admission measured using transvaginal ultrasound to diagnose pre-term birth within 7 days												
Change in cervical length > 20%												
1 (Sotiriadis 2010)	Prospective cohort	very serious[9,20]	no serious inconsistency[2]	serious[18]	serious[5]	NR	122	50.0% (18.8 to 81.2)	92.7% (85.7 to 96.4)	6.86 (1.54 to 16.61)a Moderately useful	0.54 (0.16 to 0.94)[a] Not useful	Very low
Cervical length < 15mm at admission plus change > 20% 24 hours later												
1 (Sotiriadis 2010)	Prospective cohort	very serious[9,20]	no serious inconsistency[2]	serious[18]	serious[5]	NR	122	50.0% (18.8 to 81.2)	99.0% (94.3 to 99.8)	48.00 (4.96 to 1171.37)[a] Very useful	0.51 (0.34 to 0.87)[a] Not useful	Very low
Cervical length > 15mm at admission plus change > 20% 24 hours later												
1 (Sotiriadis 2010)	Prospective cohort	very serious[9,20]	no serious inconsistency[2]	serious[18]	very serious[11]	NR	122	25.0% (0.0 to 90.3)[a]	93.0% (92.5 to 94.4)[a]	3.57 (0.00 to 16.17)[a] Not useful	0.81 (0.10 to 1.08)[a] Not useful	Very low
Cervical length < 25mm at admission plus change > 20% 24 hours later												
1 (Sotiriadis 2010)	Prospective cohort	very serious[9,20]	no serious inconsistency[2]	serious[18]	serious[5]	NR	122	50.0% (18.8 to 81.2)	97.2% (92.7 to 99.4)	24.00 (3.61 to 173.72)[a] Very useful	0.51 (0.24 to 0.88)[a] Not useful	Very low

a Calculated by the NCC-WCH technical team.
1 The primary clinician was blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics was not influenced by index test results.
2 Single study analysis.
3 Women included in the study were a mixed population where some received tocolysis and some did not.
4 52% of women (265/510) received tocolytic medication.
5 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to moderately useful (>0.1-0.5) 6 Clinicians were not blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics where decided by the attending obstetrician, may have been influenced by index test results and have affected when birth (the reference standard) occurred.
7 The proportion of women who received tocolysis was not reported.

8 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from moderately useful (>0.1-0.5) to very useful (0-0.1)
9 Blinding of clinicians to the index test was not reported.
10 Whether women received tocolytic medication was not reported.
11 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)
12 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranging from not useful (>0.5) to moderately useful (>0.1-0.5)
13 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranges from moderately useful (>0.1-0.5) to very useful (0-0.1)
14 45.4% of women who gave birth within 7 days of admission received tocolytic medication; 31.3% of women who did not give birth within 7 days received tocolytic medication.
15 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranging from not useful (>0.5) to very useful (0-0.1)
16 81% (21/26) of women with cervical length < 15mm received tocolytic medication; 52% (96/183) of women with cervical length > 15mm received tocolytic medication.
17 64.8% (70/108) of women received tocolytic medication.
18 All women received tocolytic medication.
19 Very wide CI LR+
20 No baseline characteristics were reported.
21 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranges from not useful (>0.5) to very useful (0-0.1)
22 6/51 (11.8%) women had a cervical length <20mm and these women received tocolytic therapy. 45 women had a cervical length >20mm and 10 of these women received tocolytics. Tocolytic therapy (with calcium channel blockers) was administered according to an existing protocol where women with progressive cervical change and persistent regular contractions, despite 2 hours bed rest and hydration with 500ml Ringer's lactate solution, received tocolysis. Maternal corticosteroids were given. No tocolytics or maternal steroids were used after 34 weeks gestation. The mean gestational age at enrolment was 29.5 ± 2.6.
23 12/51 (23.53%) women received tocolysis at the discretion of the attending clinician (further details of the tocolytic used or whether corticosteroids were administered were not reported).
24 Clinicians were not blinded to transvaginal ultrasound results which informed decisions regarding admission to hospital (where cervical length ≤ 30mm). However, subsequent clinical management, such as use of tocolytics, were also decided by the attending clinician, and may have been influenced by index test results, affecting when birth (the reference standard) occurred.
25 Tocolytics and corticosteroids were administered at the discretion of the attending clinician. 22/60 (37%) of women with cervical length ≤30mm and 5/42 (12%) women with cervical length >30mm received tocolysis. 28/60 (47%) of women with cervical length ≤30mm and 4/42 (10%) women with cervical length >30mm received corticosteroids.

Table 37: GRADE profile for evaluation of Bishop score plus cervical length measured using transvaginal to diagnose pre-term birth within 48 hours or 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous pre-term birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Bishop score and cervical length measured using transvaginal ultrasound to diagnose pre-term birth within 48 hours												
Bishop score between 4 and 7 and cervical length ≤ 20mm												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	serious[5]	15.5%	213	60.0% (26.0 to 88.0)	64.0% (57.0 to 71.0)	1.66 (0.75 to 2.43)[a] Not useful	0.63 (0.21 to 1.15)[a] Not useful	Very low
Bishop score between 4 and 7 and cervical length ≤ 25mm												

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous pre-term birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	very serious[6]	15.5%	213	80.0% (44.0 to 97.0)	46.0% (39.0 to 53.0)	1.48 (0.81 to 1.80)a Not useful	0.44 (0.08 to 1.23)a Moderately useful	Very low
Bishop score between 4 and 7 and cervical length ≤ 30mm												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	very serious[6]	15.5%	213	90.0% (55.0 to 100.0)	28.0% (22.0 to 34.0)	1.25 (0.75 to 1.39)[a] Not useful	0.36 (0.02 to 1.67)[a] Moderately useful	Very low
Bishop score and cervical length measured using transvaginal ultrasound to diagnose pre-term birth within 7 days												
Bishop score between 4 and 7 and cervical length ≤ 20mm												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	serious[5]	15.5%	213	55.0% (31.0 to 77.0)	65.0% (58.0 to 71.0)	1.57 (0.90 to 2.24)[a] Not useful	0.69 (0.37 to 1.06)[a] Not useful	Very low
Bishop score between 4 and 7 and cervical length ≤ 25mm												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	very serious[6]	15.5%	213	85.0% (62.0 to 97.0)	48.0% (41.0 to 55.0)	1.64 (1.16 to 1.87)a Not useful	0.31 (0.08 to 0.82)[a] Moderately useful	Very low
Bishop score between 4 and 7 and cervical length ≤ 30mm												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	very serious[6]	15.5%	213	95.0 (75.0 to 100.0)	29.0% (22.0 to 36.0)	1.34 (1.02 to 1.80)a Not useful	0.17 (0.01 to 0.94)a Moderately useful	Very low

a Calculated by the NCC-WCH technical team.
1 Clinicians were blinded to the results of transvaginal ultrasound but not blinded to Bishop score results.
2 Single study analysis.
3 Women included in the study were a mixed population where some received tocoloysis and some did not.
4 The proportion of women who received tocolysis was not reported.
5 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranging from not useful (>0.5) to moderately useful (>0.1-0.5)
6 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranging from not useful (>0.5) to very useful (0-0.1)

Table 38: GRADE profile for evaluation of a selective test based on cervical length measured using transvaginal ultrasound plus Bishop score to diagnose birth within 48 hours or 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous pre-term birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Selective test based on cervical length measured using transvaginal ultrasound and Bishop score to diagnose birth within 48 hours[a,b]												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	Very serious[5]	15.5%	213	88% (64 to 99)	58% (54 to 64)	2.08 (1.74 to 2.63) Not useful	0.20 (0.06 to 0.75) Moderately useful	Very low
Selective test based on cervical length measured using transvaginal ultrasound and Bishop score to diagnose birth within 7 days[a,b]												
1 (Schmitz 2008)	Prospective cohort	no serious risk of bias[1]	no serious inconsistency[2]	very serious[3,4]	serious[6]	15.5%	213	94% (79 to 99)	60% (55 to 65)	2.35 (2.01 to 2.74) Not useful	0.10 (0.03 to 0.40) Moderately useful	Very low

a A test was considered positive for pre-term birth if Bishop score was either ≥8 or between 4 and 7 combined with a cervical length ≤30mm.b The clinically selected population in this table only comprises women with a Bishop score of 4 to 7. Women with a Bishop score ≥8 were not included as their test results were deemed positive without additional data on cervical length from transvaginal ultrasound.
1 The primary clinician was blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics was not influenced by index test results.
2 Single study analysis.
3 Women included in the study were a mixed population where some received tocoloysis and some did not.
4 The proportion of women who received tocolysis was not reported.
5 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranging from not useful (>0.5) to very useful (0-0.1)
6 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranging from very useful (0-0.1) to moderately useful (>0.1-0.5)

Table 39: GRADE profile for evaluation of fetal fibronectin > 50ng/ml plus Bishop score to diagnose pre-term birth within 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous pre-term birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Fetal fibronectin > 50ng/ml and Bishop score >2a to diagnose birth within 7 days												
1 (Senden 1996)	Prospective cohort	serious1	no serious inconsistency	serious[2]	very serious[3]	NR	50	100%	95%	13.42 (2.16 to 23.0)b Very useful	0.13 (0.000 to 0.78) b Very useful	Very low

a Bishop score defined according to Myerscough P.R.; "Induction of labour" Chap 20 in Munro Kerr's Operative Obstetrics 10th edn. 1982, pub, Bailliere Tindall
b Calculated by the NCC-WCH technical team.

1 Clinicians were not blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics may have been influenced by index test results and have affected when birth (the reference standard) occurred
2 12/51 (23.53%) women received tocolysis at the discretion of the attending clinician (further details of the tocolytic used or whether corticosteroids were administered were not reported)
3 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranging from not useful (>0.5) to very useful (0-0.1)

Table 40: GRADE profile for evaluation of pIGFBP-1 test and cervical length measured using transvaginal ultrasound to diagnose pre-term birth within 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous pre-term birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
pIGFBP-1 test and cervical length measured using transvaginal ultrasound to diagnose birth within 7 days												
pIGFBP-1 test cervical length <20mm												
1 (Danti 2011)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	Serious[4]	NR	19	33% (1-91)	63% (35-85)	0.89 (0.16 to 4.97) Not useful	1.07 (0.44 to 2.59) Not useful	Low
pIGFBP-1 test and cervical length <25mm												
1 (Azlin 2010)	Prospective cohort	serious[6]	no serious inconsistency	Serious[2,6]	very serious[7]	NR	51	80.0% (34.4 to 98.2) [a]	97.8% (92.9 to 99.8) [a]	36.8 (4.83 to 508.35) a Very useful	0.20 (0.02 to 0.71) [a] Moderately useful	Very low
pIGFBP-1 test and cervical length 20-30mm												
1 (Danti 2011)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	very serious[7]	NR	41	75% (15-100)a	98.8% (95-100) a	61.5 (3.5 to 1083) a Very useful	0.25 (0.02 to 2.79) a Moderately useful	Very low
pIGFBP-1 test and cervical length ≤30mm												
1 (Danti 2011)	Prospective cohort	serious[1]	no serious inconsistency	Serious[2,3]	Serious[4]	NR	60	50% (7-93)	70% (56-81)	1.65 (0.57 to 4.74) Not useful	0.72 (0.27 to 1.94) Not useful	Low

a Calculated by the NCC-WCH technical team.
NR Not reported
1 Clinicians were not blinded to transvaginal ultrasound results. Women were admitted to hospital if cervical length ≤ 30mm (n=60) and offered a pIGFBP-1 test. Clinicians were blinded to the results of the pIGFBP-1 test (index test) therefore subsequent clinical management, such as use of tocolytics, which was decided by the attending clinician, was not influenced by index test results
2 Women included in the study were a mixed population where some received tocoloysis and some did not.
3 Tocolytics and corticosteroids were administered at the discretion of the attending clinician. 22/60 (37%) of women with cervical length ≤30mm and 5/42 (12%) women with cervical length >30mm received tocolysis. 28/60 (47%) of women with cervical length ≤30mm and 4/42 (10%) women with cervical length >30mm received corticosteroids. Corresponding information for subgroups of women with cervical length <20mm or 20-30mm are not reported
4 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranging from very useful (0-0.1) to moderately useful (>0.1-0.5)
5 Clinicians were not blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics may have been influenced by index test results and have affected when birth (the reference standard) occurred

6 12/51 (23.53%) women received tocolysis at the discretion of the attending clinician (further details of the tocolytic used or whether corticosteroids were administered were not reported)

7 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranging from not useful (>0.5) to very useful (0-0.1)

Table 41: GRADE profile for evaluation of fetal fibronectin > 50ng/ml plus cervical length to diagnose pre-term birth within 48 hours or 7 days

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous pre-term birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Fetal fibronectin > 50ng/ml and cervical length to diagnose birth within 48 hours[a]												
Fetal fibronectin and cervical length < 15mm												
1 (Gomez 2005)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	Serious[4]	13%	215	41.2% (20.9 to 61.6)[b]	95.5% (93.7 to 97.2)[b]	9.06 (3.32 to 22.07)[b] Moderately useful	0.62 (0.40 to 0.84)[b] Not useful	Very low
Fetal fibronectin and cervical length < 30mm												
1 (Gomez 2005)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	Serious[4]	13%	215	58.8% (34.7 to 79.8)[b]	85.9% (83.8 to 87.7)[b]	4.16 (2.14 to 6.46)[b] Not useful	0.48 (0.23 to 0.78)[b] Moderately useful	Very low
Fetal fibronectin > 50ng/ml plus cervical length to diagnose birth within 7 days[a]												
Fetal fibronectin and cervical length < 15mm												
1 (Gomez 2005)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	Serious[4]	13%	215	42.9% (28.4 to 52.2)[b]	97.7% (95.7 to 99.3)[b]	20.04 (6.60 to 69.99)[b] Very useful	0.58 (0.48 to 0.75)[b] Not useful	Very low
Fetal fibronectin and cervical length ≤ 15mm												
1 (Van Baaren 2014)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	Serious[4]	12%	714	88.7% (77.0 to to 95.7)[b]	26.7% (16.1 to 39.7)[b]	1.21 (1.01 to 1.45)[b] Not useful	0.40 (0.18 to 1.01.)[b] Moderately useful	Very low
Fetal fibronectin and cervical length < 30mm												
1 (Gomez 2005)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	Serious[4]	13%	215	60.7% (42.9 to 76.2)b	88.8% (86.1 to 91.1)[b]	5.41 (3.09 to 8.54)[b] Moderately useful	0.44 (0.26 to 0.66)[b] Moderately useful	Very low
Fetal fibronectin and cervical length 15 -20mm												
1 (Van Baaren 2014)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	no serious imprecision	12%	714	100% (42.9 to 76.2)b	47.7% (36.8 to 58.7)[b]	1.91 (1.56 to 2.34)[b] Not useful	0.00 Very useful	Very low
Fetal fibronectin and cervical length 20 -25mm												

Number of studies	Quality assessment						Sample size	Measures of diagnostic accuracy				Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations (percentage with previous pre-term birth)		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
1 (Van Baaren 2014)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	Serious[7]	12%	714	72.73% (39.1 to 93.6)[b]	54.1% (44.3 to 63.7)[b]	1.59 (1.05 to 2.40)[b] Not useful	0.50 (0.19 to 1.34) Moderately useful	Very low
Fetal fibronectin and cervical length 25 -30mm												
1 (Van Baaren 2014)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	very serious[5]	12%	714	80% (28.8 to 96.7)b	58.4% (46.6 to 69.6)b	1.93 (1.15 to 3.21)[b] Not useful	0.34 (0.06 to 2.0) Moderately useful	Very low
Fetal fibronectin and cervical length ≥ 30mm												
1 (Van Baaren 2014)	Prospective cohort	serious[1]	no serious inconsistency	very serious[2,3]	no serious imprecision	12%	714	100%	76.3% (70.5 to 81.4)[b]	4.22 (3.38 to 5.26)[b] Not useful	0.00 Very useful	Very low

a The test for fetal fibronectin was performed prior to transvaginal sonography and a digital examination carried out to ascertain cervical dilation and effacement.
b Calculated by the NCC-WCH technical team.
1 Clinicians were not blinded to the results of the index test therefore subsequent clinical management, such as use of tocolytics where decided by the attending obstetrician, may have been influenced by index test results and have affected when birth (the reference standard) occurred.
2 Women included in the study were a mixed population where some received tocolysis and some did not.
3 The proportion of women who received tocolysis was not reported.
4 Evidence was downgraded by 1 due to 95% confidence interval for the negative likelihood ratio ranging from very useful (0-0.1) to moderately useful (>0.1-0.5)
5 Evidence was downgraded by 2 due to 95% confidence interval for the negative likelihood ratio ranging from very useful (0-0.1) to not useful (>0.5)

9.1.4 Evidence statements

Likelihood ratios are reported as the primary measure of diagnostic accuracy. The positive likelihood ratio reports the number of times more likely women in preterm labour are to have that symptom than women not in preterm labour. The higher the value, the more likely it is that a woman with a positive test is actually in preterm labour. By convention, a value between 5 and 10 is regarded as moderately useful and a value of over 10 is very useful. Tests where the likelihood ratios lie close to 1 have little practical significance.

The negative likelihood ratio indicates whether the absence of a sign, score or a combination of tests is a good way of determining that a woman is not in preterm labour. The lower the value, the more likely it is that a woman with a negative test is not in preterm labour. In this case, the lower the value reported in the GRADE table the better the test may be for this diagnosis. By convention, a value of < 0.1 is regarded as very useful, and a value of 0.1 to 0.2 is moderately useful. Again, a negative likelihood ratio close to one demonstrates that a negative test is equally likely for women in preterm labour and those who are not. Hence, tests where the likelihood ratios lie close to 1 have little practical significance.

9.1.4.1 Diagnosis using clinical examination

Bishop score

Very low quality evidence from one prospective cohort study (n=70) found a not useful positive likelihood ratio and a moderately useful negative likelihood ratio for a Bishop score of 4 to 6 in diagnosing birth within 48 hours or 7 days of admission to hospital.

Evidence from a prospective cohort study (n=25) found a not useful positive likelihood ratio and a moderately useful negative likelihood ratio for a Bishop score of >2 in diagnosing birth within 7 days of admission to hospital. The quality of the evidence was very low.

Evidence from a large prospective cohort study (n=395) found a not useful positive likelihood ratio for a Bishop score ≥ 4 in diagnosing birth within 48 hours or 7 days but a moderately useful and very useful negative likelihood ratio for birth within 48 hours or 7 days, respectively. The quality of the evidence was low.

The same study found a very useful positive likelihood ratio and a not very useful negative likelihood ratio for a Bishop score ≥ 8 to diagnose birth within 48 hours or 7 days. The quality of the evidence was very low to low.

9.1.4.2 Diagnosis using biochemical tests

Phosphorylated insulin-like Growth Factor Binding Protein-1 (pIGFBP-1)

Four prospective cohort studies found (n=448) that IGFBP-1 had a not useful positive likelihood ratio and negative likelihood ratios that ranged from not useful to very useful in diagnosing birth within 48 hours of testing. The quality of the evidence was very low to low.

Low to very low quality evidence from 7 prospective cohort studies (n=567) investigating pIGFBP-1 testing reported positive likelihood ratios that ranged from not useful to very useful and negative likelihood ratios that were moderately or very useful in diagnosing birth within 7 days of testing.

Fetal fibronectin (fFN)

Two prospective cohort studies (n=333) examined fetal fibronectin test to diagnose birth within 48 hours. One study found that the test was not useful whilst the other reported that

the test was moderately useful when considering likelihood ratios. The quality of the evidence was very low.

Twelve prospective cohort studies (n=3,688) found that a fetal fibronectin test had a not useful positive likelihood ratio for diagnosing pre-term labour within 7 days of admission while 7 studies found it was moderately useful. Fourteen studies found a moderately useful, 3 studies a very useful and one study a not useful negative likelihood ratio for diagnosing pre-term labour within 7 days of admission. The quality of the evidence was very low to low.

Fetal fibronectin before or after digital cervical examination

One prospective cohort study (n=50) found a not useful positive likelihood ratio for a fetal fibronectin test both before and after digital cervical examination. The negative likelihood ratio was very useful and moderately useful for a fetal fibronectin test before and after digital cervical examination, respectively. The quality of the evidence was low for fetal fibronectin before cervical examination and very low for fetal fibronectin after cervical examination.

9.1.4.3 Diagnosis using ultrasound features

Cervical length measured using transvaginal ultrasound

Two prospective cohort studies (n=725) used a cervical length of < 15mm to diagnose birth within 48 hours. Both studies found moderately useful positive likelihood ratios; negative likelihood ratios were moderately useful and very useful. The quality of the evidence was low and very low respectively.

Two prospective cohort studies (n=610) used a cervical length of < 30mm to diagnose birth within 48 hours. Positive likelihood ratios were not useful and negative likelihood ratios moderately useful. The quality of the evidence was very low.

Low to very low quality evidence from 3 other cervical lengths of < 5mm, < 10mm and < 25mm were also used to diagnose birth within 48 hours. Positive likelihood ratios were very useful for < 5mm and < 10mm and moderately useful for < 25mm. Negative likelihood ratios were moderately useful for < 10mm and < 25mm and not useful for < 5mm.

Eight prospective cohort studies (n=1614) used a cervical length of < 15mm to diagnose birth within 7 days and found both positive and negative likelihood ratios to range from not useful to very useful. Sensitivity ranged from low to high and specificity ranged from moderate to high. The quality of evidence was very low.

Five prospective studies (n=714) used a cervical length of < 25mm to diagnose birth within 7 days. Positive likelihood ratios were not useful in all but one study which identified a very useful positive likelihood ratio for this test. Negative likelihood ratios were all moderately useful. The quality of the evidence was low in 2 studies for this test and very low in the remaining 3.

Three prospective studies (n=712) used a cervical length of < 30mm to diagnose birth within 7 days. Positive likelihood ratios were found to be not useful and negative likelihood ratios moderately useful. The quality of evidence was very low for this test in all studies.

Two further cervical lengths of < 5mm and < 10mm were used to diagnose birth within 7 days. Positive likelihood ratios were found to be very useful and negative likelihood ratios were not useful and moderately useful, respectively. Sensitivity was low and specificity high for both tests. The quality of the evidence was low and very low, respectively.

One prospective cohort study (n=122) identified evidence for change in cervical length > 20% at different cervical lengths to diagnose birth within 7 days. Tests included change score alone and in combination with cervical lengths < 15mm, > 15mm and < 25mm. Positive

likelihood ratios ranged from not useful to very useful. Negative likelihood ratios were not useful for all tests. Sensitivity was low and specificity high for all tests. The quality of the evidence was very low for all tests.

Subgroup analyses

One prospective study (n=116) provided evidence for subgroup analysis on cervical length (cut-offs of < 15mm and < 25mm) in women admitted before 32 weeks' gestation or 32 weeks' gestation and above to diagnose birth within 7 days. Positive and negative likelihood ratios were moderately useful and not useful. The quality of the evidence was very low to low.

9.1.4.4 Diagnosis using combinations of tests

Clinical examination plus ultrasound features

One prospective cohort study (n=213) found for that the combination of Bishop score of 4 to 7 and cervical lengths (measured using transvaginal ultrasound) of < 20mm, < 25mm and < 30mm provided not useful positive likelihood ratios to diagnose birth within 48 hours or 7 days. Negative likelihood ratios were not useful for this combination of tests for cervical lengths of < 20mm but moderately useful for cervical length < 25mm and < 30mm. The quality of the evidence was very low for all tests.

One prospective cohort study (n=213) found that a selective test that combined Bishop score of 4 to 7 and cervical length (measured using transvaginal ultrasound) of < 30 mm provided not useful positive likelihood ratios and moderately useful negative likelihood ratios for diagnosis of preterm labour within 48 hours or 7 days.

Biochemical tests plus clinical examination

One prospective cohort study (n=50) reported that the combination of a fetal fibronectin test > 50ng/ml and a Bishop score of >2 provided very useful positive and negative likelihood ratios to diagnose birth within 7 days. The quality of the evidence was very low.

Biochemical tests plus ultrasound features

pIGFBP1

One prospective cohort study (n=19) found that the combination of a pIGFBP-1 test and cervical length of <20mm or ≤30mm provided positive and negative likelihood ratios that were not useful for diagnosis of preterm labour within 7 days. Sensitivity and specificity were both low. The quality of this evidence was very low to low.

However, 2 prospective cohort studies (n=92) found that with cervical length measurements of <25mm or 20-30mm the same combination of tests provided positive likelihood ratios that were very useful and negative ratios were moderately useful for diagnosis of preterm labour within 7 days. The quality of this evidence was very low.

Fetal fibronectin

One study (n=215) found that the combination of fetal fibronectin and cervical length < 15mm provided moderately useful and very useful positive likelihood ratios, and not useful negative likelihood ratios, to diagnose birth within 48 hours and 7 days, respectively. Sensitivity was low and specificity high for at both time points.

However, the same study found that with cervical length measurements of < 30mm fibronectin and cervical length measurements provided not useful and moderately useful

positive likelihood ratios, and moderately useful negative likelihood ratios, to diagnose birth within 48 hours and 7 days, respectively. One other study (n=714) found that the combination of fetal fibronectin and cervical length ≤ 15mm provided not useful positive likelihood ratios, and moderately useful negative likelihood ratios with moderate sensitivity and low specificity , to diagnose birth within 7 days, The same study found that with cervical length measurement of 15-20mm, 20-25mm, 25-30 mm and ≥ 30mm this combination test provided not useful positive likelihood ratio and very useful and moderately useful negative likelihood ratios. The quality of the evidence was very low for all tests.

9.1.5 Health economics profile

This question was prioritised for health economic analysis.

A search was undertaken for health economic evidence on the diagnostic accuracy of the various tests (clinical assessment, biochemical tests, and ultrasound) used alone or in combination to identify preterm labour leading to preterm birth in women presenting with intact membranes. A total of 229 articles were identified by the search. After reviewing titles and abstracts, 15 full copies of papers were obtained but these were all excluded. Therefore, no relevant economic evidence was identified for this question.

In order to assess the cost effectiveness of alternative diagnostic strategies it is necessary to consider also the resources and interventions through which diagnosis can lead to improved health outcomes. Therefore, the analysis undertaken for this question utilised the output of the health economic model produced for the tocolytic review, as that is a treatment that could be offered as the result of a diagnostic assessment for women with suspected preterm labour and intact membranes.

The new health economic evaluation undertaken for this guideline took the form of a cost utility analysis and aimed to compare alternative diagnostic strategies in women to identify preterm labour in women with suspected preterm labour and intact membranes between the gestational ages of 24-34 weeks. Due to the limitations of the diagnostic accuracy review studies included in the clinical review, the evaluation took a “what-if” approach to diagnostic accuracy. This involved taking all 10,201 combinations of sensitivity and specificity between 0-100% with one percentage point increments and comparing their cost-utility relative to strategies of ‘treat all’ without diagnosis or ‘no diagnosis and no treat’ and determining what combinations of sensitivity and specificity were cost-effective for a given prevalence, diagnosis and treatment cost. ‘True positives’ (which in this case means those with preterm labour who are treated) were modelled as having an absolute risk of adverse outcomes as determined using the relative treatment effect of calcium channel blockers, which were assessed in 12.1.8 and 16.4 as being the most cost-effective tocolytic. ‘False positives’ (those in preterm labour not treated, either as a direct result of the strategy or a negative test result) are assumed to have the baseline risk of adverse outcomes. ‘False positives’ (those not in preterm labour but treated, either as a direct result of the strategy or a positive test result) do not receive any benefit from treatment but do incur the relevant treatment costs.

The results suggested that a treat all strategy was a cost-effective approach at lower gestational ages, but that the use of a diagnostic test was likely to be more cost-effective with increasing gestational age. Sensitivity analysis suggested that the cost of the diagnostic test (within plausible ranges) was not an important driver of cost-effective thresholds for treat all, treat based on diagnostic test and no diagnosis and no treatment. The inclusion of a cost for false negatives was also found to have little impact on model results. However, the analysis did suggest that model conclusions about cost-effective combinations of sensitivity and specificity could be sensitive to relatively small changes in prevalence.

The model is described in greater detail in Chapter 16.

9.1.6 Evidence to recommendations

9.1.6.1 Relative value placed on the outcomes considered

The Committee considered the following measures of diagnostic accuracy for decision-making for this topic: sensitivity, specificity, positive and negative likelihood ratio. The Committee considered the relative importance of having a high false positive and high false negative result in the diagnosis of preterm labour and the consequences for further management of women and babies.

Likelihood ratios were considered the most critical measures of diagnostic accuracy of different tests for preterm labour and for the Committee's decision-making. The Committee agreed that if a woman had the baby within 7 days of a positive test then she was 'truly' in labour and the test is a useful predictor.

Preterm labour (reference standard) was defined in relation to both 48 hours and 7 days. The outcome of preterm delivery at 48 hours was considered important because is related to the decision-making regarding the timing of steroid and magnesium sulfate administration.

The diagnosis of preterm labour in 7 days was considered equally important because if negative then the clinicians can be fairly confident that this women is unlikely to go into preterm labour in either 7 days or 48 hours and that would change the management strategy allowing discharge of women from hospital.

The Committee discussed in depth the importance of extending pregnancy even by days in early gestations to improve survival.

Avoiding false negative diagnoses was considered more important by the Committee compared to false positives, because the risks associated with preterm birth outside of hospital and the harms of not giving steroids and magnesium sulfate where indicated are likely to outweigh the harms of over-treatment of women incorrectly believed to be in preterm labour. Additionally, false negative diagnoses disproportionally impact women who live far from a tertiary centre and those at very early gestations.

The Committee also discussed the need for vaginal examination and concerns associated with the actual procedure as an invasive technique, as well as the role of vaginal examination in decision- making. The women's own views and circumstances are important in this decision-making (although not captured by diagnostic accuracy measures) as they are undergoing uncomfortable procedures (Bishop score and speculum exam) which might reduce the uptake rate.

9.1.6.2 Consideration of clinical benefits and harms

A number of studies were included in the review which considered different measurements.

The Bishop score was not found to be a helpful test for diagnosis of preterm labour unless the score was over 8. The Committee discussed the interpretation of these results and concluded that there was not much to gain from this test in terms of diagnosing women at preterm labour. In relation to potential harms, besides missing a diagnosis of preterm labour, the Committee discussed the risk of infection and discomfort associated with the invasive nature of this test. If used prior to sampling for fibronectin, then it could compromise the accuracy of this biochemical test. In summary, Bishop scoring was not considered an appropriate option used alone for diagnosis of preterm labour. However, the Committee did not feel a "do not do" recommendation is warranted because of the complexity of the decision, but wanted to consider the Bishop score as a 'last resort' test if other measurements for the diagnosis of preterm labour are not available.

The evidence showed that a short cervical length (<15mm) appears to have a moderately or very useful positive and negative likelihood ratio to diagnose women with preterm delivery at 48 hours whereas a cervical length < 5 mm had a very useful negative likelihood ratio. Additional usefulness of this test was found in relation to 20 mm and below for accurately diagnosing preterm labour at 7 days (useful positive likelihood ratio) and above 5mm for ruling out women without preterm labour (useful negative likelihood ratio) although confidence intervals are wide and results should be interpreted with caution.

There were mixed results in relation to the use of fetal fibronectin to diagnose PTLB. This test was found to be more useful to rule out preterm delivery.

The combination of either Bishop score or fibronectin testing and cervical length measurement were also found to be not useful to diagnose preterm labour. Although the Committee chose not to make a research recommendation, they commented that further research may be necessary because looking at individual tests would only be part of the full assessment for diagnosis of preterm labour and so combining them might be helpful.

pIGFBP1 plus cervical length was also not found to be a clinical relevant tool for diagnosis of preterm labour and of particular note was that this combination was of no more use than the use of cervical length measurements alone.

The Committee discussed the invasive nature of these techniques. They noted that in line with the NICE guideline on intrapartum care, an initial clinical assessment should include a clinical history and observations of the women and her baby (Rec 1.4.2). The Committee concluded that if recommended, biochemical or ultrasound testing should only occur if vaginal examination was being performed subsequent to these aspects of the initial assessment, hence minimising any additional discomfort associated with these diagnostic procedures.

9.1.6.3 Consideration of health benefits and resource uses

Diagnosis requires resources and therefore has opportunity costs, as those resources cannot be deployed elsewhere in the health care system with a concomitant loss in health related quality of life in those who could have benefited from such an alternative deployment. Therefore, for diagnosis to be cost-effective there is usually a minimum requirement that diagnosis can lead to improved outcomes. Therefore, the cost-effectiveness of diagnostic strategies are normally linked to the effectiveness of any treatment arising from a positive diagnosis and a consideration of health benefits and resource uses should usually be considered as part of a diagnostic and treatment package. In the analysis undertaken for this guideline diagnosis was linked to tocolysis and the costs of that treatment are included here in the consideration of health benefits and resource uses.

At low gestational ages the absolute risks of adverse outcomes are much larger and therefore the false negative rate can be particularly important in determining the most cost-effective strategy. This is because, at low gestational ages, false negatives can result in large losses of health related quality of life and, in this context, expensive lifetime NHS costs for adverse neurodevelopmental outcomes. However, absolute risks fall with increasing gestational age and therefore the relative benefits of treatment fall. As a result the false positive rate can be increasingly important as the ‘wasted’ resources from those who derive no treatment benefit from it are spread over fewer gains in health related quality of life. At the extremes a ‘treat all’ strategy minimises the false negatives and therefore is more likely to be cost-effective at the lower gestational ages. Conversely, a ‘do not diagnose, do not treat’ strategy minimises the number of false positives and therefore is more likely to be cost-effective at the higher gestational ages. The modelling undertaken for this guideline tended to bear out this logic and provided strong evidence that the cost-effective approach could vary with gestational age.

At earlier gestational ages when the absolute risks are high then treating all women with suspected preterm labour and intact membranes can be cost-effective even when allowing for the fact that 90% of those treated might not derive any treatment benefit. This is because the diagnostic accuracy of the tests is unlikely to produce a good enough trade-off in terms of reduced false positives to offset the high opportunity costs of missing false negatives at low gestational ages. At higher gestational ages, treatment can remain cost-effective at higher gestational ages when absolute risks are lower providing a diagnostic test can be applied with sufficiently good diagnostic accuracy, as additional benefit can be achieved without an unacceptable increase in cost arising from false positives.

Whilst a change in diagnostic strategy according to gestational age was indicated by the analysis undertaken for this guideline, the gestational age at which this change should take place is difficult to precisely identify given the uncertainty with respect to the diagnostic accuracy of the various tests. Nevertheless, the “what-if” data and results from the clinical review suggested that 30 weeks and beyond may be reasonable gestational age at which to require treatment to be guided by a positive diagnostic test, and thereby reduce inconvenience to women and costs to the health service when absolute risks are relatively lower. At 30 weeks there was some suggestion from the diagnostic studies reviewed that transvaginal ultrasound using a cervical length of ≤15mm could have sufficient diagnostic accuracy to be considered cost-effective relative to treat all, do not diagnose do not treat or other diagnostic tests or combinations of tests which do not have a cost-effective sensitivity/specificity combination.

9.1.6.4 Quality of evidence

The majority of evidence contributed to this section was moderate to low as assessed by the QUADAS checklist. Overall quality assessment was made based on a modified GRADE approach. The thresholds of measurements were not a priori selected based on clinical considerations but the results reported as per studies. The studies varied considerably in terms of populations’ characteristics and baseline risk therefore pooling the result was not appropriate but this is not unusual for diagnostic studies. Although the primary studies for fibronectin had used only one thresholds > 50ng/ml, the baseline characteristics of women across studies were very different to allow pooling of their results and therefore their interpretation is difficult. In addition, no complete information was given for all diagnostic measures (for example C.I. of sensitivity, specificity) to allow a diagnostic meta-analysis. In addition, the results on fibronectin should be interpreted with caution as sample sizes were small and quality of evidence was low.

9.1.6.5 Other considerations

There is not enough evidence to justify routine screening for cervical length in low-risk women’ with a reference to the National Screening Committee They noted that transvaginal ultrasound scanning is not available across the NHS because of limitations of equipment or expertise, and that investment in technology and training may be required for its universal implementation in the NHS. The Committee were aware of the importance of staff training to ensure that ultrasound measurements of cervical length were performed using consistent and standard criteria.

9.1.7 Key conclusions

The Committee concluded that measuring cervical length using transvaginal ultrasound is the most accurate way to diagnose preterm labour when used alone for women over 30 weeks in pregnancy. Fibronectin was also useful but not as good a diagnostic tool as cervical length. The Committee noted the importance of false positives and negatives and the associated harm with either missing women at risk of preterm birth who are deprived of the benefits of

treatment, or identifying wrongly that women are at risk of preterm birth resulting in unnecessary management.

They acknowledged the need for women to understand different diagnostic testing options including their associated benefits and harms and the interpretation of the results to guide possible subsequent management strategies.

The Committee made recommendations based on their interpretation of the evidence and on their clinical expert opinion.

Please see the health economics section of the tocolysis section in section 12.1.8.

9.1.8 Recommendations

22. Explain to women reporting symptoms of preterm labour who have intact membranes (and their family members or carers as appropriate):

- about the clinical assessment and diagnostic tests that are available
- how the clinical assessment and diagnostic tests are carried out
- what the benefits, risks and possible consequences of the clinical assessment and diagnostic tests are, including the consequences of false positive and false negative test results taking into account gestational age.

23. Offer a clinical assessment to women reporting symptoms of preterm labour who have intact membranes. This should include:

- clinical history taking
- the observations described for the initial assessment of a woman in labour in recommendation 1.4.2 of the NICE guideline on intrapartum care
- a speculum examination (followed by a digital vaginal examination[e] if the extent of cervical dilatation cannot be assessed).

24. If the clinical assessment suggests that the woman is in suspected preterm labour and she is 29^{+6} weeks pregnant or less, advise treatment for preterm labour as described in Chapters 10 – 12.

25. If the clinical assessment suggests that the woman is in suspected preterm labour and she is 30^{+0} weeks pregnant or more, consider transvaginal ultrasound measurement of cervical length as a diagnostic test to determine likelihood of birth within 48 hours. Act on the results as follows:

- if cervical length is more than 15 mm, explain to the woman that it is unlikely that she is in preterm labour and:
  - discuss with her the benefits and risks of going home compared with continued monitoring and treatment in hospital
  - advise her that if she does decide to go home, she should return if symptoms suggestive of preterm labour recur
- if cervical length is 15 mm or less, view the woman as being in diagnosed preterm labour and offer treatment as described in Chapters 10–12.

e Be aware that if a swab for fetal fibronectin testing is anticipated (see recommendation 26), the swab should be taken before any digital vaginal examination

26. Consider fetal fibronectin testing as a diagnostic test to determine likelihood of birth within 48 hours for women who are 30^{+0} weeks pregnant or more if transvaginal ultrasound measurement of cervical length is indicated but is not available or not acceptable. Act on the results as follows:

- if fetal fibronectin testing is negative, explain to the woman that it is unlikely that she is in preterm labour and:
  - discuss with her the benefits and risks of going home compared with continued monitoring and treatment in hospital
  - advise her that if she does decide to go home, she should return if symptoms suggestive of preterm labour recur
- if fetal fibronectin testing is positive, view the woman as being in diagnosed preterm labour and offer treatment as described in Chapters 10 – 12.

27. If a woman in suspected preterm labour who is 30^{+0} weeks pregnant or more does not have transvaginal ultrasound measurement of cervical length or fetal fibronectin testing to exclude preterm labour, offer treatment consistent with her being in diagnosed preterm labour (see Chapters 10 - 12).

28. Do not use transvaginal ultrasound measurement of cervical length and fetal fibronectin testing in combination to diagnose preterm labour.

29. Ultrasound scans should be performed by healthcare professionals with training in, and experience of, transvaginal ultrasound measurement of cervical length.

10 Maternal corticosteroids

10.1 Introduction

It has been recognised for many years that antenatal administration of corticosteroids to a mother prior to preterm birth reduces the severity of lung disease of prematurity and of other associated complications for her baby (Roberts and Dalziel 2006). This includes the severity of lung disease in babies of women with preterm birth associated with diabetes in pregnancy, hypertension in pregnancy and multiple pregnancy (see pregnancy, NICE, and NICE guideline CG129 Multiple Pregnancy). However, there remains uncertainty regarding the effectiveness of antenatal corticosteroids at the extremes of gestations of preterm birth.

In addition, women at risk of preterm birth, presenting with symptoms of preterm labour or in suspected preterm labour, do not always go on to deliver within the next 4-7 days, but may remain at high risk of preterm delivery. For these women, there is uncertainty about whether repeat courses of corticosteroids give additional benefit for fetal lung maturation, and if so, whether the risks of additional doses of corticosteroid (to both the fetus and the mother) may outweigh any benefit.

This section covers 2 aspects of maternal corticosteroids with regards to their clinical effectiveness for fetal lung maturation; their impact on neonatal outcomes given at different gestations and whether a repeated or single course is the most effective treatment option.

10.2 Single course of maternal corticosteroids at different gestations

10.2.1 Review question

What is the clinical effectiveness of a single course of maternal corticosteroids for fetal lung maturation given at different gestations in improving preterm neonatal outcomes?

10.2.2 Description of included studies

Two sources of evidence are included in the review for this question; a SR and meta-analysis (Roberts and Dalziel, 2013) with 21 component RCTs (3885 women, 4269 babies) from a variety of locations (USA [10 trials], Finland [2 trials], New Zealand [1 trial], UK [1 trial], The Netherlands [1 trial], South Africa [1 trial], Canada [1 trial], Brazil [1 trial], Jordan [1 trial], Tunisia [1 trial], Spain [1 trial]) and one RCT (Porto 2011) in Brazil. Additional data were obtained for 4 of the trials included in the SR, including IPD from the largest trial of corticosteroids for fetal lung maturation.

In the Cochrane review, the following subgroup analyses were conducted:

- gestational age at delivery (< 28 weeks, < 30 weeks, < 32 weeks, < 34 weeks, < 36 weeks, at least 34 weeks, at least 36 weeks)
- entry to delivery interval (< 24 hours, < 48 hours, 1–7 days, > 7 days)
- pre-labour rupture of membranes (at trial entry, > 24 hours before delivery, > 48 hours before delivery
- pregnancy-induced hypertension syndromes
- type of glucocorticoid (betamethasone, dexamethasone, hydrocortisone)

Post hoc subgroup analysis was performed for gestational age at entry to trial (< 26 weeks, between 26 and 29^{+6} weeks, between 30 and 32^{+6} weeks, between 33 and 34^{+6} weeks, between 35 and 36^{+6} weeks, > 37 weeks).

Fourteen of the included trials in the SR compared corticosteroids with placebo, whereas the remaining trials compared corticosteroids with expectant management. The choice of corticosteroid for the majority of trials in the SR (15 RCTs) was betamethasone and only 6 trials used dexamethasone (one trial did not report the corticosteroid used). The route and dosage of corticosteroid also varied between the trials, with the most common protocol of administration being 12mg betamethasone intramuscularly divided in 2 doses, 24 hours apart (6 trials). Eight trials in the SR allowed repeated courses of corticosteroids in their study protocols, although no clear information was given on the proportion of women who actually received repeated corticosteroids and if so, how many repeated courses per woman. Therefore, subgroup analysis for those women with single or repeated course of corticosteroids was not feasible.

The literature reports data based both on gestational age at delivery and gestational age at trial entry (i.e. gestational age at time of first corticosteroid administration). The Committee felt that both types of data would be helpful to inform clinical decision-making.

In order to estimate if the effect of maternal corticosteroids is biased by the inclusion of studies with repeated courses of corticosteroids, sensitivity analysis by excluding these studies was performed and their results were compared with the results from the overall meta-analysis.

The Committee preselected subgroup analysis at the protocol stage based on the following factors:

- gestational age at delivery (, < 24^{+0} weeks, < 26^{+0} weeks, < 28^{+0} weeks, < 30^{+0} weeks, < 32^{+0} weeks, < 34^{+0} weeks, < 36^{+0} weeks, at least 34^{+0} weeks, at least 36^{+0} weeks)
- gestational age at trial entry
- with and without intact membranes
- entry to delivery interval (< 24 hours, < 48 hours, 1 to 7 days, > 7 days)
- planned/spontaneous preterm birth

The range of women's gestational age varied considerably in the included RCTs from 24 weeks to 37 weeks. The mean gestational age at the trial entry ranged from 25.1 (SD 1.4) weeks for 11 RCTs to 32.0 (SD 3.2) weeks for 5 RCTs and ranging from 26.6 (SD 1.3) to 33.6 (SD 4.6) weeks. The mean time between corticosteroid administration and birth was often not clearly reported; with only a minority of studies (5) reporting the proportion of births within 7 days (from 50% to 76%).

The study population for this review question included women who were in spontaneous preterm labour, for whom a preterm birth was planned, or had prelabour premature rupture of membranes (P-PROM). Eight trials included all women with P-PROM whereas 4 studies included mixed populations with 23-63% of their total population having P-PROM. Details of tocolysis administration were reported in 11 trials; use of tocolytic drugs varied between trials and the percentage of women receiving tocolysis ranged from 23% to 100%.

Lastly, 10 out of 21 trials included only women with a singleton pregnancy. The remaining trials included mixed population of both single and twin pregnancies with the proportion of twin pregnancies ranging from 2% to 20%. There was no data available for sensitivity analysis on singleton pregnancies versus mixed pregnancies.

Additional source of evidence on antenatal corticosteroids before 26 weeks' gestation

One SR (Onland 2011) aimed to evaluate the effectiveness of antenatal corticosteroids before 26 weeks' gestation (extreme prematurity) and was included in this section. The

review included 9 trials whose eligibility criteria for entry in the study allowed for a lower gestational age cut-off less than 26 weeks. The majority of these studies (8 out of 9) were also included in the SR by Roberts and Dalziel 2013 whilst the ninth study studied combined corticosteroids and vitamin K therapy. Five of the 9 included trials permitted repeat courses of corticosteroids and 5 of the 9 trials included women with a multiple pregnancy.

10.2.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Study quality was assessed using the GRADE methodology. As RCTs were considered an appropriate study design for addressing this question, they were initially assigned a quality rating of 'high' and subsequently downgraded based on potential sources of bias.

The evidence is presented in the following GRADE table:

Table 42: GRADE profile for comparison of corticosteroids versus placebo or expectant management. The following subgroup factors were examined: women with premature rupture of membranes at first dose of corticosteroids, different gestations at birth, different gestations at first dose of corticosteroisin different gestations at birth (less than 28^{+0}, 30^{+0}, 32^{+0}, 34^{+0} and 36^{+0} weeks)

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

1 **Table 42: GRADE profile for comparison of corticosteroids versus placebo or expectant management**

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness[s]	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
Fetal and neonatal deaths[a]											
1 meta-analysis of 14 studies (Porto 2011; Roberts and Dalziel, 2013)	randomised trials	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	Repeat course: 4 studies Multiples: 8 studies	262/1957 (13.4%) 60/465	344/1945 (17.7%) 99/453	RR 0.77 (0.66 to 0.88)	41 fewer per 1000 (from 21 fewer to 60 fewer)	Low
Fetal and neonatal deaths – P-PROM at first dose (subgroup analysis)											
1 meta-analysis of 4 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	Repeat courses: 2 studies Multiples: 2 studies	55/368 (14.9%) 21/95	88/365 (24.1%) 42/89	RR 0.62 (0.46 to 0.82)	92 fewer per 1000 (from 43 fewer to 130 fewer)	Low
Fetal and neonatal deaths – gestational age at birth < 28 weeks (subgroup analysis)											
1 meta-analysis of 2 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias[3]	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 2 studies	39/60 (65%)	53/69 (76.8%)	RR 0.81 (0.65 to 1.01)	146 fewer per 1000 (from 269 fewer to 8 more)	Moderate
Fetal and neonatal deaths - gestational age at birth < 30 weeks (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious	Repeat courses: 0 studies Multiples: 1 study	59/99 (59.6%)	71/102 (69.6%)	RR 0.86 (0.7 to 1.05)	97 fewer per 1000 (from 209 fewer to 35 more)	Moderate
Fetal and neonatal deaths - gestational age at birth < 32 weeks (subgroup analysis)											
1 meta-analysis of 3 studies (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	serious1	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 2 studies	82/230 (35.7%)	110/223 (49.3%)	RR 0.71 (0.57 to 0.88)	143 fewer per 1000 (from 59 fewer to 212 fewer)	Low
Fetal and neonatal deaths - gestational age at birth < 34 weeks (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 1 study	90/312 (28.8%)	113/286 (39.5%)	RR 0.73 (0.58 to 0.91)	107 fewer per 1000 (from 36 fewer to 166 fewer)	Moderate
Fetal and neonatal deaths - gestational age at birth < 36 weeks (subgroup analysis)											

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness[s]	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
1 meta-analysis of 2 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	very serious4	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 2 studies	107/498 (21.5%)	135/471 (28.7%)	RR 0.75 (0.61 to 0.94)	72 fewer per 1000 (from 17 fewer to 112 fewer)	Very low
Fetal and neonatal deaths – mean gestational age < 28 weeks at trial entry (subgroup analysis)											
1 meta-analysis of 2 studies (Onland 2011)	randomised trials	no serious risk of bias	NR	no serious indirectness	very serious[6]	Repeat courses: 2 studies Multiples: 2 studies	NR	NR	RR 0.98 (0.57 to 1.67)	NC	Low
Fetal and neonatal deaths - first dose given before 26 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	None	15/23 (65.2%)	17/26 (65.4%)	RR 1.00 (0.66 to 1.50)	0 fewer per 1000 (from 222 fewer to 327 more)	Low
Fetal and neonatal deaths - first dose given between 26 and < 30 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	None	50/140 (35.7%)	54/121 (44.6%)	RR 0.80 (0.59 to 1.08)	89 fewer per 1000 (from 183 fewer to 36 more)	Moderate
Fetal and neonatal death - first dose given between 30 and < 33 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	None	19/165 (11.5%)	30/154 (19.5%)	RR 0.59 (0.35 to 1.01)	80 fewer per 1000 (from 127 fewer to 2 more)	Moderate
Fetal and neonatal death - first dose given between 33 and < 35 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	None	18/168 (10.7%)	18/185 (9.7%)	RR 1.10 (0.59 to 2.05)	10 more per 1000 (from 40 fewer to 102 more)	Low
Fetal and neonatal death - first dose given between 35 and < 37 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	None	3/87 (3.4%)	3/107 (2.8%)	RR 1.23 (0.25 to 5.94)	6 more per 1000 (from 21 fewer to 139 more)	Low
Fetal and neonatal death - first dose given after 36 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	None	3/18 (16.7%)	0/24 (0%)	RR 9.21 (0.51 to 167.82)	NC	Low

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness[s]	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
Intraventricular haemorrhage – all grades											
1 meta-analysis of 13 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	serious[1]	no serious indirectness	no serious imprecision	Repeat courses: 6 studies Multiples: 7 studies	88/1445 (6.1%)	155/1427 (10.9%)	RR 0.54 (0.43 to 0.69)	50 fewer per 1000 (from 34 fewer to 62 fewer)	Moderate
Intraventricular haemorrhage - premature rupture of membranes at first dose (subgroup analysis)											
1 meta-analysis of 5 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 3 studies Multiples: 2 studies	19/454 (4.2%) 15/195	38/441 (8.6%) 31/182	RR 0.47 (0.28 to 0.79)	46 fewer per 1000 (from 18 fewer to 62 fewer)	Moderate
Intraventricular haemorrhage - gestational age at birth < 28 weeks (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 1 study	5/34 (14.7%)	12/28 (42.9%)	RR 0.34 (0.14 to 0.86)	283 fewer per 1000 (from 60 fewer to 369 fewer)	Moderate
Intraventricular haemorrhage - gestational age at birth < 30 weeks (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 1 study	11/76 (14.5%)	19/74 (25.7%)	RR 0.56 (0.29 to 1.1)	113 fewer per 1000 (from 182 fewer to 26 more)	Moderate
Intraventricular haemorrhage - gestational age at birth < 32 weeks (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 1 study	13/144 (9%)	23/133 (17.3%)	RR 0.52 (0.28 to 0.99)	83 fewer per 1000 (from 2 fewer to 125 fewer)	Moderate
Intraventricular haemorrhage - gestational age at birth < 34 weeks (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 1 study	16/273 (5.9%)	27/242 (11.2%)	RR 0.53 (0.29 to 0.95)	52 fewer per 1000 (from 6 fewer to 79 fewer)	Moderate
Intraventricular haemorrhage - gestational age at birth < 36 weeks (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies	16/394 (4.1%)	27/373 (7.2%)	RR 0.56 (0.31 to 1.02)	32 fewer per 1000 (from 50 fewer to 1 more)	Moderate

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness[s]	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
						Multiples: 1 study					
Intraventricular haemorrhage (all grades) – mean gestational age < 28 weeks at trial entry in non-intervention arm (subgroup analysis)											
1 meta-analysis of 2 studies (Onland 2011)	randomised trials	no serious risk of bias	NR	no serious indirectness	very serious[6]	Repeat courses: 2 studies Multiples: 2 studies	NR	NR	RR 0.90 (0.45 to 1.78)	NC	Low
Intraventricular haemorrhage - first dose given before 26 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	None	3/15 (20%)	2/12 (16.7%)	RR 1.20 (0.24 to 6.06)	33 more per 1000 (from 127 fewer to 843 more)	Low
Intraventricular haemorrhage - first dose given between 26 to 29^{+6}weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	None	9/121 (7.4%)	18/108 (16.7%)	RR 0.45 (0.21 to 0.95)	92 fewer per 1000 (from 8 fewer to 132 fewer)	Moderate
Intraventricular haemorrhage - first dose given between 30 and < 33 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	None	1/155 (0.65%)	4/140 (2.9%)	RR 0.23 (0.03 to 2.00)	22 fewer per 1000 (from 28 fewer to 29 more)	Low
Intraventricular haemorrhage - first dose given between 33 and < 35 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	None	3/161 (1.9%)	3/178 (1.7%)	RR 1.11 (0.23 to 5.40)	2 more per 1000 (from 13 fewer to 74 more)	Low
Intraventricular haemorrhage - first dose given between 35 and < 37 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[10]	none	0/85 (0%)	0/106 (0%)	NC	NC	Low
Intraventricular haemorrhage - first dose given after 36 weeks gestation (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious10	None	0/18 (0%)	0/24 (0%)	NC	NC	Low
Intraventricular haemorrhage – grades 3 or 4b (subgroup analysis)											

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness[s]	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
1 meta-analysis of 4 studies (Garite 1992; Lewis 1996; Morales 1989; Silver 1996)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	Repeat courses: 4 studies Multiples: 2 studies	6/186 (3.2%) 3/125	30/187 (16%.0%) 15/117	RR 0.22 (0.10 to 0.49)	125 fewer per 1000 (from 82 fewer to 144 fewer)	High
Intraventricular haemorrhage grades 3 or 4 – mean gestational age < 28 weeks at trial entry in non-intervention arm (subgroup analysis)											
1 meta-analysis of 2 studies (Onland 2011)	randomised trials	no serious risk of bias	NR	no serious indirectness	no serious imprecision	Repeat courses: 2 studies Multiples: 2 studies	NR	NR	RR 0.20 (0.06 to 0.64)	NC	Moderate
Chronic lung disease											
1 meta-analysis of 6 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	very serious[4]	no serious indirectness	serious[2]	Repeat courses: 4 studies Multiples: 3 studies	48/413 (11.6%)	50/405 (12.3%)	RR 0.86 (0.61 to 1.22)	17 fewer per 1000 (from 48 fewer to 27 more)	Very low
Chronic lung disease - premature rupture of membranes at first dose (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	serious5	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 1 study Multiples: 0 studies	23/87 (26.4%)	41/78 (52.6%)	RR 0.5 (0.33 to 0.76)	263 fewer per 1000 (from 126 fewer to 352 fewer)	Low
Bronchopulmonary dysplasia at 28 days postnatal age –											
1 meta-analysis of 2 studies (Onland 2011)	randomised trials	no serious risk of bias	NR	no serious indirectness	serious[2]	Repeat courses: 2 studies Multiples: 2 studies	NR	NR	RR 1.18 (0.78 to 1.79)	NC	Moderate
Need for mechanical ventilation/CPAP											
1 meta-analysis of 5 studies (Porto 2011; Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious2	Repeat courses: 2 studies Multiples: 2 studies	64/430 (14.9%) 35/301	93/414 (22.5%) 58/284	RR 0.7 (0.54 to 0.91)	67 fewer per 1000 (from 20 fewer to 103 fewer)	Moderate
Need for mechanical ventilation/CPAP - premature rupture of membranes at first dose (subgroup analysis)											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	Repeat courses: 0 studies	15/105 (14.3%)	16/101 (15.8%)	RR 0.9 (0.47 to 1.73)	16 fewer per 1000 (from 84 fewer to 116 more)	Low

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness[s]	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
						Multiples: 1 study					
Neonatal sepsis											
1 meta-analysis of 6 studies (Porto 2011; Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 2 studies Multiples: 3 studies	38/809 (4.7%) 19/267	65/785 (8.3%) 37/253	RR 0.57 (0.39 to 0.83)	36 fewer per 1000 (from 14 fewer to 51 fewer)	Moderate
Neonatal sepsis - premature rupture of membranes at first dose (subgroup analysis)											
1 meta-analysis of 2 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency[7]	no serious indirectness	very serious[6]	Repeat courses: 1 study Multiples: 1 study	11/128 (8.6%)	11/123 (8.9%)	RR 0.96 (0.44 to 2.12)	4 fewer per 1000 (from 50 fewer to 100 more)	Low
Neonatal sepsis – mean gestational age < 28 weeks at trial entry in non-intervention arm (subgroup analysis)											
1 study (Onland 2011)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	Repeat courses: unclear Multiples: Unclear	NR	NR	RR 0.40 (0.04 to 3.70)	NC	Low
Cerebral palsy in childhood (at 2-year follow up)											
1 meta-analysis of 5 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Multiples: 4 studies	20/490 (4.1%)	28/414 (6.8%)	RR 0.6 (0.34 to 1.03)	27 fewer per 1000 (from 45 fewer to 2 more)	Moderate
Cerebral palsy in childhood (at 2-year follow up) (subgroup analysis with studies without repeated courses of corticosteroids											
Roberts and Dalziel 2013	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	Repeat courses: 1 study	1/60	2/34	RR 0.28 (0.03 to 3.01)	42 fewer per 1000 (from 57 fewer to 123 more)	Low
Visual impairment in childhood (at 2-year follow up)											
1 meta-analysis of 2 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 0 studies Multiples: 2 studies	9/100 (9%)	11/66 (16.7%)	RR 0.55 (0.24 to 1.23)	75 fewer per 1000 (from 127 fewer to 38 more)	Moderate
Hearing impairment in childhood (at 2-year follow up)											
1 meta-analysis of 2 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency[7]	no serious indirectness	very serious[6]	Repeat courses: 0 studies	1/100 (1%)	1/66 (1.5%)	RR 0.64 (0.04 to 9.87)	5 fewer per 1000 (from 15 fewer to 134 more)	Low

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness[s]	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
						Multiples: 2 studies					
Neurodevelopmental delay in childhood (at 24-month follow up ; defined as tetraplegic cerebral palsy and/or a score < 70 on Bayley Scales for 2-year children)[c]											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	Repeat courses: 0 studies Multiples: 1 study	3/50 (6%)	3/32 (9.4%)	RR 0.64 (0.14 to 2.98)	34 fewer per 1000 (from 81 fewer to 186 more)	Low
Developmental delay in childhood (at 18 to 24-month follow up ; defined as Psychomotor Developmental Index of the Bayley Scales at 18 months of age (50 ≤ Index ≤ 67)[d]											
1 meta-analysis of 2 studies (Roberts and Dalziel 2013) 1 study remains when multiples removed – no effect remains, confidence interval widens	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	Repeat courses: 1 study Multiples: 1 study	11/266 (4.1%) 4/60	19/252 (7.5%) 7/34	RR 0.49 (0.24 to 1) RR 0.32 (0.10 to 1.03)	38 fewer per 1000 (from 57 fewer to 0 more)	Moderate
Intellectual impairment in childhood (at 18 to 24-month follow up ; defined in Collaborative 1981 as Mental Developmental Index of the Bayley Scales at 18 months of age (50 ≤ Index ≤ 67); Liggins 1972 - ≤ 70 on Stanford-Binet Intelligence Scale and in Schutte 1980 as < 70 on Weschler Intelligence Scale for Children-Revised full-scale IQ)											
1 meta-analysis of 3 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency	serious8	serious[2]	Repeat courses: 0 studies Multiples: 3 studies	16/409 (3.9%)	17/369 (4.6%)	RR 0.86 (0.44 to 1.69)	6 fewer per 1000 (from 26 fewer to 32 more)	Low
Behavioural/learning difficulties in childhood (at 24-month follow up ; defined as children who had to repeat a class or required special education[f]											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	Repeat courses: 0 studies Multiples: 1 study	9/54 (16.7%)	7/36 (19.4%)	RR 0.86 (0.35 to 2.09)	27 fewer per 1000 (from 126 fewer to 212 more)	Low
Maternal death (where all women had severe preeclampsia) [g]											
1 meta-analysis of 3 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	no serious inconsistency[9]	no serious indirectness	very serious[6]	Repeat courses: 1 study Multiples: 2 studies	1/188 (0.53%)	1/177 (0.56%)	RR 0.98 (0.06 to 15.5)	0 fewer per 1000 (from 5 fewer to 82 more)	Low
Side-effects of therapy in women											
1 study (Roberts and Dalziel 2013)	randomised trial	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[8]	Repeat courses: 0 studies Multiples: 1 study	0/50 (0%)	0/51 (0%)	NC	NC	Low
Puerperal sepsis											

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Corticosteroids	Placebo or expectant management	Relative (95% CI)	Absolute (95% CI)	Quality
1 meta-analysis of 8 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	serious[1]	no serious indirectness	serious[2]	Repeat courses: 4 studies Multiples: 2 studies	57/496 (11.5%)	44/507 (8.7%)	RR 1.35 (0.93 to 1.95)	30 more per 1000 (from 6 fewer to 82 more)	Low
Puerperal sepsis – premature rupture of membranes at first dose (subcommittee analysis)											
1 meta-analysis of 4 studies (Roberts and Dalziel 2013)	randomised trials	no serious risk of bias	serious[1]	no serious indirectness	very serious[6]	Repeat courses: 2 studies Multiples: 2 studies	16/242 (6.6%)	14/235 (6%)	RR 1.11 (0.55 to 2.25)	7 more per 1000 (from 27 fewer to 74 more)	Very low

1 *CI confidence interval, RR relative risk, NC not calculable, NR not reported, CPAP continuous positive airway pressure1 Evidence was downgraded by 1 due to serious*
2 *heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%)*
3 *2 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID 3 In one trial (contributes 15% to meta-analysis) significantly*
4 *lower gestational age in control than experimental group at entry to study and birth, significantly more women in control group received tocolysis*
5 *4 Evidence was downgraded by 2 due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%)*
6 *5 Unclear method of randomisation and allocation concealment*
7 *6 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs 8*
8 *7 Each of the 3 trials used a different scale to measure intellectual impairment (3/3 trials included women with a multiple pregnancy)*
9 *8 Confidence interval could not be calculated: zero events in both arms of the trial so effect estimate cannot be calculated; trial underpowered for outcome*

10

10.2.4 Evidence statements

Neonatal outcomes

Fetal and neonatal deaths

In the overall analysis, findings from a meta-analysis of 14 RCTs (n=3902) suggested there were significantly fewer fetal and neonatal deaths in the group of women who had received corticosteroids compared with women who had placebo or expectant management. Subgroup analysis in women with P-PROM had also showed significantly fewer fetal and neonatal deaths in babies born to women who had received corticosteroids compared with women who had placebo or expectant management. Subgroup analysis by gestational age at birth suggested that there were significantly fewer fetal and neonatal deaths in babies born at less than 32 weeks, less than 34 weeks and less than 36 weeks to women who had corticosteroids compared with placebo or expectant management, however there was no significant difference in fetal and neonatal deaths in babies born at less than 28 weeks or less than 30 weeks. Subgroup analysis by gestational age at first dose of corticosteroids suggested that there was no significant difference in fetal and neonatal death in the steroid versus the placebo group at any gestational age cut-off. The evidence was of moderate to low quality across all outcomes.

Intraventricular haemorrhage

Evidence from a meta-analysis of 13 RCTs (n=2872) suggested there were significantly fewer any grade of intraventricular haemorrhage and a significantly lower proportion of grade 3 or 4 intraventricular haemorrhage, in babies born to women who had received corticosteroids compared with women who had placebo or expectant management. Subgroup analysis in women with P-PROM had also showed fewer intraventricular haemorrhage in babies born to women who had received corticosteroids compared with women who had placebo or expectant management. Subgroup analysis by gestational age at birth from one study suggested that there were fewer babies with intraventricular haemorrhage born at less than 28 weeks, less than 32 weeks, and less than 34 weeks to women who had corticosteroids compared with placebo or expectant management, however subgroup analysis from one other study showed no significant difference in intraventricular haemorrhage in babies born at less than 30 weeks and less than 36 weeks. Fewer babies who were between 26 weeks and 30 weeks gestation at first corticosteroid dose had intraventricular haemorrhage compared with placebo or expectant management. Further analysis of trials where the mean gestation age of babies was less than 28 weeks found that a significantly lower proportion of babies born to women who had received corticosteroids had an intraventricular haemorrhage grade 3 and 4 compared with babies born to women who had received placebo or expectant management. No significant difference was found in the number of babies born with any grade of intraventricular haemorrhage at 28 days. The quality of the evidence for all outcomes ranged from moderate to low (with the majority rated moderate)

Bronchopulmonary dysplasia/Chronic lung disease

Evidence from a meta-analysis of 2 RCTs (n=not reported) suggested there was no significant difference in bronchopulmonary dysplasia at 28 days following birth, in babies born to women who had received corticosteroids compared with women who had placebo or expectant management. Evidence from meta-analysis of 5 RCTs (n=818) suggested there was no significant difference in chronic lung disease, in babies born to women who had received corticosteroids compared with women who had placebo or expectant management. However, subgroup analysis in women with P-PROM showed significantly lower rate of chronic lung disease in babies born to women who had received corticosteroids compared

with women who had placebo or expectant management. The evidence was of low to very low quality.

Need for mechanical ventilation

Moderate and low quality evidence from meta-analysis of 6 RCTs (n=844) suggested there was significantly less need for mechanical ventilation in babies born to women who had received corticosteroids compared with women who had placebo or expectant management. However, subgroup analysis in women with P-PROM showed no significant difference in need for mechanical ventilation in babies born to women who had received corticosteroids compared with women who had placebo or expectant management.

Neonatal sepsis

Moderate and low quality evidence from meta-analysis of 6 RCTs (n=1594) suggested there were significantly lower rates of neonatal sepsis in babies born to women who had received corticosteroids compared with women who had placebo or expectant management. However, subgroup analysis in women with P-PROM and by gestational age at less than 28 weeks showed no significant difference in neonatal sepsis in babies born to women who had received corticosteroids compared with women who had placebo or expectant management.

Neurodevelopmental disability

There was no difference in the proportion of babies who developed cerebral palsy at 2 year follow up (5 studies, n=904) or the proportion of children with visual or hearing impairment (2 studies n=166), neurodevelopmental (1 study, n=82)or developmental delay (2 studies, n=518), intellectual impairment (3 studies, n=778) or behavioural/learning difficulties (1 study, n=90) born to women who had corticosteroids compared with women who had placebo or expectant management. The quality of the evidence ranged from moderate to very low.

Maternal outcomes

Low quality evidence from meta-analysis of 3 RCTs (n=365) showed there was no significant difference in maternal death (where all women had severe pre-eclampsia), and in puerperal sepsis (8 studies, n=1003) between women who had corticosteroids compared with women who had placebo or expectant management. One study (n=101) reported that there were no side effects in both women who had corticosteroids and women who had placebo or expectant management. The evidence was of low quality.

Further sensitivity analyses

Single and repeat course trials

Sensitivity analysis was conducted to test any differences in the direction of results from only single-course corticosteroid trials. When we excluded the studies with repeat-course corticosteroids studies from the analysis, we found the following differences in the results of the main analysis:

- the beneficial effect of corticosteroids on fetal and neonatal deaths and intraventricular haemorrhage remained significant ,
- the beneficial effect of corticosteroids on the need for mechanical ventilation and neonatal sepsis was no longer significant, and
- it remained that no difference was observed in effect on chronic lung disease, cerebral palsy, developmental delay and puerperal sepsis.

Singleton and multiple pregnancy trials

Sensitivity analysis was done to compare the results of trials with only singleton pregnancies and trials with mixed populations (both single and multiple pregnancies). When trials that only included women with single pregnancies were considered, the beneficial effect of corticosteroids on fetal and neonatal deaths, intraventricular haemorrhage (all grades and grades 3 and 4), need for mechanical ventilation and neonatal sepsis remained significant. When trials that included women with multiple pregnancies were excluded from the analysis, corticosteroids were shown to have a beneficial effect in reducing chronic lung disease.

10.2.5 Health economics profile

A single search was undertaken for health economic evidence on a course or repeat courses of maternal corticosteroids for fetal lung maturation given at different gestations in improving preterm neonatal outcomes. A total of 136 articles were identified by the search. After reviewing titles and abstracts, 9 full papers were obtained and one was included for review.

An old UK study (Mugford 1991) used a decision analytic approach to assess the cost-effectiveness of antenatal corticosteroids relative to no treatment in order to prevent death of a preterm infant and to prevent respiratory distress syndrome in surviving preterm infants. The model considered the cost-effectiveness of treatment across 2 different populations of women with threatened preterm labour: <31 weeks and <35 weeks. The authors reported that antenatal corticosteroids was dominant (cheaper and more effective) across the 2 different populations.

This question was initially prioritised for health economic analysis although new analysis was ultimately not undertaken as there were more important topics to be addressed in terms of health economic analysis. Furthermore, corticosteroids are relatively cheap and a single course is current practice. The review did not find clinical evidence to suggest that repeat courses were beneficial and therefore there cannot be economic evidence that would justify the routine use of repeat courses.

10.2.6 Evidence to recommendations

10.2.6.1 Relative value placed on the outcomes considered

Given that the primary intention of administering maternal corticosteroids is to accelerate fetal lung maturation in babies who are likely to be born preterm, the Committee prioritised measures that would be likely to indicate whether the drug had been successful in achieving this end. This included any pulmonary-specific adverse event or need for mechanical ventilation and bronchopulmonary dysplasia/chronic lung disease, as well as proxy measures of neonatal lung disease (mortality, all death up to one year), neurodevelopmental disability and intraventricular haemorrhage, or white matter injury (periventricular leucomalacia [PVL]).Although the Committee agreed *a priori* that all manifestations of neurodevelopmental disability could be reported as a single outcome, the data were available for particular developmental disabilities and results were presented individually for each type of disability.

Neonatal sepsis was prioritised as an important outcome because corticosteroids are used in other contexts to suppress the body's normal immune response to infection. Therefore the Committee considered that a rise in infection rates might be a potential, unintended adverse outcome of steroid use.

In terms of maternal outcomes, the Committee prioritised mortality because they felt any change in the incidence of this outcome would affect clinical decision-making. They also agreed that the composite of all maternal adverse events was important. They were aware of the possible effect of maternal corticosteroids on blood sugar control in women with diabetes

and the potential for adverse events related to the immunosuppressive effect of corticosteroids.

10.2.6.1.1 Consideration of clinical benefits and harms

There was consistent evidence from randomised trials that maternal corticosteroids are beneficial from 26 weeks' gestation in terms of reducing neonatal morbidity including intraventricular haemorrhage (all grades and grade 3 or 4), need for mechanical ventilation and neonatal sepsis. Given that some of these benefits (specifically reduced IVH and lung disease of prematurity) were found for gestations down to 26 weeks, the Committee considered that they would also likely apply to babies born before 26 weeks' gestation.

There was also some evidence that maternal corticosteroids may reduce fetal and neonatal mortality from 26 weeks but the Committee interpreted these results with caution due to low quality of the evidence and the concerns around its generalisation. Also these studies are likely to be underpowered for low event rate outcomes such as these.

The Committee was reassured that any concerns regarding the potential risk of neonatal sepsis were not supported by the included evidence. Infection rates were reduced in the intervention group (corticosteroids) compared to the 'no treatment' arm. It was noted that the reduction in the incidence of neonatal sepsis was less marked in the sub-group of women who had preterm pre-labour rupture of membranes (P-PROM). However, the Committee concluded that the potential risk of infection in this sub-group does not negate the benefits of antenatal corticosteroids.

There was limited evidence available for any of the outcomes when maternal corticosteroids were given before 26 weeks. There were only data for 2 of the prioritised outcomes: fetal and neonatal deaths and intraventricular haemorrhage. Nevertheless, the Committee considered it reasonable to extrapolate from later gestational ages down to 23 weeks because the physiology of lung disease and IVH are the same at 23-25 weeks gestation, as compared to 26 weeks gestation. Therefore the Committee p considered that it would be plausible to expect the same benefits to be seen at lower gestations as were reported for babies of 26 weeks gestation and above. However, it was acknowledged that any benefits from the use of maternal corticosteroids in this group may be relatively small in the context of the significant morbidities and mortality rate of babies who are born at such extremely preterm gestations. The Committee noted that in current clinical practice the decision to give maternal corticosteroids at extreme preterm gestations is often influenced by the decision on whether or not to offer neonatal life support at birth, following discussion with parents regarding the risks and likely outcomes. The Committee considered that this should be discussed carefully on an individual basis with the woman and her partner and with reference to other relevant guidance such as that published by the British Association of Perinatal Medicine (2008) and the Royal College of Obstetrics and Gynaecology (2014).

There was no reliable evidence of benefit of antenatal corticosteroids in terms of fetal or neonatal death, intraventricular haemorrhage, or chronic lung disease above 34 weeks gestation. The effectiveness of antenatal corticosteroids in reducing requirement for ventilation or pressure support has not been reported. The Committee considered that any potential short-term benefits to the baby of antenatal corticosteroid administration should be balanced against potential risk of delaying delivery at these gestations.

10.2.6.1.2 Consideration of health benefits and resource uses

The Committee noted that corticosteroids are relatively inexpensive when viewed in isolation but it was necessary to take into account the cost of the other associated management strategies (such as the use of drugs to delay birth to allow corticosteroids to be administered) when assessing cost-effectiveness. The cost of the combined strategy was considered in the section on tocolysis.

However, the Committee considered that the use of maternal corticosteroids in women whose babies were most likely to benefit had the potential to reduce the number of ventilated days and that this would result in a cost saving that would outweigh the initial costs incurred.

10.2.6.1.3 Quality of evidence

The evidence included in the review was mostly of moderate or low quality. The Committee had some reservations about the appropriateness of included studies. The Committee considered that although the participants in the included trials reflected the study population of the guideline, within these trials were many sub-populations who might potentially vary in terms of their response to corticosteroids (for example, women in spontaneous preterm labour and those having a planned preterm birth). The Committee acknowledged that multiple sub-group analyses would not give any precision in the estimate of effects. The sensitivity analysis excluding women with multiple pregnancy indicated that findings were broadly the same as those in overall analysis so separate recommendations were not warranted.

The Committee had concerns regarding the relevance of data derived from trials that were conducted in the 1970s. The Committee noted that clinical practice had changed significantly in the interim so that these findings may not be currently relevant. For example, the availability of neonatal interventions in contemporary practice may lessen the impact of maternal corticosteroid administration to prevent death in preterm neonates. However, the Committee considered that, because more preterm babies are surviving, the effects of corticosteroids on neonatal complications may have become more important. In light of these considerations, the Committee considered that it was reasonable to consider this evidence in the context of the meta-analyses where data from more recent studies contributed an equal or dominant amount of data to the overall result.

The Committee also noted that the SR by Onland added little value in terms of specific information regarding the effectiveness of steroids a lower gestational ages (below 26 weeks) because the data were only reported in terms of 2 gestational age groups - those with a mean gestational age below or equal to 28 weeks and those with a mean gestational age of over 28 weeks. However, although the data were included to given an estimate of subgroup analysis to lower gestational ages, results should be interpreted with caution given the effect of data double counting (as some of the data were also included in the SR by Roberts and Dalziel, 2013. It was highlighted that the length of follow-up in the included studies was too short to report accurately outcomes such as cerebral palsy and learning difficulties, but the consensus view of the Committee was that IVH grades 3 and 4 are likely to lead to these outcomes and could therefore be accepted as useful short-term proxies.

10.2.6.1.4 Other considerations

The Committee noted the lack of available evidence on which to judge the optimal timing of administration of corticosteroids in relation to the time of birth and particularly the ‘latest point’ at which the drug could most effectively be given. The Committee acknowledged that this had not been prioritised as an aim of the review but, taking account of the drug’s pharmacological mechanism of action, the Committee suspected that any benefits would be likely to be transferred even if there was only a limited amount of time (such as less than 24-48 hours) between administration and time of birth. However, the Committee could not make any recommendations to this effect.

The Committee was aware of the existence of 2 other studies (EPICURE and EPIPHASE) that did not meet the inclusion criteria but might provide further information about use of corticosteroids at low gestations.

10.2.6.1.5 Key conclusions

There was sufficient evidence of benefit without concomitant harm to justify a strong recommendation for the use of corticosteroids in women who are thought to be in spontaneous preterm labour, having planned a preterm birth or have preterm prelabour rupture of membranes between 26 and 34 weeks gestation. The Committee concluded that some of these benefits would be seen in babies born at lower and higher gestational ages, but that the evidence was less robust at these gestations. The extrapolation of findings to the groups outside the gestational age range of 26-34 weeks was more complex in terms of clinical effectiveness, which warranted less strong recommendations at gestations below 26 weeks and above 34 weeks.

10.2.6.2 Recommendations

The recommendations on corticosteroids are in Section 10.3.7.

10.3 Repeat course corticosteroids for fetal lung maturation

10.3.1 Review question

What is the clinical effectiveness of repeat courses of maternal corticosteroids for fetal lung maturation in improving preterm neonatal outcomes?

10.3.2 Description of included studies

Three studies are included for this review question; a SR of 10 RCTs (Crowther 2013), an RCT (Atarod 2014), and a 5-year follow up study (Asztalos 2013) of a RCT already included in the SR (Murphy 2008).

The total sample size of SR was 4733 women (with 5700 babies) from a variety of locations (USA [5 trials], Canada [1 trial], India [1 trial], Finland [1 trial], multicentre [2 trials]). Three of the trials stopped recruitment early: in one trial this was due to the small probability of detecting a difference in primary outcome and safety concerns that were reported in the literature; in one trial it was due to a decrease in intact survival in the repeat corticosteroids group; and in one trial it was due to a tendency towards lower birth weight in the repeat corticosteroids group. The long term follow up study included 1728 children.

The intervention arm in all studies was a single course of corticosteroids which often taken more than 7 days previously in women who continued to be at high risk of preterm birth. The definition of high risk of preterm birth was not clearly reported in the majority of studies and none of the studies included women with planned preterm births. Betamethasone was the type of corticosteroid used in the studies. The course varied between the trials, with the most common course being 2 doses, 24 hours apart, of 12mg betamethasone intramuscularly (IM), repeated weekly until 33 to 34 weeks or birth (5 trials). One trial used this course but repeated it fortnightly until 33^{+6} weeks or birth. In 3 trials the protocol allowed only one repeat course of 12mg betamethasone IM (2 doses, 24 hours apart).

Six trials reported the mean gestational age at trial entry ranging from 28.0 (2.4) weeks to 30.7 (2.5) weeks (in one trial only 5% of births were at a gestational age of more than 36 weeks whereas in other trials, less than one third of the population, 18% and 32% of births, were at a gestational age of 37 weeks or more).

The proportion of women receiving tocolysis and the timing to delivery since study entry were not clearly reported. One trial reported that 49% of women had received tocolysis in the 2 weeks preceding trial entry. One trial reported that 79% of women gave birth less than 24 hours after receiving a single repeat course. One trial reported that the interval between study entry and birth was significantly lower in the repeat corticosteroids arm than the control

(single course corticosteroids) arm. The number of repeat courses administered in the trials varied from one repeat course to more than six repeat courses, with the majority of women in the repeat corticosteroids arm of the trials receiving one or 2 repeat courses. Subgroup analysis performed for women received only one repeat course of antenatal corticosteroid.

Women with a multiple pregnancy were included in 9 out of 10 trials in the SR in a proportion ranging from 9% to 33%. The follow up study included 22.3% of women with a multiple pregnancy.

10.3.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The following GRADE profiles and subgroups analyses are presented:

- Table 43: GRADE profile for comparison of repeat course corticosteroids versus single course corticosteroids –
- Table 44: GRADE profile for comparison of repeat course corticosteroids versus single course corticosteroids – follow up study

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

1 **Table 43: GRADE profile for comparison of repeat course corticosteroids versus single course corticosteroids – RCTs**

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Repeat corticosteroids	Single course corticosteroids	Relative (95% CI)	Absolute (95% CI)	Quality
Fetal and neonatal mortality -											
1 meta-analysis of 9 studies (Crowther 2011)	randomised trials	no serious risk of bias[1]	no serious inconsistency	no serious indirectness	serious[2]	none	96/2791 (3.4%)	102/2763 (3.7%)	RR 0.94 (0.71 to 1.23)	2 fewer per 1000 (from 11 fewer to 8 more)	Moderate
Fetal and neonatal mortality (subgroup analysis for women with PPROM)											
1 study (Crowther 2011)	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	3/81 (3.7%)	6/79 (7.6%)	RR 0.49 (0.13 to 1.88)	39 fewer per 1000 (from 66 fewer to 67 more)	Very low
Fetal and neonatal mortality (subgroup analysis for one repeat course vs. single course of corticosteroids											
1 meta-analysis 3 studies (Crowther 2011)	randomised trials	no serious risk of bias[5]	no serious inconsistency	no serious indirectness	very serious[4]	none	14/504 (2.8%)	10/511 (2%)	RR 1.41 (0.64 to 3.08)	8 more per 1000 (from 7 fewer to 41 more)	Low
Need for mechanical ventilation-											
1 meta-analysis of 6 studies (Crowther 2011)	randomised trials	no serious risk of bias	serious[6]	no serious indirectness	serious[2]	none	556/2463 (22.6%)	668/2455 (27.2%)	RR 0.84 (0.71 to 0.99)	44 fewer per 1000 (from 3 fewer to 79 fewer)	Low
Chronic lung disease) [a]											
1 meta-analysis of 8 studies (Crowther 2011)	randomised trials	no serious risk of bias[7]	no serious inconsistency	no serious indirectness	serious[2]	none	181/2709 (6.7%)	170/2684 (6.3%)	RR 1.06 (0.87 to 1.3)	4 more per 1000 (from 8 fewer to 19 more)	Moderate
Chronic lung disease[a] (subgroup analysis for women with PPROM)											
1 study (Crowther 2011)	randomised trials	seriou[s3]	no serious inconsistency	no serious indirectness	very serious[4]	none	15/81 (18.5%)	19/79 (24.1%)	RR 0.77 (0.42 to 1.41)	55 fewer per 1000 (from 139 fewer to 99 more)	Very low
Chronic lung disease[a] (subgroup analysis for one repeat course vs. single course of corticosteroids)											
1 meta-analysis of 2 studies (Crowther 2011)	randomised trials	no serious risk of bias[5]	no serious inconsistency	no serious indirectness	serious[2]	none	42/432 (9.7%)	34/445 (7.6%)	RR 1.27 (0.83 to 1.96)	21 more per 1000 (from 13 fewer to 73 more)	Moderate
Intraventricular haemorrhage											
1 meta-analysis of 6 studies (Crowther 2011)	randomised trials	no serious risk of bias[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	129/1533 (8.4%)	137/1532 (8.9%)	RR 0.94 (0.75 to 1.18)	5 fewer per 1000 (from 22 fewer to 16 more)	High
Intraventricular haemorrhage - subgroup analysis for one repeat course vs. single course of corticosteroids											
1 meta-analysis of 2 studies (Crowther 2011)	randomised trials	no serious risk of bias[5]	serious[8]	no serious indirectness	very serious[4]	none	50/431 (11.6%)	52/441 (11.8%)	RR 0.99 (0.69 to 1.42)	1 fewer per 1000 (from 37 fewer to 50 more)	Very low
Intraventricular haemorrhage grade 3/4											

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Repeat corticosteroids	Single course corticosteroids	Relative (95% CI)	Absolute (95% CI)	Quality
1 meta-analysis of 6 studies (Crowther 2011)	randomised trials	no serious risk of bias[7]	no serious inconsistency	no serious indirectness	very serious[4]	none	32/2419 (1.3%)	28/2400 (1.2%)	RR 1.13 (0.69 to 1.86)	2 more per 1000 (from 4 fewer to 10 more)	Low
Periventricular leucomalaci[a]											
1 meta-analysis of 7 studies (Crowther 2011)	randomised trials	no serious risk of bias[7]	no serious inconsistency	no serious indirectness	very serious[4]	none	20/2453 (0.82%)	26/2435 (1.1%)	RR 0.77 (0.43 to 1.37)	2 fewer per 1000 (from 6 fewer to 4 more)	Low
Early systemic neonatal infection (defined by Crowther systemic infection suspected < 48h after birth; Mazumder – early onset sepsis; Peltoniemi – sepsis at age of < 3 days)											
1 meta-analysis of 3 studies (Crowther 2011)	randomised trials	no serious risk of bias9	no serious inconsistency	no serious indirectness	no serious imprecision	none	177/763 (23.2%)	193/781 (24.7%)	RR 0.93 (0.79 to 1.11)	17 fewer per 1000 (from 52 fewer to 27 more)	High
Birthweight adjusted for gestational age											
1 meta-analysis of 2 studies (Crowther 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision[10]	none	-0.40g ± 1.05	-0.27g ± 1.14	NC	MD 0.11 lower (0.23 lower to 0 higher)	High
Birthweight adjusted for gestational age - subgroup analysis for one repeat course vs. single course of corticosteroids											
1 study (Crowther 2011)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision[10]	none	-0.14g ± 0.86	-0.14g ± 0.98	NC	MD 0 higher (0.34 lower to 0.34 higher)	High
Major neurosensory disability at early childhood follow-up includes: developmental delay or intellectual impairment, blindness, deafness, or cerebral palsy after 18 months of age)[c]											
1 meta-analysis of 2 studies (Crowther 2011)	randomised trials	no serious risk of bias[5]	serious[6]	no serious indirectness	very serious[4]	none	56/613 (9.1%)	71/643 (11%)	RR 1.08 (0.31 to 3.76)	9 more per 1000 (from 76 fewer to 305 more)	Very low
Major neurosensory disability at early childhood follow-up - subgroup analysis for one repeat course vs. single course of corticosteroids											
1 study (Crowther 2011)	randomised trials	serious5	no serious inconsistency	no serious indirectness	very serious[4]	none	3/118 (2.5%)	1/139 (0.72%)	RR 3.53 (0.37 to 33.52)	18 more per 1000 (from 5 fewer to 234 more)	Very low
Any maternal side-effects of therapy											
1 meta-analysis of 2 studies (Crowther 2011)	randomised trials	no serious risk of bias[11]	very serious[12]	no serious indirectness	very serious[4]	None	115/739 (15.6%)	159/735 (21.6%)	RR 0.97 (0.24 to 3.9)	6 fewer per 1000 (from 164 fewer to 627 more)	Very low
Puerperal sepsis											
1 meta-analysis of 5 studies (Crowther 2011)	randomised trials	serious[13]	no serious inconsistency	no serious indirectness	serious[2]	none	72/1565 (4.6%)	61/1526 (4%)	RR 1.15 (0.83 to 1.60)	6 more per 1000 (7 fewer to 24 more)	Low
Puerperal sepsis (subgroup analysis for women with PPROM)											
1 study (Crowther 2011)	randomised trials	serious3	no serious inconsistency	no serious indirectness	very serious[4]	none	4/81 (4.9%)	6/79 (7.6%)	RR 0.65 (0.19 to 2.22)	27 fewer per 1000 (62 fewer to 93 more)	Very low
Puerperal sepsis - subgroup analysis for one repeat course vs. single course of corticosteroids											

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Repeat corticosteroids	Single course corticosteroids	Relative (95% CI)	Absolute (95% CI)	Quality
1 study (Crowther 2011)	randomised trials	serious5	no serious inconsistency	no serious indirectness	serious[2]	none	19/125 (15.2%)	12/124 (9.7%)	RR 1.57 (0.80 to 3.10)	55 more per 1000 (19 fewer to 203 more)	Low

CI confidence interval, RR relative risk, NC not calculable

[a] *Definitions of chronic lung disease varied among the studies: need for oxygen at 36 weeks; Crowther – need for oxygen at 36 weeks post conception; Garite – oxygen support at 30 days of life; Guinn – requiring oxygen and usually ventilator therapy for at least 28 days of life; Mazumder – bronchopulmonary dysplasia; Murphy – needing oxygen at a postmenstrual age of 36 completed weeks and radiographic scan compatible with bronchopulmonary dysplasia; Peltoniemi – bronchopulmonary dysplasia (supplemental oxygen or any form of ventilation with continuous distending pressures at postmenstrual age of 36 weeks or at postnatal age of 4 weeks for those born after postmenstrual age of 31 weeks; Wapner – bronchopulmonary dysplasia*

1 Three of the trials terminated recruitment early and one trial had no placebo control

2 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID

3 Trial terminated early, interim analysis suggested only 2% probability of detecting one-third reduction in composite morbidity: stopped at 502 women, planned to recruit 1000 women

4 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs

5 One trial was terminated early due to decrease in intact survival in repeat corticosteroids group; stopped at 249 women, planned to recruit 440 women

6 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%)

7 Two of the trials terminated early and one trial had no placebo control

8Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%)

9 One trial was terminated early and one trial had no placebo control so no blinding of participants or staff

10 Confidence interval smaller than half of the combined SD

11 One trial was terminated early due to tendency towards lower birthweights in the repeat corticosteroids group; stopped at 495 women, planned to recruit 2400 women

12 Evidence was downgraded by 2 due to very serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of >75%)

13 Three of the trials terminated recruitment early

Table 44: GRADE profile for comparison of repeat course corticosteroids versus single course corticosteroids – follow up study

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Repeat corticosteroids	Single course corticosteroids	Relative (95% CI)	Absolute (95% CI)	Quality
Child mortality up to 5 years of age (follow-up mean 5 years)											
1 study (Asztalos 2013)	observational studies	serious[1]	no serious inconsistency	no serious indirectness	Serious[3]	none	96/2791 (3.4%)	102/2763 (3.7%)	RR 0.94 (0.71 to 1.23)	2 fewer per 1000 (from 11 fewer to 8 more)	Very low
Neuromotor disability (nonambulatory cerebral palsy) at 5 years of age (follow-up mean 5 years)											

Quality assessment							Number of women/babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Repeat corticosteroids	Single course corticosteroids	Relative (95% CI)	Absolute (95% CI)	Quality
1 study (Asztalos 2013)	observational studies	serious1	no serious inconsistency	no serious indirectness	very serious[2]	none	4/287 (1.4%)	11/808 (1.4%)	RR 1.02 (0.33 to 3.19)	0 more per 1000 (from 9 fewer to 30 more)	Very low
Neurosensory disability (blindness, deafness, visual aids, hearing aids) at 5 years of age (follow-up mean 5 years)											
1 study (Asztalos 2013)	observational studies	serious1	no serious inconsistency	no serious indirectness	Serious[3]	none	70/827 (8.5%)	61/808 (7.5%)	RR 1.12 (0.81 to 1.56)	9 more per 1000 (from 14 fewer to 42 more)	Very low
Neurocognitive/ neurobehavioural disability (abnormal attention, memory or behaviour) at 5 years of age (follow-up mean 5 years)											
1 study (Asztalos 2013)	observational studies	serious1	no serious inconsistency	no serious indirectness	no serious imprecision	none	108/822 (13.1%)	109/793 (13.7%)	RR 0.96 (0.75 to 1.22)	5 fewer per 1000 (from 34 fewer to 30 more)	Low

Abbreviations: CI confidence interval, RR relative risk, NC not calculable
1 high risk of selection and attrition bias
2 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs
3 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID

Aratord 2015

This study was added to the review following final re-run searches for this review question (n=1348). The result of the trial was only reported by percentages. The total number of participants for each neonatal outcome was not reported and also it is not clear if the result is for babies born at term or just those born preterm. Therefore the result could not be meta-analysed with the other included trials in this review and should be interpreted withcaution. There was no significant difference in neonatal mortality between babies born to women who had repeat courses of corticosteroids (21.4%) and babies born to women who had a single course of corticosteroids (29.4% p>0.05). The evidence indicated that significantly fewer babies born to women who had repeat courses of corticosteroids required mechanical ventilation (27.7%) compared with babies who were born to women who had a single course of corticosteroids (39.6% p=0.002). The evidence was of very low quality across all outcomes.

10.3.4 Evidence statements

Neonatal outcomes

Moderate quality evidence from one meta-analysis of 9 RCTs (n=5554) suggested that there was no significant difference in fetal and neonatal mortality in babies born to women who had repeat courses of corticosteroids compared with babies who were born to women who had a single course of corticosteroids. Subgroup analysis in women with P-PROM and women who had one repeat course of corticosteroids had also showed no significant difference in fetal and neonatal mortality between the 2 groups. The evidence for the subgroup analysis was of very low quality.

Low quality evidence from one meta-analysis of 6 studies (n=4918) suggested that there were significantly fewer babies born to women who had repeat courses of corticosteroids required mechanical ventilation compared with babies who were born to women who had a single course of corticosteroids.

Moderate quality evidence from one meta-analysis of 8 RCTs (n=5393) suggested that there was no significant difference in the outcome of chronic lung disease in babies born to women who had repeat courses of corticosteroids compared with babies who were born to women who had a single course of corticosteroids. Subgroup analysis in women with P-PROM and women who had one repeat course of corticosteroids had also showed no significant difference in proportion of babies with chronic lung disease between the 2 groups.

High quality evidence from one meta-analysis of 6 RCTs (n=3065) suggested that there was no significant difference in intraventricular haemorrhage of any grade in babies born to women who had repeat courses of corticosteroids compared with babies who were born to women who had a single course of corticosteroids. Subgroup analysis in women with P-PROM and intraventricular haemorrhage of grades 3/4 only had also showed no significant difference in in proportion of babies with in intraventricular haemorrhage between the 2 groups.

Evidence from meta-analysis of several RCTs showed that there was no significant difference in periventricular leucomalacia (7 studies, n=4888), early systemic neonatal infection (3 studies, n=1544), birthweight adjusted for gestational age (2 studies, n=1256) or major neurosensory disability in early childhood (2 studies, n=4888) between babies born to women who had repeat courses of corticosteroids and babies born to women who had a single course of corticosteroids. The quality of the evidence ranged from high to very low across the outcomes.

Evidence from one observational study with over 6,000 participants showed that at 5 years of age, there was no significant difference in child mortality (very low quality evidence), neuromotor disability (very low quality evidence), neurosensory disability (very low quality evidence), or neurocognitive disability (low quality evidence) between babies born to women who had repeat courses of corticosteroids and babies born to women who had a single course of corticosteroids. There was no difference in maternal side effects (very low quality) or puerperal sepsis (low quality) between women who had repeat courses of corticosteroids and women who had a single course of corticosteroids.

Maternal outcomes

Very low quality evidence from one meta-analysis of 2 studies (n=1474) suggested no significant difference in maternal side effects between women who had repeat courses of corticosteroids and women who had a single course of corticosteroids. Low quality evidence from one meta-analysis of 5 studies (n=3091) suggested no significant difference in puerperal sepsis between women who had repeat courses of corticosteroids and women who had a single course of corticosteroids. Subgroup analysis of women with PPROM and women who had only one repeat course of corticosteroids had also showed no significant difference in puerperal sepsis between 2 groups.

10.3.5 Health economics profile

A single search was undertaken for health economic evidence on a single course or repeat courses of maternal corticosteroids for fetal lung maturation given at different gestations in improving preterm neonatal outcomes. A total of 136 articles were identified by the search. After reviewing titles and abstracts, 9 full papers but they were all excluded. Therefore no health economic evidence was found for this question.

This question was initially prioritised for health economic analysis although new analysis was ultimately not undertaken because there were more important topics to be addressed. Furthermore, corticosteroids are relatively cheap and offering a single course is current practice. The review did not find clinical evidence to suggest that repeat courses were beneficial and therefore there cannot be economic evidence that would justify the routine use of repeat courses. Health economic evidence statements

10.3.6 Evidence to recommendations

10.3.6.1 Relative value placed on the outcomes considered

The Committee selected the same outcomes for this review as it did for the review of the use of maternal corticosteroids at different gestational ages (see Section 10.2). The Committee prioritised measures that would be likely to indicate whether the drug had been successful in reducing pulmonary-specific adverse events such as need for mechanical ventilation and bronchopulmonary dysplasia/chronic lung disease, as well as other measures of outcome (mortality, all death up to one year, neurodevelopmental disability and intraventricular haemorrhage/periventricular leucomalacia/white matter injury). In addition it prioritised both birth weight for gestation and cognitive functioning in infancy or early childhood because it was aware of associations inconsistently observed previously in the literature between repeat steroid use and these adverse outcomes.

In terms of maternal outcomes the Committee prioritised mortality because it felt any change in the incidence of this outcome would impact on clinical decision making. It also prioritised all adverse events and agreed that these could be reported as a single outcome.

10.3.6.2 Consideration of clinical benefits and harms

The Committee considered that that there was very little evidence to support the use of repeat courses of maternal corticosteroids.

There was little difference between the groups that received repeat courses of corticosteroids and those that did not for most of the outcomes of chronic lung disease, intraventricular haemorrhage of any grade, periventricular leucomalacia, early systemic neonatal infection, birthweight adjusted for gestational age or major neurosensory disability in early childhood.

The Committee was concerned that trial investigators terminated some studies early on the basis of possible harm.

The only significant reported benefit of repeated courses of corticosteroids compared to a single course was a reduction in the need for mechanical ventilation although there is lack of precision in the estimate of effects and results should be interpreted with caution.

However, one of the included studies in the overall analysis did show benefit in terms of respiratory outcomes when the original course was given at 24 weeks and the repeat was given at 34 weeks.

Based on the evidence reviewed, the Committee concluded that there was insufficient evidence of benefit to support a recommendation that courses of steroids should be repeated routinely, but that this should not rule out the judicious use of repeat courses of corticosteroids in circumstances where clinical judgement suggested that it might be beneficial given the lack of clear evidence that such practice would cause harm. Decisions should be based upon gestation, likelihood of imminent birth, and time period since the last course of steroids of around 10 weeks.

10.3.6.3 Consideration of health benefits and resource uses

Offering repeat courses of maternal corticosteroids carries an opportunity cost. Furthermore, the Committee were not persuaded that there was sufficient clinical evidence to justify repeat courses and therefore a cost-effectiveness case for repeat courses of maternal corticosteroids has yet to be demonstrated.

10.3.6.4 Quality of evidence

The majority of evidence was of low to very low due to high risk of selection bias and imprecision. The Committee noted that the quality of the evidence varied across outcomes but concluded that the fact that some trials were stopped early was a cause for concern.

It also noted that the length of follow-up was not sufficient to identify all potential neurodevelopmental outcomes for the children of mothers being treated with single or repeated doses of corticosteroids.

10.3.6.5 Other considerations

The recommendations were based on both the evidence reviewed and the Committee's expert opinion.

10.3.7 Recommendations

30. For women between 23^{+0} and 23^{+6} weeks of pregnancy who are in suspected or established preterm labour, are having a planned preterm birth or have P-PROM (see Chapter 5), discuss with the woman (and her family members or carers as appropriate) the use of maternal corticosteroids in the context of her individual circumstances.

31. Consider maternal corticosteroids for women between 24^{+0} and 26^{+0} weeks of pregnancy who are in suspected or established preterm labour, are having a planned preterm birth or have P-PROM (see Chapter 5).

32. Offer maternal corticosteroids to women between 26^{+1} and 35^{+6} weeks of pregnancy who are in suspected, diagnosed or established preterm labour, are having a planned preterm birth or have P-PROM.

33. When offering or considering maternal corticosteroids, discuss with the woman (and her family members or carers as appropriate):

- how corticosteroids may help
- the potential risks associated with them.

34. Do not routinely offer repeat courses of maternal corticosteroids, but take into account:

- whether the interval since the end of last course is more than 10 weeks
- gestational age
- the likelihood of birth within 48 hours.

11 Magnesium sulfate for neuroprotection

11.1 Introduction

With advances in neonatal care in recent years, more babies born preterm are surviving and in particular those born at extremely preterm gestations now commonly survive the neonatal period (over 90% of those born before 28 weeks in high income countries) (Blencow 2013). These children frequently suffer long-term complications of prematurity. Neurological effects are common and may cause severe disability. They include cerebral palsy (which is associated with intraventricular haemorrhage and periventricular leukomalacia), developmental delay, cognitive problems, behavioural and learning difficulties. Agents to protect the developing fetal brain would therefore have great potential to reduce disability. Neuroprotection is offered by administration of maternal corticosteroids (see Chapter 11) and this chapter reviews the use of magnesium sulfate.

11.1.1 Review question

What is the clinical and cost effectiveness of magnesium sulfate given to women at high risk of giving birth preterm (defined as those suspected to be in preterm labour or diagnosed as being in preterm labour and those having planned preterm birth) for preventing cerebral palsy and other neurological disorders in babies born at different preterm gestations?

11.1.2 Description of included studies

Six studies were included in this review (Crowther 2003; Marret 2007; Marret 2008; Mittendorf 2002; Rouse 2008; Doyle 2014); 4 of them were RCTs and 2 studies (Marret 2008, Doyle 2014) were follow up studies of included RCTs (Marret 2007, Crowther 2003)

The setting of studies was France (Marret 2007; Marret 2008), USA (Mittendorf 2003; Rouse 2008), Australia and New Zealand (Crowther 2003; Doyle 2014).

All of the included studies evaluated the use of magnesium sulfate ($MgSO_4$) given to women at high risk of preterm labour, either because they were already thought to be in labour (with or without ruptured membranes) or they were having a planned preterm birth within 24 hours (only 1 study did not define the timing of planned preterm birth). Only trials where magnesium sulfate was given for the purposes of neuroprotection of the baby were included. One trial included women at over 24 but less than 34 completed weeks gestation (Mittendorf 2003), one trial included all women at less than 33 weeks gestation (Marret 2007), one trial included women at 24 to 31 weeks gestation (Rouse 2008), and one trial only included women at less than 30 weeks gestation (Crowther 2003). We reported separate results from sub-group analyses by gestation weeks at randomisation as a proxy to estimate any difference in the effect of magnesium sulfate by gestation of babies at birth. The administration of the intervention (magnesium sulfate) commenced from 16 to 36 weeks in pregnancy.

The initial (loading) dose of magnesium sulfate was 4 grams in 3 trials (Crowther 2003; Marret 2007; Mittendorf 2002) and 6 grams in one trial (Rouse 2008). In 2 trials, women then received a maintenance infusion of 1 gram per hour (Crowther 2003) or 2 grams per hour (Rouse 2008). (For full details see evidence tables in Appendix H).

11.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The following GRADE profiles present the comparison of magnesium sulfate with no magnesium sulfate:

- Table 45: GRADE profile for comparison of magnesium sulfate (MgSO4) with no magnesium sulfate
- Table 46: GRADE profile for comparison of magnesium sulfate (MgSO4) with no magnesium sulfate (long term child outcomes)
- Table 47: GRADE profile for comparison of magnesium sulfate (MgSO4) with no magnesium sulfate (maternal outcomes)

In 4 of the studies (Crowther 2003; Marret 2008; Rouse 2008; Doyle 2014) long-term outcomes (cerebral palsy, gross motor dysfunction, developmental delay, cognitive dysfunction, vision and hearing) were reported after excluding babies who died from the denominator. Therefore, the risks reported by the authors represent the risk of the outcome among children alive and available for follow-up, rather than being a reflection of the actual risk of the specific outcome occurring following the decision to administer magnesium sulfate or not (intention to treat analysis). Where possible, the denominator was changed to include the babies who died (stillbirths + neonatal deaths before discharge + neonatal/paediatric deaths after discharge) in the denominators, in order to provide an accurate reflection of long-term risk to inform decision making.

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

Table 45: GRADE profile for comparison of magnesium sulfate (MgSO4) with no magnesium sulfate (neonatal outcomes)

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	MgSO	No $MgSO_4$	Relative (95% CI)	Absolute (95% CI)	Quality
Stillbirth (randomised before 34 weeks)											
1 meta-analysis of 3 studies (Crowther 2003; Marret 2007; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	16/2160 (0.74%)	22/2214 (0.99%)	RR 0.74 (0.39 to 1.4)	3 fewer per 1000 (from 6 fewer to 4 more)	Moderate
Stillbirth (randomised before 30 weeks) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[3]	none	9/629 (1.4%)	11/626 (1.8%)	RR 0.81 (0.34 to 1.95)	3 fewer per 1000 (from 12 fewer to 17 more)	Very low
Neonatal mortality: before discharge											
1 meta-analysis of 3 studies (Crowther 2003; Marret 2007; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	187/2160 (8.7%)	195/2214 (8.8%)	RR 0.97 (0.8 to 1.18)	3 fewer per 1000 (from 18 fewer to 16 more)	Moderate
Neonatal mortality: before discharge (randomised before 30 weeks) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	76/629 (12.1%)	92/626 (14.7%)	RR 0.82 (0.62 to 1.09)	26 fewer per 1000 (from 56 fewer to 13 more)	Very low
Neonatal/paediatric mortality: between discharge and follow-up[a]											
1 meta-analysis of 2 studies (Crowther 2003; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	20/1808 (1.1%)	21/1878 (1.1%)	RR 1 (0.55 to 1.84)	0 fewer per 1000 (from 5 fewer to 9 more)	Moderate
Neonatal/paediatric mortality: between discharge and follow-up[a] (randomised before 30 weeks) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[3]	none	2/629 (0.32%)	4/626 (0.64%)	RR 0.5 (0.09 to 2.71)	3 fewer per 1000 (from 6 fewer to 11 more)	Very low
Total perinatal, neonatal and paediatric mortality[b]											
1 meta-analysis of 4 studies (Crowther 2003; Marret 2008; Mittendorf 2002; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	226/2190 (10.3%)	242/2243 (10.8%)	RR 0.95 (0.8 to 1.13)	5 fewer per 1000 (from 22 fewer to 14 more)	Moderate
Total perinatal, neonatal and paediatric mortality[b] (randomised before 30 weeks) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	87/629 (13.8%)	107/626 (17.1%)	RR 0.81 (0.62 to 1.05)	32 fewer per 1000 (from 65 fewer to 9 more)	Low
Total perinatal, neonatal and paediatric mortality[b] (Randomised at or after 28 weeks) [a] (subgroup analysis)											
1 study (Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[3]	None	21/599 (3.5%)	15/599 (2.5%)	RR 1.4 (0.73 to 2.69)	10 more per 1000 (from 7 fewer to 42 more)	Very low
Total perinatal, neonatal and paediatric mortality[b] (Randomised before 28 weeks)[a] (subgroup analysis)											
1 study (Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	None	78/442 (17.6%)	78/496 (15.7%)	RR 1.12	19 more per 1000	Low

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	MgSO	No $MgSO_4$	Relative (95% CI)	Absolute (95% CI)	Quality
									(0.84 to 1.49)	(from 25 fewer to 77 more)	
Grades III or IV intracranial haemorrhage[c] (findings on cranial ultrasound)											
1 meta-analysis of 3 studies (Crowther 2003; Mittendorf 2002; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	None	72/1738 (4.1%)	90/1799 (5%)	RR 0.81 (0.6 to 1.09)	10 fewer per 1000 (from 20 fewer to 5 more)	Low
Grades III or IV intracranial haemorrhage[c] (findings on cranial ultrasound) (randomised before 28 weeks) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[3]	None	49/596 (8.2%)	50/586 (8.5%)	RR 0.96 (0.66 to 1.4)	3 fewer per 1000 (from 29 fewer to 34 more)	Very low
Periventricular leukomalacia[c] (findings on cranial ultrasound)											
1 meta-analysis of 3 studies (Crowther 2003; Mittendorf 2002; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious[3]	None	44/1738 (2.5%)	48/1799 (2.7%)	RR 0.94 (0.63 to 1.4)	2 fewer per 1000 (from 10 fewer to 11 more)	Very low
Periventricular leukomalaciac (findings on cranial ultrasound) **(Randomised before 30 week [a]) (subgroup analysis)**											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Very serious[3]	None	22/596 (3.7%)	21/586 (3.6%)	RR 1.03 (0.57 to 1.85)	1 more per 1000 (from 15 fewer to 30 more)	Very low
Cerebral palsy: any[d]											
1 meta-analysis of 4 studies (Crowther 2003; Marret 2007; Mittendorf 2002; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Serious[2]	None	102/2130 (4.8%)	146/2184 (6.7%)	RR 0.71 (0.56 to 0.91)	19 fewer per 1000 (from 6 fewer to 29 fewer)	Low
Cerebral palsy: any[d] (randomised before 30 week a) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Very serious[3]	none	36/620 (5.8%)	42/621 (6.8%)	RR 0.86 (0.56 to 1.32)	9 fewer per 1000 (from 30 fewer to 22 more)	Very low

a. These are the deaths reported as occurring between the point of initial discharge and later follow-up (1 year in the case of Rouse 2008 and 2 years in the case of Crowther 2003. (Note: Because of the way the data are split between the trial and follow-up papers, these data are not reported for the Marret trial)
b. Deaths are reported up to the age of 1 year in Rouse 2008 and 2 years in Crowther 2003 and Marret 2008. There was no long-term follow-up in Mittendorf 2002 and the point at which deaths occurred is not reported.
c. Reported as a proportion of babies who received a cranial ultrasound
d. Mittendorf 2002 did not have long term follow-up. Follow-up was at 2 years in Crowther 2003, Marret 2008 and Rouse 2008.
e. Rouse 2008 reported this outcome for pregnancies rather than babies, and insufficient data are reported to convert it. Crowther 2003 reported the data for moderate and severe cerebral palsy separately, and these were pooled by the NCC-WCH technical team.
f. Crowther 2003 reported data for minimal and substantial gross motor dysfunction separately, and these were pooled by the NCC-WCH technical team. 18/616 (2.9%) of babies in the magnesium sulfate arm and 34/620 (5.5%) of babies in control arm had substantial gross motor dysfunction.
g. Developmental delay was defined according to Mental Development Index (MDI) scores. It was classified as: mild (MDI - 2 SDs to less than - 1 SD), moderate (MDI - 3 SDs to - 2 SDs) or severe (MDI < 3 SDs).
h. Children were classified as blind if their vision in both eyes was worse than 6/60
i. Children were classified as deaf if they required hearing aids
1 All trials in the meta-analysis included a proportion of women with multiple pregnancy

2 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID
3 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs

Table 46: GRADE profile for comparison of magnesium sulfate (MgSO4) with no magnesium sulfate (long term child outcomes)

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	MgSO	No $MgSO_4$	Relative (95% CI)	Absolute (95% CI)	Quality
Cerebral palsy: moderate or severe (at 2 years)[e]											
1 meta-analysis of 2 studies (Crowther 2003; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Serious[2]	none	35/1661 (2.1%)	59/1715 (3.4%)	RR 0.61 (0.4 to 0.92)	13 fewer per 1000 (from 3 fewer to 21 fewer)	Moderate
Cerebral palsy: moderate or severe (at 2 years)[e] (randomised before 30 week) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Very serious[3]	none	15/620 (2.4%)	21/620 (3.4%)	RR 0.71 (0.37 to 1.37)	10 fewer per 1000 (from 21 fewer to 13 more)	Very low
Cerebral palsy: moderate or severe (at 2 years)[e] (randomised at or after 28 week) (subgroup analysis)											
1 study (Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Very serious3	none	8/599 (1.3%)	8/599 (1.3%)	RR 1 (0.38 to 2.65)	0 fewer per 1000 (from 8 fewer to 22 more)	Very low
Cerebral palsy: moderate or severe (at 2 years)[e] (randomised before 28 week[a]) (subgroup analysis)											
1 study (Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Serious[2]	none	12/442 (2.7%)	30/496 (6%)	RR 0.45 (0.23 to 0.87)	33 fewer per 1000 (from 8 fewer to 47 fewer)	Low
Cerebral Palsy: school-age (6-11 years of age)											
1 study (Doyle 2014)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Very serious[3]	none	23/295 (7.8%)	21/314 (6.7%)	RR 1.17 (0.66 to 2.06)	11 more per 1000 (from 23 more to 71 more)	Very Low
Gross motor dysfunction (at 2 years)[f]											
1 meta-analysis of 2 studies (Crowther 2003; Marret 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	157/963 (16.3%)	171/951 (18%)	RR 0.91 (0.74 to 1.1)	16 fewer per 1000 (from 47 fewer to 18 more)	Low
Gross motor dysfunction (at 2 years)f (randomised before 30 week) (subgroup analysis)											
Crowther 2003	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	Very serious[3]	none	102/616 (16.6%)	107/620 (17.3%)	RR 0.96 (0.75 to 1.23)	7 fewer per 1000 (from 43 fewer to 40 more)	Very low
Motor function: school-age (6-11 years of age)											
1 study (Doyle 2014)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	80/297 (26.9%)	80/300 (26.7%)	RR 1.01 (0.77 to 1.32)	267 fewer per 1000 (from 267	Low

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	MgSO	No $MgSO_4$	Relative (95% CI)	Absolute (95% CI)	Quality
										more to 267 more)	
Developmental delay: any (at 2 years)[g]											
1 study (Crowther 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	none	176/581 (30.3%)	170/585 (29.1%)	RR 1.04 (0.87 to 1.24)	12 more per 1000 (from 38 fewer to 70 more)	Moderate
Cognitive dysfunction (at 2 years)											
1 study (Marret 2008)	randomised trial	serious[4,5,6]	no serious inconsistency	serious[7]	serious[2]	none	57/347 (16.4%)	62/331 (18.7%)	RR 0.88 (0.63 to 1.22)	22 fewer per 1000 (from 69 fewer to 41 more)	Very low
Vision: blindness (at 2 years)[h]											
1 study (Crowther 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	none	1/620 (0.16%)	1/621 (0.16%)	RR 1 (0.06 to 15.98)	0 fewer per 1000 (from 2 fewer to 24 more)	Moderate
Hearing: deafness (at 2 years)[i]											
1 study (Crowther 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	none	8/620 (1.3%)	7/621 (1.1%)	RR 1.14 (0.42 to 3.14)	2 more per 1000 (from 7 fewer to 24 more)	Moderate

MgSO4 magnesium sulfate, CI confidence interval, RR relative risk, NC not calculable

a. These are the deaths reported as occurring between the point of initial discharge and later follow-up (1 year in the case of Rouse 2008 and 2 years in the case of Crowther 2003. (Note: Because of the way the data are split between the trial and follow-up papers, these data are not reported for the Marret trial)

b. Deaths are reported up to the age of 1 year in Rouse 2008 and 2 years in Crowther 2003 and Marret 2008. There was no long-term follow-up in Mittendorf 2002 and the point at which deaths occurred is not reported.

c. Reported as a proportion of babies who received a cranial ultrasound

d. Mittendorf 2002 did not have long term follow-up. Follow-up was at 2 years in Crowther 2003, Marret 2008 and Rouse 2008.

e. Rouse 2008 reported this outcome for pregnancies rather than babies, and insufficient data are reported to convert it. Crowther 2003 reported the data for moderate and severe cerebral palsy separately, and these were pooled by the NCC-WCH technical team.

f. Crowther 2003 reported data for minimal and substantial gross motor dysfunction separately, and these were pooled by the NCC-WCH technical team. 18/616 (2.9%) of babies in the magnesium sulfate arm and 34/620 (5.5%) of babies in control arm had substantial gross motor dysfunction.

g. Developmental delay was defined according to Mental Development Index (MDI) scores. It was classified as: mild (MDI - 2 SDs to less than - 1 SD), moderate (MDI - 3 SDs to - 2 SDs) or severe (MDI < 3 SDs).

h. Children were classified as blind if their vision in both eyes was worse than 6/60

i. Children were classified as deaf if they required hearing aids

1 All trials in the meta-analysis included a proportion of women with multiple pregnancy

2 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID

3 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs

Table 47: GRADE profile for comparison of magnesium sulfate (MgSO4) with no magnesium sulfate (maternal outcomes)

Quality assessment							Number of women or babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	$MgSO^4$	No $MgSO^4$	Relative (95% CI)	Absolute (95% CI)	Quality
Maternal death											
1 meta-analysis of 3 studies (Crowther 2003; Marret 2007; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	very serious [4]	none	0/1917 (0%)	1/1950 (0.05%)	RR 0.32 (0.01 to 7.92)	0 fewer per 1000 (from 1 fewer to 4 more)	Very low
Maternal adverse effects: any											
1 meta-analysis of 2 studies (Crowther 2003; Rouse 2008)	randomised trials	serious[9]	serious[3]	serious[1]	no serious imprecision	none	1309/1613 (81.2%)	339/1652 (20.5%)	RR 3.82 (1.38 to 10.59)	579 more per 1000 (from 78 more to 1000 more)	Low
Maternal adverse effects: leading to stopping of infusion											
1 meta-analysis of 2 studies (Crowther 2003; Rouse 2008)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	none	123/1613 (7.6%)	44/1652 (2.7%)	RR 2.81 (2.01 to 3.93)	48 more per 1000 (from 27 more to 78 more)	Moderate
Maternal adverse effects: cardiac or respiratory arrest											
1 meta-analysis of 2 studies (Crowther 2003; Marret 2007)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	None	0/821 (0%)	0/805 (0%)	NC	NC	Moderate
Maternal adverse effects: drop in diastolic blood pressure of more than 15 mmHg											
1 study (Crowther 2003)	randomised trial	no serious risk of bias	no serious inconsistency	serious[5]	serious[2]	None	77/535 (14.4%)	52/527 (9.9%)	RR 1.46 (1.05 to 2.03)	45 more per 1000 (from 5 more to 102 more)	Low
Maternal adverse effects: hypotension											
1 study (Marret 2007)	randomised trial	serious12	no serious inconsistency	serious[7]	very serious[4]	None	3/286 (1%)	0/278 (0%)	RR 6.8 (0.35 to 131.14)	NC	Very low

MgSO4 magnesium sulfate, CI confidence interval, RR relative risk, NC not calculable
a. These are the deaths reported as occurring between the point of initial discharge and later follow-up (1 year in the case of Rouse 2008 and 2 years in the case of Crowther 2003. (Note: Because of the way the data are split between the trial and follow-up papers, these data are not reported for the Marret trial)
b. Deaths are reported up to the age of 1 year in Rouse 2008 and 2 years in Crowther 2003 and Marret 2008. There was no long-term follow-up in Mittendorf 2002 and the point at which deaths occurred is not reported.
c. Reported as a proportion of babies who received a cranial ultrasound
d. Mittendorf 2002 did not have long term follow-up. Follow-up was at 2 years in Crowther 2003, Marret 2008 and Rouse 2008.
e. Rouse 2008 reported this outcome for pregnancies rather than babies, and insufficient data are reported to convert it. Crowther 2003 reported the data for moderate and severe cerebral palsy separately, and these were pooled by the NCC-WCH technical team.

f. Crowther 2003 reported data for minimal and substantial gross motor dysfunction separately, and these were pooled by the NCC-WCH technical team. 18/616 (2.9%) of babies in the magnesium sulfate arm and 34/620 (5.5%) of babies in control arm had substantial gross motor dysfunction.
g. Developmental delay was defined according to Mental Development Index (MDI) scores. It was classified as: mild (MDI - 2 SDs to less than - 1 SD), moderate (MDI - 3 SDs to - 2 SDs) or severe (MDI < 3 SDs).
h. Children were classified as blind if their vision in both eyes was worse than 6/60
i. Children were classified as deaf if they required hearing aids
1 All trials in the meta-analysis included a proportion of women with multiple pregnancy
2 Evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID
3 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%)
4 Evidence was downgraded by 2 due to very serious imprecision as 95% confidence interval crossed 2 default MIDs
5 16.7% of women had a multiple pregnancy
6 21.6% of women had a multiple pregnancy
7 16.7% of women had a multiple pregnancy

11.1.4 Evidence statements

11.1.4.1 Overall analysis

Moderate quality evidence from 3 RCTs with a total sample size of over 3000 babies showed that those whose mothers were treated with Mg sulfate during pregnancy were significantly less likely to be diagnosed with cerebral palsy (CP) at birth compared to those who were not treated. However, subgroup analysis demonstrated that this protective effect of Mg sulfate on the risk of cerebral palsy was not significant for those who were randomised before 30 weeks gestation (very low quality evidence). No other neonatal outcome (still birth, neonatal, perinatal and paediatric mortality, intracranial haemorrhage and periventricular leukomalaciac) was found significantly different between the groups.

The risk of cerebral palsy at 2 years was also found significantly lower for those children whose mothers were treated with Mg sulfate during pregnancy compared to control group (moderate quality) and this finding remained significant in the subgroup analysis by gestational age only for the subgroup of children whose mothers were randomised at the intervention/control arms before 28 weeks of gestation. No other long term developmental outcome (gross motor dysfunction, developmental delay, hearing and vision difficulties) was found significantly different children whose mothers received treatment with Mg sulfate and control groups (moderate to very low quality evidence).

Low to moderate quality evidence from 3 RCTs with a total sample size of over 3000 women showed that significantly more women who were treated with Mg sulfate were likely to experience adverse effects (including any adverse effect, leading to stopping of infusion or dropping of diastolic pressure by more than 15mmHg) compared to those who were not treated. No significant difference was found for the maternal outcome of hypotension (very low quality) between the 2 groups.

No evidence was found on studies comparing the different effectiveness of different doses of magnesium sulfate.

11.1.5 Health economics profile

This question was prioritised for health economic analysis.

A systematic search found 2 studies (Cahill 2011; Bickford 2013) which considered the cost effectiveness of magnesium for neuroprotection in women with imminent or threatened preterm labour. These studies both reported that magnesium sulfate for neuroprotection dominated no magnesium sulfate for neuroprotection, meaning that savings in the cost of adverse outcomes more than offset treatment costs and that the intervention resulted in gains in health related quality of life. These studies are reported in more detail in Appendix H.

In addition an original health economic model was developed using the evidence from the clinical review undertaken for this guideline on neonatal mortality, cerebral palsy, periventricular leukomalacia and intraventricular haemorrhage. This took the form of a cost-utility analysis and compared magnesium sulfate for neuroprotection against no magnesium sulfate for neuroprotection in women with between 24^{+0} and 34^{+0} weeks of pregnancy and at high risk of preterm birth. The base case analysis reached the same conclusion as the 2 identified published studies, that magnesium sulfate for neuroprotection dominated no magnesium sulfate being cheaper when considering lifetime costs and also offering increased in health related quality of life. Probabilistic sensitivity analysis suggested that magnesium sulfate for neuroprotection had an 86% probability of being cost effective when compared against no magnesium sulfate.

Given this finding, sensitivity analyses were designed to subject this conclusion to challenge, for example by finding the threshold for input parameters when magnesium sulfate for

neuroprotection would cease to be cost-effective even if the input value fell outside a plausible range. These sensitivity analyses found that the base case input values were markedly below these thresholds for cost-effectiveness, suggesting that the model results were robust with respect to uncertainty not directly related to treatment effect size, the uncertainty of which was assessed with probabilistic sensitivity analysis using Monte Carlo simulation.

The model is described in detail in Chapter 16.

11.1.6 Evidence to recommendations

11.1.6.1 Relative value placed on the outcomes considered

The Guideline Committee selected both any form of neonatal mortality and stillbirth as the priority outcomes for this review. The rationale for this was that the mechanisms leading to stillbirth might be different from those that cause death after birth and therefore, the Committee felt that it was feasible that the drug might be more beneficial or harmful in relation to one outcome and not necessarily for the other.

Although the Committee considered many different developmental outcomes for this review question, it ultimately concluded that a reduction in the incidence of cerebral palsy would be the most useful measure. It was hypothesised that if there was a significant reduction in the risk of cerebral palsy in a RCT, it would be reasonable to attribute this difference to the action of magnesium sulfate. They acknowledged that cerebral palsy covers a wide spectrum of disability and therefore a reduction in the more severe forms would show a particular benefit for the use of magnesium sulfate. Another concern that prevented the Committee from prioritising other developmental outcomes over cerebral palsy was that assessment techniques used in research to measure the severity of such developmental problems are often unreliable and vary from study to study.

In terms of maternal outcomes the Committee felt that it was important to consider maternal adverse effects as well as maternal mortality because in their clinical experience there is an appreciable incidence of adverse maternal effects from the use of magnesium sulfate. They did not, however, think it would be particularly helpful to know what proportion of women suffering adverse effects had chosen to stop the infusion because they thought that any results might be affected by the trial setting. In addition, this outcome would vary from what might be expected in usual clinical circumstances. This is due to the fact that women who are taking part in a trial would be likely to have access to a higher level of information and support than might be the case outside of a trial setting and thus be more likely to tolerate adverse effects. Also, in their experience, women would tolerate a high level of adverse effects if they felt that this was likely to improve outcomes for the baby.

11.1.6.2 Consideration of clinical benefits and harms

The Committee noted that there was consistent evidence of benefit to the baby in terms of significantly lowering the risk of cerebral palsy resulting from the use of magnesium sulfate. Although there was overall evidence of benefit when women with babies up to 34 weeks gestation being included in these trials, subgroup analysis by gestational age showed significant benefit only for babies born before 28 weeks. However, the Committee decided to recommend magnesium sulfate for babies born before 34 weeks in the absence of evidence of harm and in the belief that this would widen any potential long term benefit to a larger group of babies.

The Committee also noted that the use of magnesium sulfate showed no difference in effect between the experimental and control groups in terms of perinatal, neonatal and paediatric mortality and concluded from this that the use of magnesium sulfate is not associated with a higher risk of harm to the baby.

With regards to maternal outcomes, limited available data, showed no firm conclusions about the risk of cardiac or respiratory arrest or maternal mortality. However, the Committee was noted that there was some evidence of harm to the woman treated with magnesium sulfate in terms of adverse effects and these results were in keeping with their clinical experience. While they acknowledged that the results showed a statistically significant difference in blood pressure, with a drop of more than 15mmHg for women receiving magnesium sulfate compared to controls, the Committee were uncertain as to whether this difference is significant in clinical terms. These findings support current practice that women being treated with magnesium sulfate should continue to routinely have their vital signs monitored.

The Committee observed that, in their clinical experience, different women had different responses including adverse events to magnesium sulfate and therefore it was difficult to capture the real balance of risks and benefits for the individual pregnant woman but that there would always need to be some clinical judgement involved about the decision to initiate the treatment of magnesium sulfate.

In light of all of these considerations the Committee felt that the benefits to the baby outweighed the potential harms to the woman and that the evidence supported a strong recommendation to offer magnesium sulfate for neonatal neuroprotection to women in labour at or before 32 weeks gestation, and women who were at or before 32 weeks gestation having a planned preterm birth within the next 24 hours.

The Committee acknowledged that the review had not been designed to look at the comparative effectiveness of different doses of magnesium sulfate. Given the lack of evidence for a specific dose of this treatment, it would be preferable and pragmatic in terms of improving safety and reducing the likelihood of errors being made, if the recommendations could be consistent with other established drug protocols. In light of this the Committee decided to recommend the dose of magnesium sulfate used in the treatment of pre-eclampsia (4 grams bolus, 1 gram per hour IV) because this seemed both clinically applicable and was the one used in many of the studies.

The Committee also concluded that the potential risk of harm to the woman warranted further recommendations being made to ensure that appropriate monitoring of this treatment is carried out. For the reasons given above, it was agreed that maternal and fetal monitoring should be consistent with the monitoring protocols for magnesium toxicity used in women who are receiving magnesium sulfate for the treatment of pre-eclampsia. In addition, the frequency of monitoring for toxicity in the presence of symptoms of oliguria or other signs of renal failure and adjustment on magnesium's toxicity was noted by the Committee as a separate recommendation.

The Committee also discussed whether repeated doses of magnesium sulfate should be offered and, if so, whether both the bolus and/or the intravenous infusion should be repeated as these were areas of clinical uncertainty in current practice. The Committee noted that the review had not been designed to look at the effectiveness of repeat courses and so did not make any recommendations to this effect, but were aware that repeated administration sometimes happens in practice.

11.1.6.3 Consideration of health benefits and resource uses

Magnesium sulfate is a relatively inexpensive treatment and the evidence identified in the SR demonstrated a significant clinical benefit to the baby in terms of a significantly reduced incidence of cerebral palsy.

The Committee were aware that the need to monitor women who are being treated with magnesium sulfate makes the intervention more costly than the cost of the drug alone.

Nevertheless, given that the management of cerebral palsy is extremely costly, the Committee considered that overall the health benefits not only justified the resource use, but

that the initial costs incurred would be likely to be off-set by large cost savings downstream and the health economic model produced for this guideline provided support for this viewpoint.

11.1.6.4 Quality of evidence

The majority of randomised evidence included for this review was of moderate to low quality and not overly subject to bias.

The Committee noted that blinding of the participants and the assessors would be difficult because the adverse effects of magnesium sulfate can be very obvious in some women. This concern was mitigated somewhat by the fact that some adverse effects were also reported in the placebo group. Overall the Committee felt that neither the concerns about blinding nor the observed placebo effects would change their interpretation of the results.

11.1.6.5 Other considerations

It was also noted that the length of follow-up in the trials (2 years) meant that the results were unlikely to capture the impact of the drug on cognitive function as this aspect of development is unlikely to become clear until the child reaches school age. Therefore, the interpretation of results in the short term should be interpreted with caution.

11.1.7 Recommendations

35. Offer intravenous magnesium sulfate for neuroprotection of the baby to women between 24^{+0} and 34^{+0} weeks of pregnancy who are:

- in established preterm labour or
- having a planned preterm birth within 24 hours.

36. Give a 4 g intravenous bolus of magnesium sulfate over 15 minutes, followed by an intravenous infusion of 1 g per hour until the birth or for 24 hours (whichever is sooner).

37. For women on magnesium sulfate, monitor for clinical signs of magnesium toxicity at least every 4 hours by recording pulse, blood pressure, respiratory rate and deep tendon (for example, patellar) reflexes.

38. If a woman has or develops oliguria or other signs of renal failure:

- monitor more frequently for magnesium toxicity
- think about reducing the dose of magnesium sulfate.

11.1.8 Research recommendations

Research question	5. What is the clinical effectiveness of a bolus plus infusion of magnesium sulfate compared with a bolus alone for preventing neurodevelopmental injury in babies born preterm?
Why this is needed	
Importance to 'patients' or the population	There is evidence from randomised studies that magnesium sulfate has neuroprotective properties for the baby when given to women who will deliver preterm up to 34+0 weeks of pregnancy. However, there is uncertainty about the best method of administering magnesium sulfate for this purpose, with different studies using different strategies. There are significant advantages for the woman and for reducing healthcare costs if a bolus is as effective as a bolus

Research question	5. What is the clinical effectiveness of a bolus plus infusion of magnesium sulfate compared with a bolus alone for preventing neurodevelopmental injury in babies born preterm?
	plus infusion, because magnesium sulfate has side effects for the woman, and more monitoring is needed for infusion, with additional associated healthcare costs. A randomised controlled trial would best address this question by assessing the effects of each method on neonatal and maternal outcomes.
Relevance to NICE guidance	The current guideline recommends the use of magnesium sulfate for neuroprotection, and the research is unlikely to change the recommendation, but will address the best protocol for administration.
Relevance to the NHS	If a bolus is as effective as a bolus plus infusion this will reduce healthcare costs, because additional monitoring is needed for infusion
National priorities	N/A
Current evidence base	There is existing evidence of the effectiveness of magnesium sulfate in neuroprotection but uncertainty about the best method of administration, with different studies using different protocols.
Equality	The population is defined as pregnant woman at imminent risk of preterm birth before 34+0 weeks gestational age.
Feasibility	The Committee decided that a RCT would best address this research recommendation. This could realistically be carried out within a reasonable timescale and cost. There are no particular ethical issues beyond other perinatal trials, and no technical issues since participating centres will have availability of monitoring.
Other comments	N/A

1

2

12 Tocolysis

12.1 Introduction

This review considers the clinical and cost effectiveness of medicines given to women who are in suspected or diagnosed preterm labour with the aim of delaying birth to improve outcomes. By definition a drug which stops or delays the progress of labour once it is believed to have started is a 'tocolytic'. However, which drugs should be classified as tocolytic is debated. Drugs which are used to prevent preterm labour are not considered tocolytics. Progesterone is a good example of a drug which is normally used as a prophylactic agent for women at risk of preterm labour during the antenatal period (see Section 4.2). However, progesterone is also used increasingly in the intrapartum context for its potential 'tocolytic effect'. In contrast, ethanol, one of the first agents used as a tocolytic, would no longer be considered a therapeutic option even if found to be effective for neonatal outcomes because of known maternal side effects. Therefore, this chapter will review the relative effectiveness of all medicines which have been used to delay or stop preterm labour and will refer to them as tocolytics.

There is currently variation in clinical practice with respect to the use of tocolytic medicines both in terms of the choice of medicine and the selection of population that receives treatment (all pregnant women in preterm labour or a selected sub-group).

Where multiple treatment options exist, it is very difficult to determine which intervention is most effective in improving outcomes based on the results of conventional pair-wise meta-analyses of direct evidence. The challenge of interpretation of direct evidence for assessing the most effective intervention for improving outcomes arises for 2 reasons:

- some pairs of alternative interventions may not have not been directly compared in a RCT (for example, in the case of tocolytics there are no studies that have directly comparing oxytocin receptor blocker to magnesium sulfate)
- a head-to-head analysis usually only provides information about the relative effect of a maximum of 2 treatments, it does not provide an estimate of the relative effects across multiple treatment options.

To overcome these issues, Mixed Treatment Comparison (MTC) meta-analytic techniques, also termed NMA were performed. Advantages of performing this type of analysis are:

- It allows the synthesis of data from direct and indirect comparisons without breaking randomisation, to produce measures of treatment effect and ranking of different interventions. If treatment A has never been compared against treatment B head to head, but these 2 interventions have been compared to a common comparator (treatment C), then an indirect treatment comparison can use the relative effects of the 2 treatments versus the common comparator. This is also the case whenever there is a path linking 2 treatments through a set of common comparators. All the randomised evidence is considered within the same model. NMA is a generalisation of standard pairwise meta-analysis for A versus B trials, to data structures that include, for example, A versus B, B versus C, and A versus C trials. A basic assumption of NMA methods is that direct and indirect evidence estimate the same parameter, that is, the relative effect between A and B measured directly from an A versus B trial, is the same as the relative effect between A and B estimated indirectly from A versus C and B versus C trials. This is often termed the consistency assumption and should be assessed and taken into account when interpreting the results of a NMA. NMA techniques strengthen inference concerning the relative effect of 2 treatments by including both direct and indirect comparisons between treatments, and, at the same time, allow simultaneous inference on all treatments while respecting randomisation.

- For every intervention in a connected network, a relative effect estimate (with its 95% credible intervals) can be estimated versus any other intervention. These estimates provide a useful clinical summary of the results and facilitate the formation of recommendations based on all of the best available evidence, whilst appropriately accounting for uncertainty. Furthermore, these estimates will be used to parameterise treatment effectiveness in the de novo cost-effectiveness modelling.

For details of the methods, results and interpretation of the NMA, see 'Methods' below and Appendix J.

Given the interventional nature of this review question, only RCTs of women at high risk of or suspected to be in preterm labour that compared different interventions as tocolytics between each other or between a tocolytic and a placebo or usual care for delaying preterm delivery were considered. Trials that compared combination interventions as tocolytics were excluded for the scope of this review question. The Guideline Committee decided upon the selection of following outcomes (maternal and neonatal) and the hierarchy of their importance for decision making (the smaller the number, the higher this outcome's importance)

1. Maternal mortality
2. Neonatal mortality
3. Perinatal mortality
4. Maternal infection
5. Delay of birth by at least 48 hours
6. Neonatal sepsis
7. Chronic lung disease/bronchopulmonary dysplasia (CLD)
8. Intraventricular haemorrhage (IVH)
9. Mothers with adverse events requiring cessation of treatment
10. Neurodevelopmental disability (combined outcome including: developmental delay, intellectual, gross motor, visual or hearing impairment; cerebral palsy; learning difficulties)
11. Periventricular leucomalacia (PVL)/white matter injury
12. Gestational age at birth
13. Respiratory distress syndrome (RDS)
14. Quality of life

However, given the paucity of data for some outcomes and the time constraints in the guideline development, the Committee prioritised the following outcomes for the NMA:

- neonatal mortality
- perinatal mortality
- RDS
- IVH
- mothers with adverse events requiring cessation of treatment
- delay of birth by at least 48 hours
- neonatal sepsis
- gestational age at birth

A class effect model was adopted for the new NMA because it was hypothesised that treatments within class would borrow similar clinical characteristics and mechanisms of effect. In other words, results for one member of the class in relation to efficacy and side effects were considered to be generalisable to other members of that same class. For that reason, trials with non UK licensed interventions were included in the NMA to allow the maximum use of available evidence and borrow strength of loops in the network only if there

was another trial that included licensed (for pre-term labour or for other conditions) interventions for the same class. The Committee confirmed a priori that they would only consider for decision making UK licensed medicines from each class depending on the clinical and cost effectiveness analysis. Some other considerations in the design of the NMA:

- the Committee discussed that although dosage, mode of administration and timing of treatment may influence the effectiveness of different tocolytics interventions, it was considered unlikely for this factor to change the direction of relative effect for the different interventions tested in the analysis. Therefore, the Committee decided not to consider any confounding effect of these factors in the NMA
- some of the included studies examined medicines that are not licensed as tocolytics for use in pregnancy (including nylidrin and barusiban). These medicines were included in the NMA to increase the size of the network, and because it is not uncommon for medicines that are not licenced for pregnancy indications to be prescribed for use in this context
- it is anticipated that the economic analysis will be based on the lowest cost treatment in a class that the Committee would be willing to recommend
- the Committee decided to have separate classes for alcohol/ethanol and combination treatments (classed as 'other') in the new NMA. The only study that considered the hormone human chorionic gonadotropin (HCG) was excluded because it is not used in current practice and did not form a loop to any other trials in the network. Both the new separate alcohol/ethanol and 'others' classes were removed from final ranking and health economic analysis as the Committee did not aim to consider making recommendations about these treatment options
- placebo was used as the reference treatment in the NMA as there is no universally recognised "standard" tocolytic. The effect in the placebo group is used as an indicator for the effect when a tocolytic has not been administered

Further details on the protocol for this review question are given in Appendix D.

12.1.1 Review question

What is the clinical and cost effectiveness of tocolytics given to women with suspected or diagnosed preterm labour to improve outcomes:

- **progesterone/progestogens**
- **beta-sympathomimetics**
- **oxytocin receptor antagonists**
- **calcium channel blockers**
- **cyclo-oxygenase enzyme inhibitors**
- **non-steroidal anti-inflammatory drugs**
- **nitric oxide donors**
- **magnesium sulfate?**

12.1.2 Description of included evidence

Nine studies were included for this review question overall: a SR and NMA (Haas 2012) of 95 RCTs from a variety of settings, and 8 individual primary RCTs; one from the Netherlands (Houtzager 2006) one from India (Jaju 2011), one from Israel (Salim 2012), 2 from the USA (Klauser 2012; Klauser 2014), and 3 from Iran (Kashanian 2014; Nankali 2014; Nikbakht 2014).

12.1.2.1 Summary of included studies

Details on the nature of included studies and their own characteristics are given in the following table (Table 48). The mean gestational age of women across all included studies

was 26 weeks and the population was predominately women with no P-PROM. For full details of included studies, see Evidence Tables in Appendix H.

Table 48: Summary of included studies

Included studies	Type of study design/N (sample size)	Interventions/number of studies	Outcomes
Haas 2012	SR and NMA of 95 RCTs (N=10860)	• betamimetics (ritodrine, terbutaline, nylidrin, salbutamol, fenoterol, hexoprenaline, isoxsuprine) (60 RCTs) • calcium channel blockers (nifedipine, nicardipine) (29 RCTs) • magnesium sulfate (29 RCTs) • nitrates (nitroglycerin, nitric oxide) (4 RCTs) • oxytocin receptor blockers (atosiban, barusiban) (13 RCTs) • others (alcohol, human chorionic gonadotropin) (5 RCTs) • prostaglandin inhibitors (indomethacin, indomethacin plus alcohol*, celecoxib, sulindac, ketorolac, rofecoxib) (18 RCTs) • placebo (25 RCTs)	• delayed delivery by 48 hours (primary outcome) • neonatal mortality • neonatal respiratory distress syndrome • maternal adverse events (all cause)
Klauser 2012 Klauser 2014 (further analysis of Klauser 2012)	RCT (N=301)	• prostaglandin inhibitors (indomethacin) • magnesium sulfate • calcium channel blockers (nifedipine)	• neonatal adverse effects
Houtzager 2006	Follow up study of Paptsonis 1997 and 2000 included in Haas 2012 (N=102)	• calcium channel blockers (nifedipine) • betamimetics (ritodrine)	• long-term psychosocial and motor effects on children
Jaju 2011	RCT (N=210)	• calcium channel blockers (nifedipine) • betamimetics (ritodrine)	• adverse events
Salim 2012	RCT (N=145)	• calcium channel blockers (nifedipine) • oxytocin receptor blockers (atosiban)	• adverse events
Kashanian 2014	RCT (N=120)	• calcium channel blockers (nifedipine) • nitrates (nitro-glycerine [NG])	
Nankali 2014	RCT (N=84)	• Nitrates (glyceryl trinitrate [GTN])	

Included studies	Type of study design/N (sample size)	Interventions/number of studies	Outcomes
		• Placebo	
Nikbakht 2014	RCT (N=100)	• magnesium sulfate nifedipine • placebo	• efficacy

** Indomethacin plus alcohol was not detailed in the published paper but was included in Haas's analysis and is recorded here for completeness.*

12.1.3 Introduction to the new network meta-analysis

A standard network meta-analysis (NMA) model in WinBUGS was carried out by the NICE Technical Support Unit (TSU) at Bristol University with the support of technical team at NCC-WCH.

The NMA for this review analysis was structured around the database of the published NMA (Haas 2012) with the following changes:

- new data from the additional studies (Klauser 2012; Klauser 2014; Jaju 2011; Salim 2012; Kashanian 2014; Nankali 2014; Nikbakht 2014) that were not originally included in Haas 2012
- further data relating to additional outcomes prioritised by the Committee were extracted and added to the data set from 8 studies that had already been included by Haas et.al
- we separated alcohol, and human chorionic gonadotropin (HCG) as different class interventions whereas these were placed in one class of medicines ('other' class) in Haas 2012
- Three studies were removed from the original dataset for the following reasons:
  - Grignaffini 2007 as it was an observational study
  - Lorzadeh 2007 because the included intervention (Human chorionic gonadotropin [HCG]) is not used in current practice and it was only the only study for that loop in the network.
  - Roy 1992 because it only reported results for one outcome of interest and for that outcome zero events were reported in one arm of the trial (see below).

Following the above changes, the final dataset from Haas 2012 NMA was based on 91 RCTs and included data for 35 different medicines across 9 out of 10 classes of interest as follows:

- placebo (placebo or usual or standard care without a tocolytic medicine)
- beta mimetics (ritodrine, terbutaline, nylidrin, salbutamol, fenoterol, hexoprenaline, isoxsuprine)
- calcium channel blockers (nifedipine, nicardipine)
- magnesium sulfate
- nitrates (nitroglycerin, nitric oxide)
- oxytocin receptor blockers (atosiban, barusiban)
- others (treatments defined as 'tocolysis' and 'other tocolytics' by study authors)
- prostaglandin inhibitors (indomethacin, celecoxib, sulindac, ketorolac, rofecoxib)
- alcohol/ethanol

The available data allowed for a NMA to be undertaken for 8 out of the 13 outcomes prioritised by the Committee as follows (separate NMAs were carried out for each outcome):

1. neonatal mortality
2. perinatal mortality

3. RDS
4. IVH
5. mothers with adverse events sufficient to require cessation of treatment
6. delay of birth by at least 48 hours
7. neonatal sepsis
8. gestational age at birth

Limited data were available for 4 out of the 5 remaining outcomes and were analysed using conventional pair-wise meta-analysis (see below).

12.1.3.1 Methods

A NMA class model was used to estimate the relative effects of each treatment class compared to placebo/control. Since there was no evidence of within-class variability (see Appendix J.) for any of the outcomes considered, all the results presented assume that all treatments in a class have the same relative effect.

A binomial / logit model was used to model outcomes 1 to 7 and a normal model with identity link was used to model EGA.

The final dataset consisted of data from 93 trials comparing 35 treatments, although not all trials report all the outcomes of interest. Studies reporting zero events on all arms were removed from the NMA as they do not contribute information on the relative treatment effects. A Bayesian framework is used to estimate all parameters, using Markov chain Monte Carlo simulation methods implemented in WinBUGS 1.4.3. (WinBUGS is software / code for performing the computation). Under this framework, and unlike in standard meta-analysis packages, it is not necessary to add a continuity correction (add 0.5 to arms of studies that report zero events in one arm). For detailed description of methods (baseline variability, relative effects model, NMA model for Binary and continuous data) and sample WinBUGS code see Appendix J.

12.1.3.2 Limitations in the data

Most comparisons were only made in one or 2 trials. Furthermore, not all trials report all outcomes. Some networks were very sparse in terms of patient numbers contributing to each loop.

Because some studies included multiple births, allowing more than one infant per mother, it was not always clear which was the most appropriate number of individuals to consider for outcomes on the infant. Where available we used the number of infants as the denominator. Although this does not account for the expected correlation in outcomes of infants from the same mother, it prevents double counting of infants from the same mother who may both have had an event.

12.1.4 Introduction to pair-wise meta-analysis

There was insufficient data to undertake NMA for the other outcomes set up in the review protocol. Limited data allowed for conventional meta-analysis to be undertaken for the following outcomes:

- neurodevelopmental disability (developmental delay, intellectual, gross motor, visual or hearing impairment, learning difficulties) - results for this outcome were derived from Houtzager 2006
- periventricular leucomalacia (PVL) - results for this outcome were derived from studies included in Haas 2012

- chronic lung disease (CLD) - results for this outcome were derived from studies included in Haas 2012 and have been presented separately
- maternal infection - results for this outcome were derived from studies included in Haas 2012 and have been presented separately
- three studies reported maternal mortality, however, all 3 studies reported zero events in either arm and so no further analysis was possible

12.1.5 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Firstly, the results of the updated NMA are presented by outcome (Sections 12.1.5.1 - 12.1.6.6).

For each outcome included in the NMA, the following information is presented:

- a text description of the studies included in the outcome-specific network
- a network diagram providing a graphic representation of the treatments compared in the outcome-specific network and the 'weight' of the data in terms of the number of included studies and number of participants
- a table summarising the relative effects derived from the direct and the overall NMA for each outcome
- a table summarising the probability rankings of effectiveness for the medicine classes included in the NMA, excluding the classes of 'Alcohol/ethanol' and 'Others'
- a set of rankograms that provide a graphic representation of the probability rankings of effectiveness for the medicine classes included in the NMA
- a modified GRADE profile summarising the quality assessment of the studies included in the outcomes-specific network

Evidence for the pairwise comparisons for the outcomes of neurodevelopmental disability, PVL and maternal infection, CLD is presented by comparison in GRADE profiles.

Full description of the characteristics and results of the included studies can be found in the Evidence Tables in Appendix H.

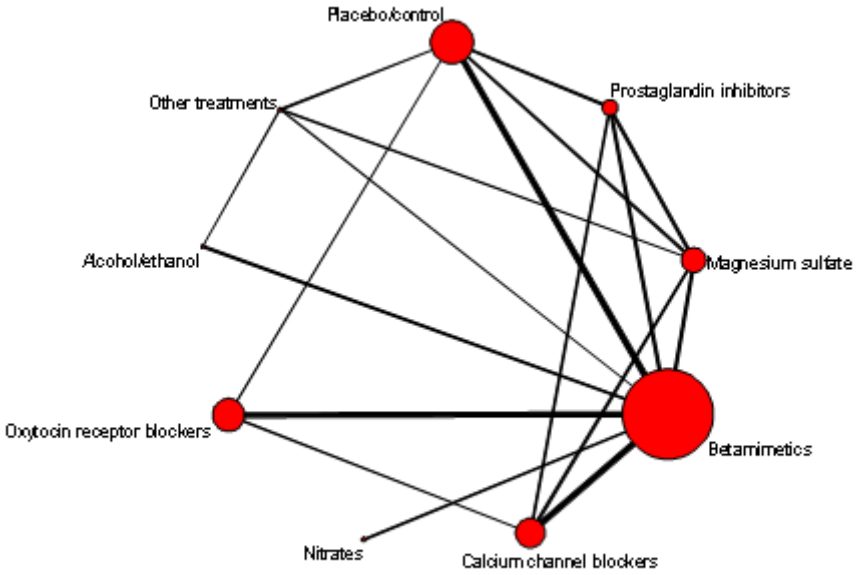
12.1.5.1 Neonatal mortality

Out of the 98 studies included in the review, 59 studies reported neonatal mortality as an outcome:

- 9 studies observed no events and were removed
- 1 study only reported events in one arm and was also removed.

The remaining 49 studies (Cotton, 1984; Klauser, 2012; Goodwin,1996; Niebyl, 1980; Panter, 1999; Zuckerman, 1984; Cox, 1990; Spellacy, 1979; Merkatz, 1980; Leveno, 1986; CPLIG, 1992; Romero, 2000; Weiner, 1988; Morales, 1993; Parilla, 1997; Morales, 1989; Kurki, 1991; McWhorter, 2004; Lyell, 2007; Essed, 1978; Holleboom,1996; Caritis, 1984; Maitra, 2007; Cararach, 2006; VandeWater, 2008; Papatsonis, 1997/ 2000; Shim, 2006; Moutquin, 2000; Lauersen, 1977; French/Australian, 2001; Laohapojanart, 2007; European, 2001; Nassar, 2009; Al-Omari, 2006; Koks, 1998; Kupferminc, 1993; Fan, 2003: Rayamajhi, 2003; Bisits, 2004; Mittendorf MAGnet, 2002; Glock, 1993; Surichamorn, 2001; Zhu, 1996; Kashanian, 2011; Besinger, 1991; Larson, 1980; Adam, 1966; Ma, 1992; Spearing, 1979) examined 19 medicines treatments allowing for 9 out of the 9 treatment classes to be assessed against each other.

Figure 3: Graphic representation of tocolytics trials for the NMA for neonatal mortality



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the lines is proportional to the number of studies contributing to the comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 49: Posterior median of the log odds ratios (OR) and 95% credible Intervals (CrI) for neonatal mortality

Upper diagonal: values shown are the log odds-ratios (OR) for the columns header versus the row header and are derived from the NMA. Given that this table relates to an adverse outcome, values lower than 1 favour the column defining treatment and values higher than 1 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

Lower diagonal: values shown are the log OR for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to an adverse outcome, values higher than 1 favour the column defining treatment and values lower than 1 favour the row defining the treatment.

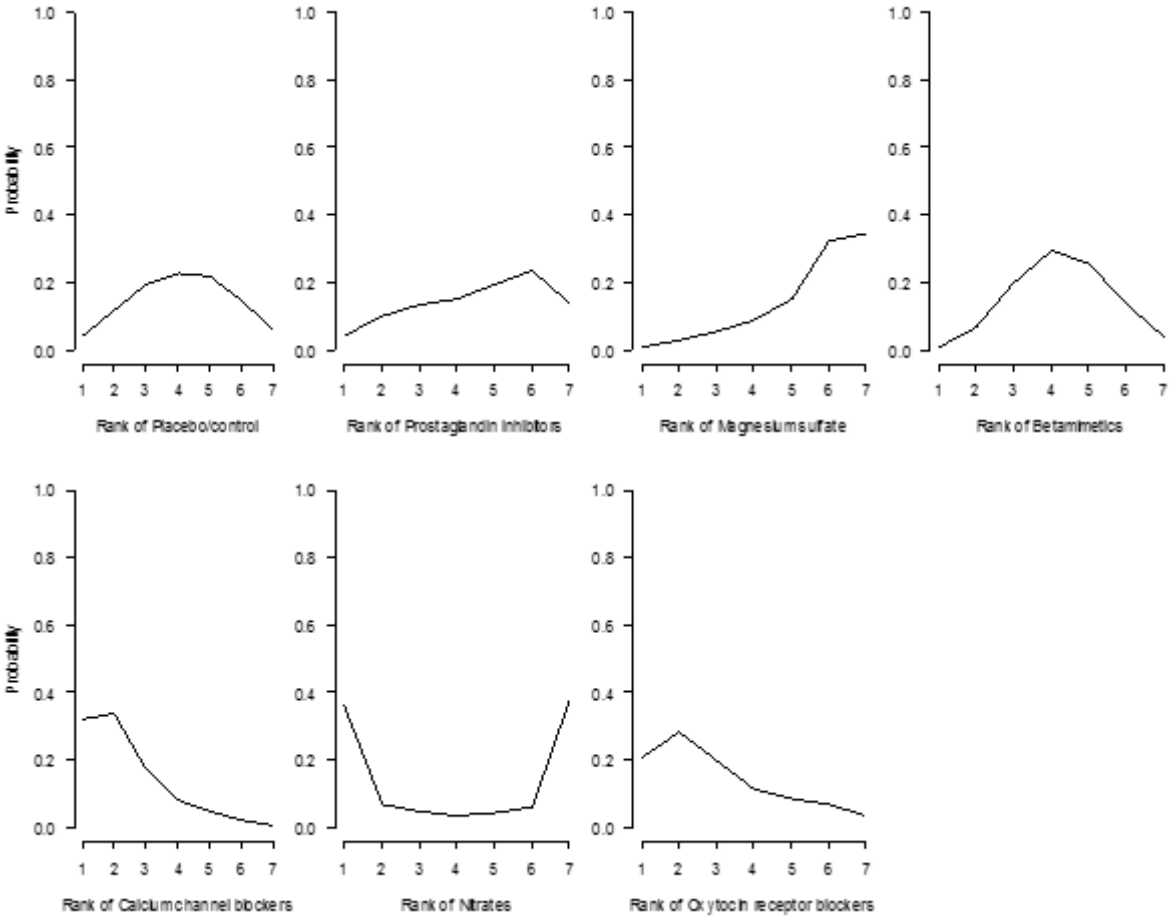
	Placebo/control	Prostaglandin inhibitors	Magnesium sulfate	Betamimetics	Calcium channel blockers	Nitrates	Oxytocin receptor blockers	Alcohol/ethanol	Other treatments
Placebo/control		1.13 (0.39 to 3.40)	1.49 (0.56 to 4.09)	1.02 (0.49 to 2.15)	0.62 (0.21 to 1.80)	0.98 (0.02 to 62.47)	0.73 (0.23 to 2.19)	2.33 (0.41 to 13.99)	0.56 (0.11 to 2.60)
Prostaglandin inhibitors	1.08 (0.15 to 7.80)		1.32 (0.45 to 3.81)	0.90 (0.32 to 2.41)	0.55 (0.16 to 1.70)	0.86 (0.01 to 55.83)	0.64 (0.16 to 2.37)	2.06 (0.29 to 13.93)	0.49 (0.08 to 2.74)
Magnesium sulfate	1.18 (0.23 to 5.58)	1.42 (0.35 to 7.13)		0.68 (0.26 to 1.75)	0.42 (0.13 to 1.23	0.65 (0.01 to 42.22)	0.49 (0.13 to 1.72)	1.56 (0.24 to 10.17)	0.37 (0.06 to 1.83)
Betamimetics	0.79 (0.31 to 1.97)	1.15 (0.21 to 6.02)	0.91 (0.14 to 6.53)		0.61 (0.25 to 1.43)	0.96 (0.02 to 56.79)	0.71 (0.26 to 1.83)	2.28 (0.44 to 12.21)	0.55 (0.10 to 2.56)
Calcium channel blockers	-	0.25 (0.03 to 1.62)	2.33 (1.27 to 4.87)	0.59 (1.18 to 1.74)		1.57 (0.02 to 106.20)	1.17 (0.36 to 3.84)	3.74 (0.60 to 25.10)	0.89 (0.15 to 5.07)
Nitrates	0 cell	-	-	0.95 (0.02 to 58.44)	-		0.74 (0.01 to 49.77)	2.39 (0.03 to 192.10)	0.57 (0.01 to 43.68)
Oxytocin receptor blockers	4.98 (0.45 to 74.44)	-	-	0.43 (0.13 to 1.29)	1.15 (0.11 to 12.54)	-		3.21 (0.49 to 22.37)	0.76 (0.12 to 4.70)
Alcohol/ethanol	-	-	-	3.74 (0.63 to 26.36)	-	-	-		0.24 (0.03 to 1.65)
Other treatments	0.69 (0.06 to 8.18)	-	0 cell	2.97 (0.18 to 58.38)	-	-	-	-	

Table 50: Probability rankings for medicines for neonatal mortality. Ranking probability of best medicine classes for improving the outcome. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Nitrates	36%	4.0	4	(1 to7)
Calcium channel blockers	32%	2.3	2	(1 to6)
Oxytocin receptor blockers	21%	2.9	3	(1 to 7)
Placebo/control	4%	4.1	4	(1 to7)
Prostaglandin inhibitors	4%	4.6	5	(1 to7)
Betamimetics	1%	4.3	4	(2 to7)
Magnesium sulfate	1%	5.7	6	(2 to7)

CrI credible interval

Figure 4: Graphic representation of the each medicine's effectiveness probability ranking for reducing neonatal death



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (7). The line indicates the probability that medicine will achieve each rank, for example placebo/control has a 0.01 probability of being ranked most effective and 0.2 probability of being ranked fifth most effective.

12.1.5.1.1 NMA Quality assessment

There was some evidence of inconsistency (conflict between direct and indirect evidence) in the comparison of placebo/control versus oxytocin receptor blockers (Direct OR = 4.95 (95%CrI 0.83 to 40.45) versus indirect OR = 0.44 (95%CrI 0.17, 1.14), p-value = 0.022)

In relation to heterogeneity, 23/59 studies in the network included women with multiple pregnancy, therefore the results should be interpreted with the perspective of mixed populations.

Table 51: Quality assessment of the evidence contributing to the NMA for neonatal death

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
RDS							
1 NMA of 49 studies (Original data from Haas 2012)	49 RCTs[a]	serious[1]	serious[2]	serious[3]	serious[4]	None	Very low

[a] Cotton, 1984; Klauser, 2012; Goodwin,1996; Niebyl, 1980; Panter, 1999; Zuckerman, 1984; Cox, 1990; Spellacy, 1979; Merkatz, 1980; Leveno, 1986; CPLIG, 1992; Romero, 2000; Weiner, 1988; Morales, 1993; Parilla, 1997; Morales, 1989; Kurki, 1991; McWhorter, 2004; Lyell, 2007; Essed, 1978; Holleboom,1996; Caritis, 1984; Maitra, 2007; Cararach, 2006; VandeWater, 2008; Papatsonis, 1997/ 2000; Shim, 2006; Moutquin, 2000; Lauersen, 1977; French/Australian, 2001; Laohapojanart, 2007; European, 2001; Nassar, 2009; Al-Omari, 2006; Koks, 1998; Kupferminc, 1993; Fan, 2003: Rayamajhi, 2003; Bisits, 2004; Mittendorf MAGnet, 2002; Glock, 1993; Surichamorn, 2001; Zhu, 1996; Kashanian, 2011; Besinger, 1991; Larson, 1980; Adam, 1966; Ma, 1992; Spearing, 1979.
[1] Analysis was based on the class therefore different doses and co-treatment were combined together
[2] There were some evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of Placebo/control v Oxytocin receptor blockers:
Placebo/control v Oxytocin receptor blockers
Bayesian p-value = 0.022
Direct OR = 4.95 (95%CrI 0.83 to 40.45)
Indirect OR = 0.44 (95%CrI 0.17, 1.14)
[3] Women with multiple pregnancy were included in 23/59 studies
[4] Wide and very wide CrI across all comparisons

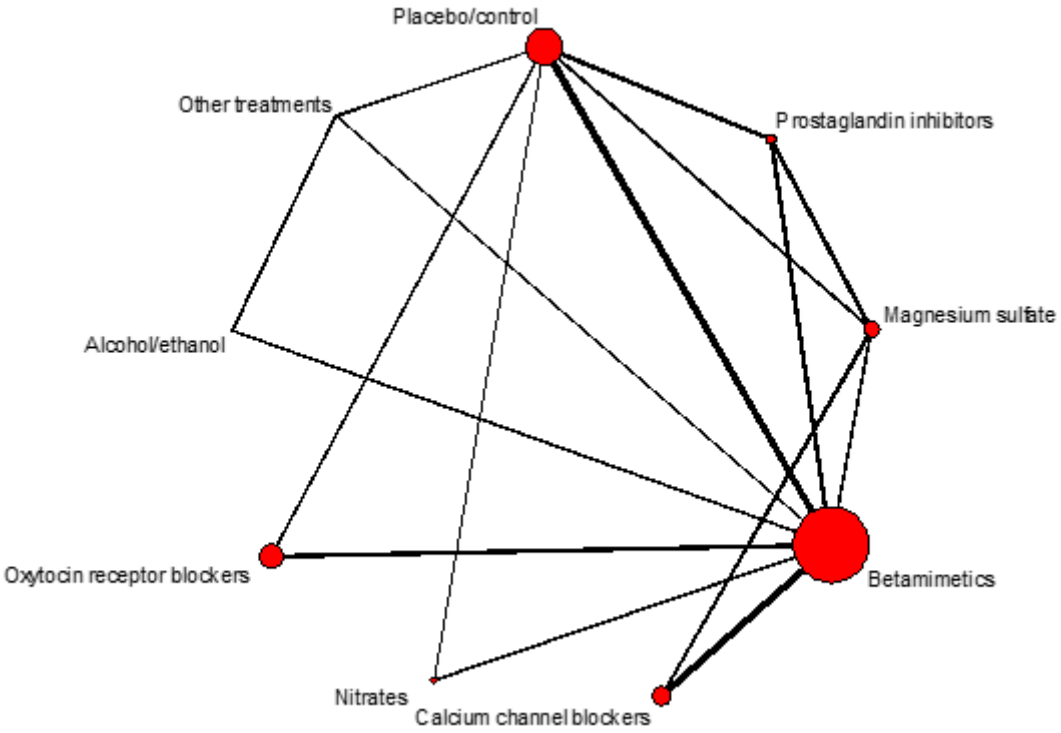
12.1.6 Perinatal Mortality

Out of the 98 studies included in the review, 47 studies reported perinatal mortality as an outcome:

- 3 studies observed no events and were removed

The remaining 44 studies examined 20 medicines treatments allowing for 9 out of the 9 treatment classes to be assessed against each other.

Figure 5: Graphic representation of tocolytics trials for the NMA for perinatal mortality.



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the lines is proportional to the number of studies contributing to the comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 52: Posterior median of the log odds ratios (OR) and 95% credible Intervals (CrI) for perinatal death

Upper diagonal: values shown are the log OR for the columns header versus the row header and are derived from the NMA. Given that this table relates to an adverse outcome, values lower than 1 favour the column defining treatment and values higher than 1 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

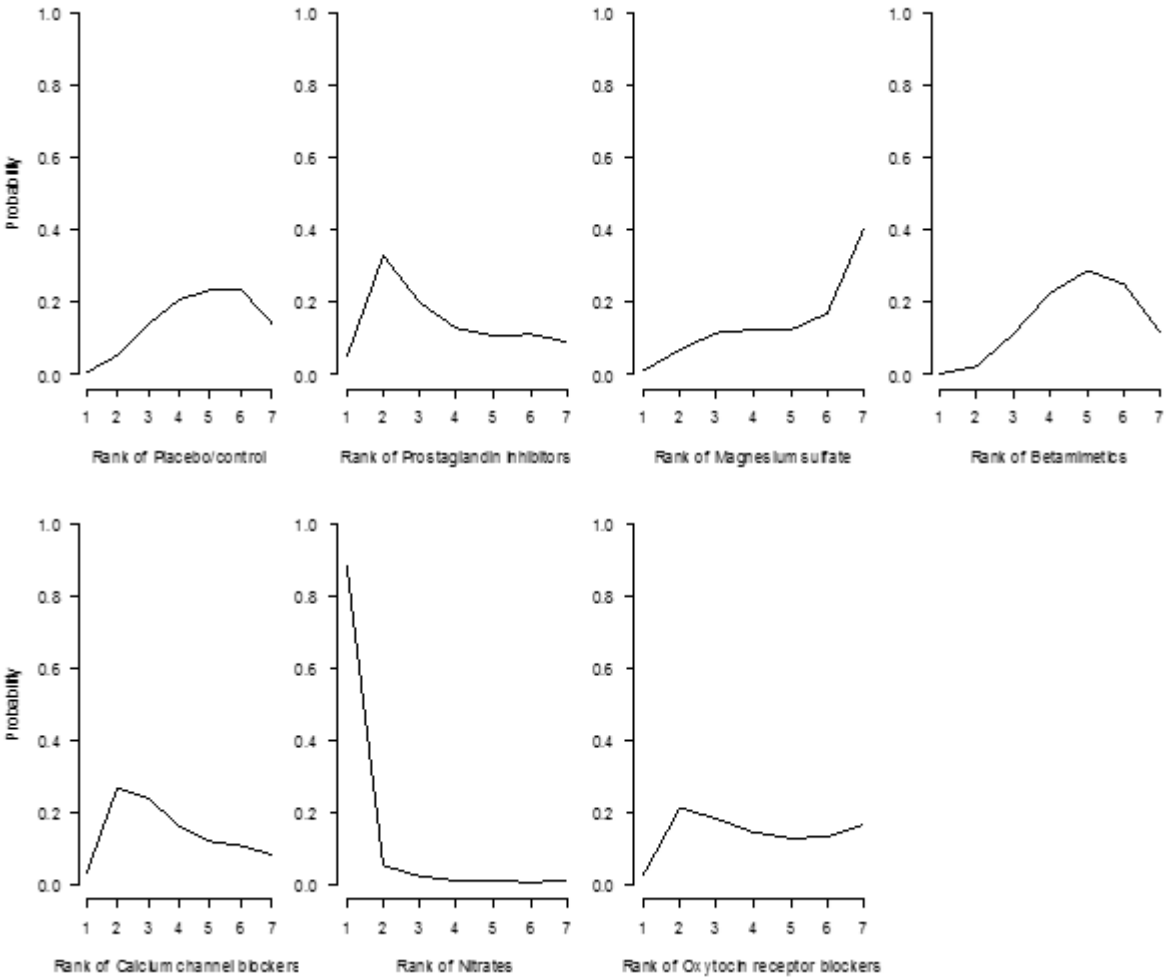
Lower diagonal: values shown are the log OR for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to an adverse outcome, values higher than 1 favour the column defining treatment and values lower than 1 favour the row defining the treatment.

	Placebo/control	Prostaglandin inhibitors	Magnesium sulfate	Betamimetics	Calcium channel blockers	Nitrates	Oxytocin receptor blockers	Alcohol/ethanol	Other treatments
Placebo/control		0.72 (0.22 to 2.28)	1.19 (0.35 to 3.73)	1.01 (0.48 to 1.99)	0.76 (0.25 to 2.24)	0.10 (0.00 to 1.07)	0.86 (0.25 to 2.59)	2.59 (0.50 to 13.84)	2.00 (0.41 to 9.74)
Prostaglandin inhibitors	0.78 (0.13 to 4.96)		1.65 (0.44 to 6.34)	1.40 (0.43 to 4.49)	1.07 (0.25 to 4.34)	0.14 (0.00 to 1.86)	1.19 (0.25 to 5.39)	3.59 (0.54 to 25.48)	2.80 (0.41 to 18.54)
Magnesium sulfate	2.05 (0.36 to 10.82)	2.97 (0.46 to 28.36)		0.85 (0.28 to 2.736)	0.64 (0.18 to 2.48	0.08 (0.00 to 1.14)	0.72 (0.16 to 3.29)	2.20 (0.34 to 15.72)	1.67 (0.26 to 11.72)
Betamimetics	0.89 (0.35 to 1.87)	0.93 (0.14 to 5.63)	2.62 (0.13 to 120.7)		0.75 (0.31 to 1.83)	0.10 (0.00 to 1.05)	0.85 (0.28 to 2.42)	2.56 (0.57 to 12.94)	1.98 (0.42 to 9.75)
Calcium channel blockers	-	-	3.99 (0.32 to 143.2)	0.61 (0.23 to 1.51)		0.13 (0.00 to 1.65)	1.13 (0.27 to 4.35)	3.41 (0.60 to 21.24)	2.63 (0.43 to 16.10)
Nitrates	0 cell	-	-	0.24 (0.01 to 3.80)	-		8.42 (0.65 to 308.20)	26.41 (1.57 to 1163.00)	20.1 (1.23 to 874.00)
Oxytocin receptor blockers	2.44 (0.36 to 16.93)	-	-	0.51 (0.14 to 1.78)	-	-		3.03 (0.50 to 21.71)	2.32 (0.37 to 16.03)
Alcohol/ethanol	-	-	-	3.70 (0.76 to 21.61)	-	-	-		0.77 (0.13 to 4.53)
Other treatments	0.70 (0.07 to 6.37)	-	-	5.12 (0.53 to 53.36)	-	-	-	-	

Table 53: Probability rankings for medicines for perinatal mortality. Ranking probability of best medicine classes for improving the outcome. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Nitrates	89%	1.3	1	(1 to 5)
Prostaglandin inhibitors	5%	3.6	3	(1 to 7)
Calcium channel blockers	3%	3.7	3	(1 to 7)
Oxytocin receptor blockers	3%	4.2	4	(1 to 7)
Magnesium sulfate	1%	5	6	(2 to 7)
Placebo/control	0%	4.9	5	(2 to 7)
Betamimetics	0%	5	5	(2 to 7)

Figure 6: Graphic representation of the each medicine's effectiveness probability ranking for reducing perinatal mortality



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (7). The line indicates the probability that medicine will achieve each rank.

Table 54: Quality assessment of the evidence contributing to the network analysis for perinatal death

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
RDS							
1 NMA of 49 studies (Original data from Haas 2012)	44 RCTs[a]	serious[1]	no serious inconsistency	serious[2]	serious[3]	None	Very low

[a] *Cotton 1984; Niebyl 1980; Panter 1999; Zuckerman 1984; Cox 1990; Spellacy 1979; Leveno 1986; CPLIG 1992; Romero 2000; Weiner 1988; Morales 1993; Parilla 1997; Morales 1989; Kurki 1991; McWhorter 2004; Lyell, 2007; Essed 1978; Holleboom1996; Caritis 1984; Cararach 2006; VandeWater 2008; Papatsonis 1997/ 2000; Shim 2006; Moutquin 2000; Lauersen 1977; French/Australian 2001; European 2001; Koks 1998; Fan 2003: Rayamajhi 2003; Bisits 2004; Glock 1993; Besinger 1991; Larson 1980; Larson 1986; Smith 2007 ; Floyd 1995; Gummerus 1983; Sirohiwal 2001; Trabelsi 2008; Adam 1966; Spearing 1979, Jaju 2011.*

[1] *Analysis was based on the class therefore different doses and co-treatment were combined together*

[2] *Women with multiple pregnancy were included in 37/44 studies*

[3] *Wide and very wide CrI across all comparisons*

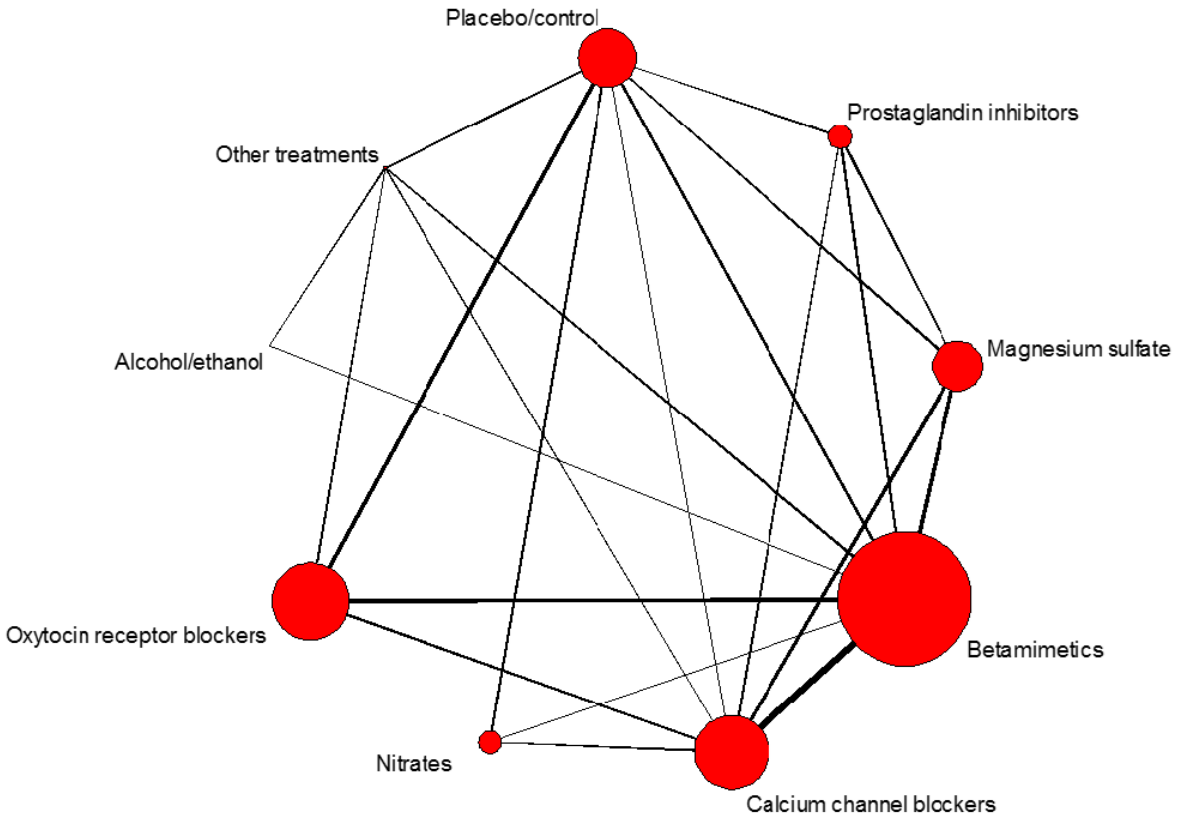
12.1.6.1 Delay of birth by more than 48 hours

Out of the 98 studies included in the review, 69 studies reported delay of birth by more than 48 hours as an outcome:

- 2 studies observed same events rate in both arms and were removed.

The reminding 67 studies examined 26 medicine treatments allowing for 9 out of the 9 treatment classes to be assessed against each other.

Figure 7: Graphic representation of tocolytics trials for the NMA for delay by 48 hours



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the lines is proportional to the number of studies contributing to the comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 55: Posterior median of the odds ratios (OR) and 95% credible Intervals (CrI) for delay birth by more than 48 hours

Upper diagonal: values shown are the OR for the columns header versus the row header and are derived from the NMA. Given that this table relates to a positive outcome, values higher than 1 favour the column defining treatment and values lower than 1 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

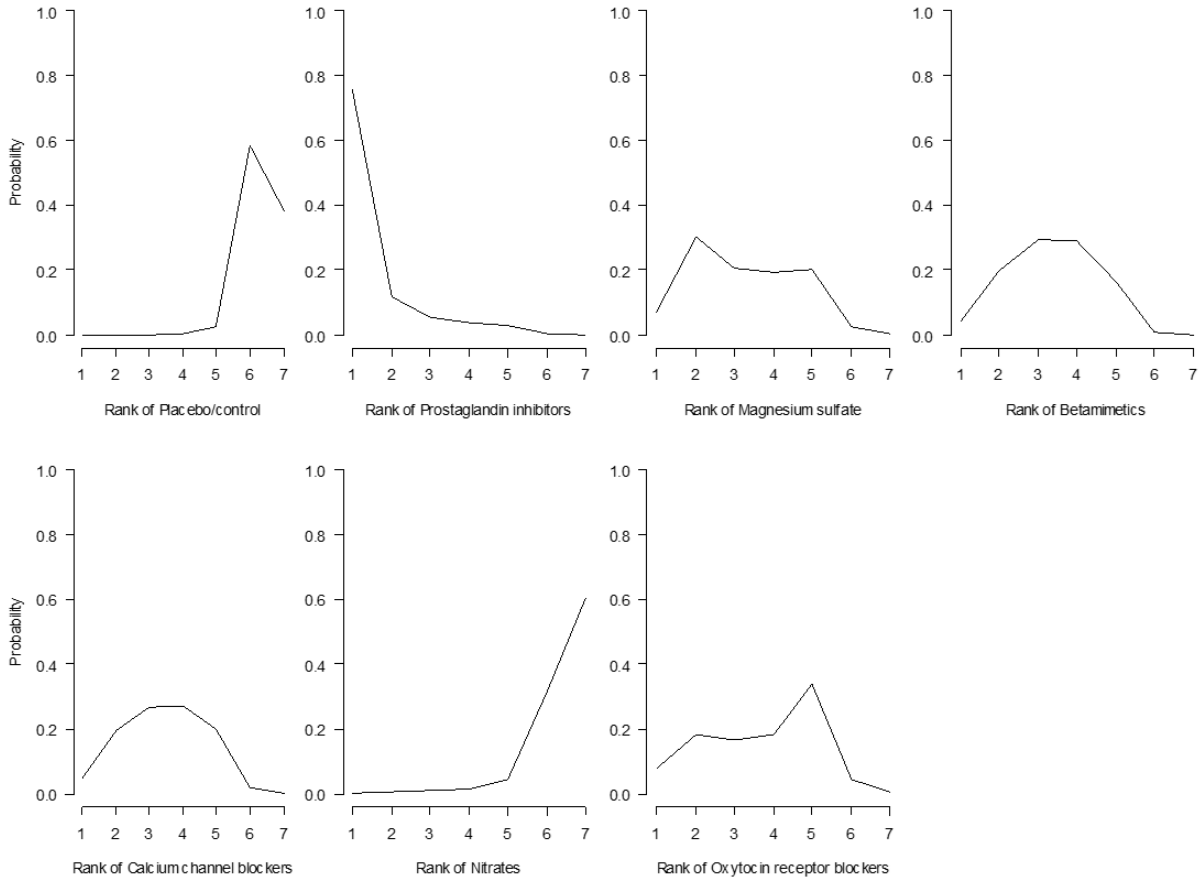
Lower diagonal: values shown are the OR for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to a positive outcome, values lower than 1 favour the column defining treatment and values higher than 1 favour the row defining the treatment.

	Placebo/control	Prostaglandin inhibitors	Magnesium sulfate	Betamimetics	Calcium channel blockers	Nitrates	Oxytocin receptor blockers	Alcohol/ethanol	Other treatments
Placebo/control		3.14 (1.45,7.05)	2.10 (1.10,4.07)	2.04 (1.17,3.59)	2.02 (1.11,3.76)	0.89 (0.40,2.02)	1.93 (1.02,3.65)	0.83 (0.12,5.64)	1.10 (0.38,3.24)
Prostaglandin inhibitors	14.51 (2.87,86.23)		0.67 (0.33,1.33)	0.65 (0.32,1.29)	0.64 (0.31,1.30)	0.28 (0.10,0.78)	0.61 (0.26,1.40)	0.26 (0.04,1.89)	0.35 (0.10,1.19)
Magnesium sulfate	2.65 (0.91,7.85)	0.88 (0.36,2.10)		0.97 (0.57,1.66)	0.96 (0.56,1.65)	0.43 (0.17,1.07)	0.92 (0.45,1.86)	0.40 (0.06,2.73)	0.53 (0.17,1.65)
Betamimetics	2.68 (1.22,6.15)	0.32 (0.09,1.12)	0.90 (0.41,1.98)		0.99 (0.65,1.50)	0.44 (0.19,1.01)	0.95 (0.54,1.63)	0.41 (0.06,2.63)	0.54 (0.18,1.57)
Calcium channel blockers	1.68 (0.30,9.26)	2.06 (0.63,6.82)	1.24 (0.56,2.73)	0.91 (0.56,1.49)		0.44 (0.19,1.03)	0.96 (0.52,1.74)	0.41 (0.06,2.73)	0.54 (0.18,1.63)
Nitrates	0.35 (0.13,0.93)	-	-	0.74 (0.15,3.62)	1.85 (0.46,7.27)		2.16 (0.85,5.47)	0.93 (0.12,7.02)	1.23 (0.34,4.52)
Oxytocin receptor blockers	1.51 (0.70,3.15)	-	-	1.02 (0.50,2.06)	1.06 (0.36,3.13)	-		0.43 (0.06,2.96)	0.57 (0.19,1.78)
Alcohol/ethanol	-	-	-	0.38 (0.05,2.74)	-	-	-		1.33 (0.21,8.59)
Other treatments	1.12 (0.33,3.90)	-	-	0.48 (0.06,3.56)	-	-	-	-	

Table 56: Probability rankings for medicines by class to increase delay of birth by > 48 hours. Probability rankings for medicines to increase delay of birth by > 48 hours. Ranking probability of best medicine classes for improving the outcome. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Prostaglandin inhibitors	76%	1.474	1	(1 to 5)
Oxytocin receptor blockers	8%	3.683	4	(1 to 6)
Magnesium sulfate	7%	3.232	3	(1 to 6)
Calcium channel blockers	5%	3.444	3	(1 to 5)
Betamimetics	4%	3.371	3	(1 to 5)
Nitrates	0%	6.46	7	(4 to 7)
Placebo/control	0%	6.337	6	(5 to 7)

Figure 8: Graphic representation of the each medicine's effectiveness probability ranking for improving outcome



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (7).

Table 57: Quality assessment of the evidence contributing to the outcome of delay of birth by > 48 hours

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
RDS							
1 NMA of 67 studies (Original data from Haas 2012)	67 RCTs[a]	serious[1]	serious[2]	serious[3]	serious[4]	none	Very low

[a] Thornton, 2009; Cotton, 1984; Goodwin,1996; Panter, 1999; Zuckerman, 1984; Cox, 1990; Larsen, 1986; CPLIG, 1992; Goodwin, 1994; Romero, 2000; Weiner, 1988; Morales, 1993; Morales, 1989; Kurki, 1991; Lyell, 2007; Maitra, 2007; Cararach, 2006; VandeWater, 2008; Shim, 2006; Lin, 2009; Moutquin, 2000; Trabelsi, 2008; French/Australian, 2001; Laohapojanart, 2007; European, 2001; Nassar, 2009; Al-Omari, 2006; Ma, 1992; Spearing, 1979; Husslein, 2007; Larson, 1980; Ingemarsson 1976; Koks, 1998; Smith , 2007; Borna, 2007; Kashanian, 2011; Besinger, 1991; McWhorter, 2004; Larmon, 1999; Papatsonis, 1997; Papatsonis , 2000; Al-Qattan, 2000; Kupferminc, 1993; Fan, 2003: Rayamajhi, 2003; Bisits, 2004; Holleboom,1996; Kashanian, 2011; Amorim, 2009; Weerakul, 2002; Mawaldi, 2008; Motazedian, 2010; Beall, 1985; Haghighi, 1999; Taherian, 2007; Chau, 1992; Surichamorn, 2011; Aramayo, 1990; Wilkins, 1988; Tchilinguirian, 1984; Garcia-Velasco, 1998; Jaju, 2011; Nankali, 2014; Kashanian, 2014; Salim, 2012; Nikbakhat, 2014.

[1] Analysis was based on the class therefore different doses and co-treatment were combined together

[2] There were some evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of Placebo/control v Prostaglandin inhibitors

Bayesian p-value = 0.049
Direct OR= 14.51 (95%CrI 2.87, 86.23) 2 small studies compare these classes directly
Indirect OR=3.14 (95%CrI 1.45, 7.05)

The direct and indirect are not contradictory to each other (both on the same side of one)

Strong evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of Placebo/Control v Nitrates

Bayesian p-value = 0.007
Direct OR= 0.35 (95%CrI 0.13, 0.93) 2 medium studies compare these classes directly
Indirect OR=0.89 (95%CrI 0.40, 2.02)

The direct and indirect are not contradictory to each other (both on the same side of one)

Some evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of Prostaglandin inhibitors v Calcium channel blockers

Bayesian p-value = 0.036
Direct OR= 2.06 (95%CrI 0.63, 6.82) 2 medium studies compare these classes directly
Indirect OR=0.64 (95%CrI 0.31, 1.30)

The direct and indirect are contradictory to each other (both on opposite sides of one). Results from this network should therefore be treated with caution.

Some evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of Calcium channel blockers v Nitrates

Bayesian p-value = 0.024
Direct OR= 1.85 (95%CrI 0.46, 7.27) 2 small studies compare these classes directly
Indirect OR=0.44 (95%CrI 0.19, 1.03)
The direct and indirect are contradictory to each other (both on opposite sides of one). Results from this network should therefore be treated with caution.[3] Women with multiple pregnancy were included in 17/63 studies

[4] Wide and very wide CrIs across all comparisons except one direct comparison (Placebo/control vs Prostaglandin inhibitors)

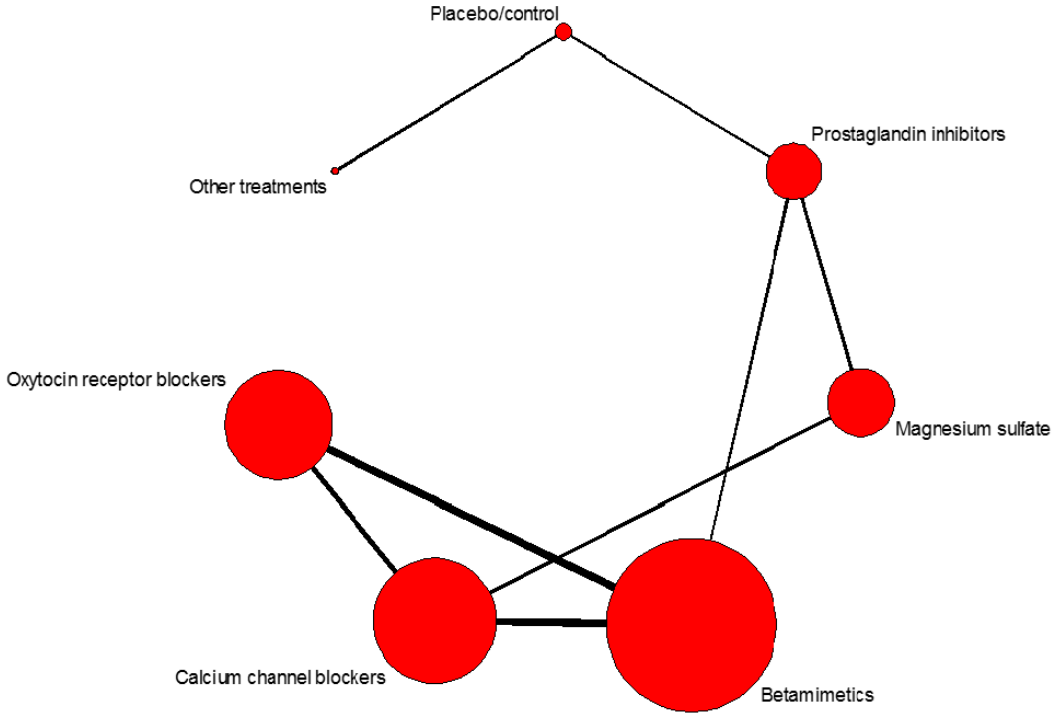
12.1.6.2 Neonatal sepsis

Out of the 98 studies included in the review, 22 studies reported neonatal sepsis as an outcome:

- 3 studies observed no events and were removed

The remaining 19 studies examined 12 medicines allowing for 7 out of the 9 treatment classes to be assessed against each other.

Figure 9: Graphic representation of tocolytics trials for the NMA for neonatal sepsis.



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the lines is proportional to the number of studies contributing to the comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 58: Posterior median of the odds ratios (OR) and 95% credible Intervals (CrI) for neonatal sepsis

Upper diagonal: values shown are the OR for the columns header versus the row header and are derived from the NMA. Given that this table relates to an adverse outcome, values lower than 1 favour the column defining treatment and values higher than 1 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

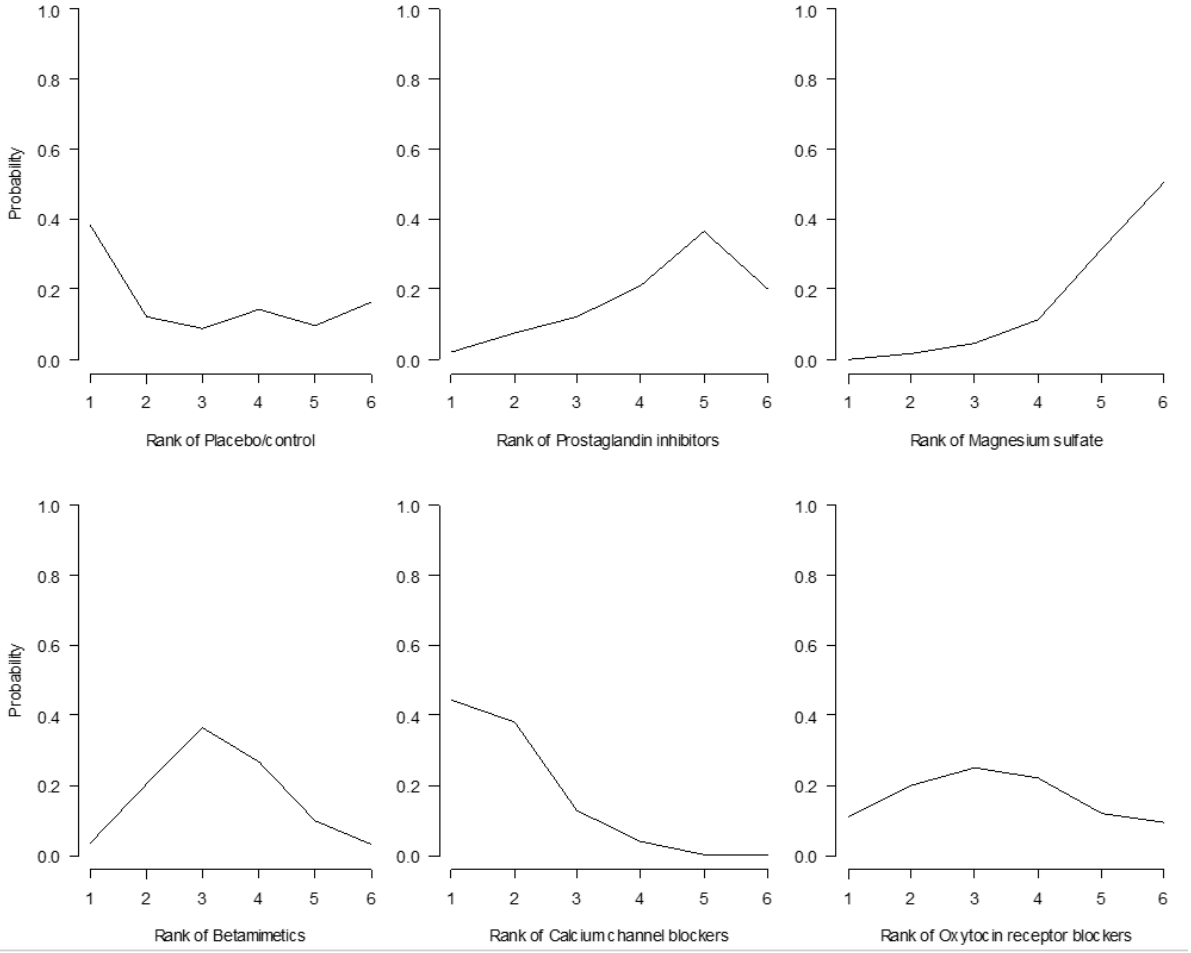
Lower diagonal: values shown are the OR for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to an adverse outcome, values higher than 1 favour the column defining treatment and values lower than 1 favour the row defining the treatment.

	Placebo/control	Prostaglandin inhibitors	Magnesium sulfate	Betamimetics	Calcium channel blockers	Oxytocin receptor blockers	Other treatments
Placebo/control		1.59 (0.33,9.33)	1.93 (0.43,10.77)	1.15 (0.25,6.56)	0.83 (0.18,4.75)	1.16 (0.22,7.15)	1.31 (0.21,8.05)
Prostaglandin inhibitors	0.00 (0.00,0.52)		1.21 (0.63,2.37)	0.72 (0.29,1.77)	0.52 (0.23,1.14)	0.73 (0.25,2.10)	0.81 (0.07,8.96)
Magnesium sulfate	7.91 (1.42,69.55)	0.96 (0.47,1.94)		0.59 (0.26,1.35)	0.43 (0.21,0.86)	0.60 (0.22,1.62)	0.67 (0.06,7.04)
Betamimetics	0.00 (0.00,0.18)	1.00 (0.02,39.17)			0.72 (0.42,1.23)	1.01 (0.55,1.87)	1.13 (0.09,12.27)
Calcium channel blockers		0.68 (0.28,1.57)	0.54 (0.10,2.36)	0.54 (0.29,0.98)		1.40 (0.65,3.03)	1.56 (0.13,16.94)
Oxytocin receptor blockers				1.06 (0.56,2.05)	0.96 (0.16,5.80)		1.12 (0.09,12.97)
Other treatments	1.30 (0.21,8.09)						

Table 59: Probability rankings for medicines by class to reduce neonatal sepsis. Probability rankings for medicines for neonatal sepsis. Ranking probability of best medicine classes for improving the outcome. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Calcium channel blockers	45%	1.779	2	(1 to 4)
Placebo/control	38%	2.942	2	(1 to 6)
Oxytocin receptor blockers	11%	3.319	3	(1 to 6)
Betamimetics	4%	3.287	3	(1 to 6)
Prostaglandin inhibitors	2%	4.428	5	(2 to 6)
Magnesium sulfate	0%	5.245	6	(3 to 6)

Figure 10: Graphic representation of the each medicine's effectiveness probability ranking for improving outcome



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (6). The line indicates the probability that medicine will achieve each rank.

Table 60: Quality assessment of the evidence contributing to the outcome of neonatal sepsis

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Neonatal sepsis							
1 NMA of 19 studies (Original data from Haas 2012)	19 RCTs[a]	serious[1]	no serious inconsistency	serious[2]	serious[3]	none	Very low

[a] *Cotton, 1984; Klauser, 2012; Goodwin,1996; Niebyl, 1980; Weiner, 1988; Stika, 2002; Al-Omari, 2006; Kurki, 1991; McWhorter, 2004; Lyell, 2007; Holleboom, 1996; Maitra, 2007; VandeWater, 2008; Papatsonis, 1997/2000; Moutquin, 2000; French/Australian, 2001; European, 2001; Nassar, 2009; Salim, 2012.*

[1] *Analysis was based on the class therefore different doses and co-treatment were combined together*

[2] *Women with multiple pregnancy were included in 7/19 studies 4 Wide and very wide CrI across all comparisons except on (Placebo/control vs magnesium sulfate)*

[3] *There were strong evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of Magnesium sulfate v Calcium channel blockers*

Bayesian p-value = 0.005
Direct OR=0.54 (95%CrI 0.10, 2.36) 2 medium studies compare these classes directly
Indirect OR= 0.43 (95%CrI 0.21, 0.86)

The direct and indirect are not contradictory to each other (both on the same side of one)

Some evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of betamimetics v calcium channel blockers

Bayesian p-value= 0.025
Direct OR = 0.54 (95%CrI 0.29, 0.98) 3 small studies compare these classes directly
Indirect OR= 0.72 (95%CrI 0.42, 1.23)

The direct and indirect are not contradictory to each other (both on the same side of one)

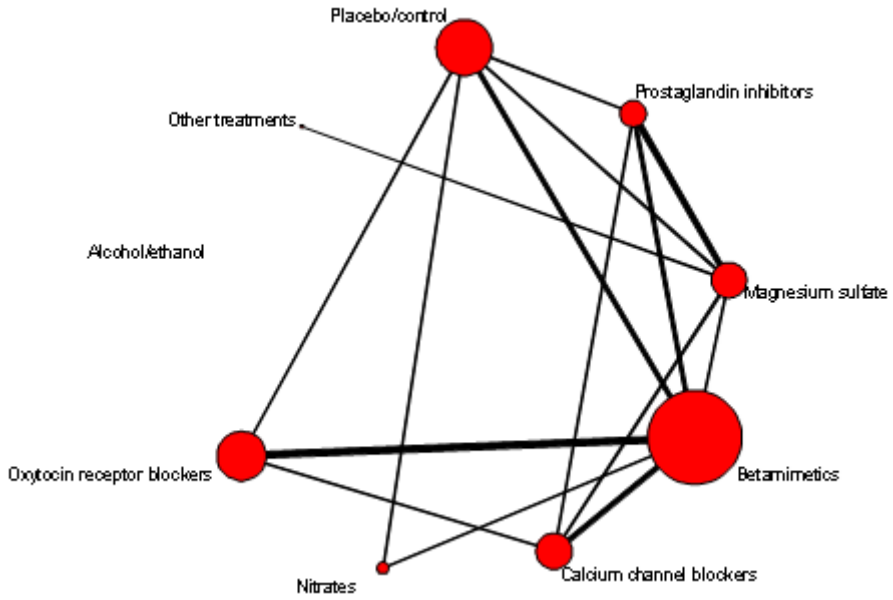
12.1.6.3 Intraventricular haemorrhage

Out of the 98 studies included in the review, 34 studies reported IVH as an outcome:

- 4 studies observed no events and were removed

The remaining 30 studies examined 14 medicines allowing for 8 out of the 9 treatment classes to be assessed against each other.

Figure 11: Graphic representation of tocolytics trials for the NMA for IVH



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the lines is proportional to the number of studies contributing to the comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 61: Posterior median of the odds ratios (OR) and 95% credible Intervals (CrI) for intraventricular haemorrhage

Upper diagonal: values shown are the OR for the columns header versus the row header and are derived from the NMA. Given that this table relates to an adverse outcome, values lower than 1 favour the column defining treatment and values higher than 1 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

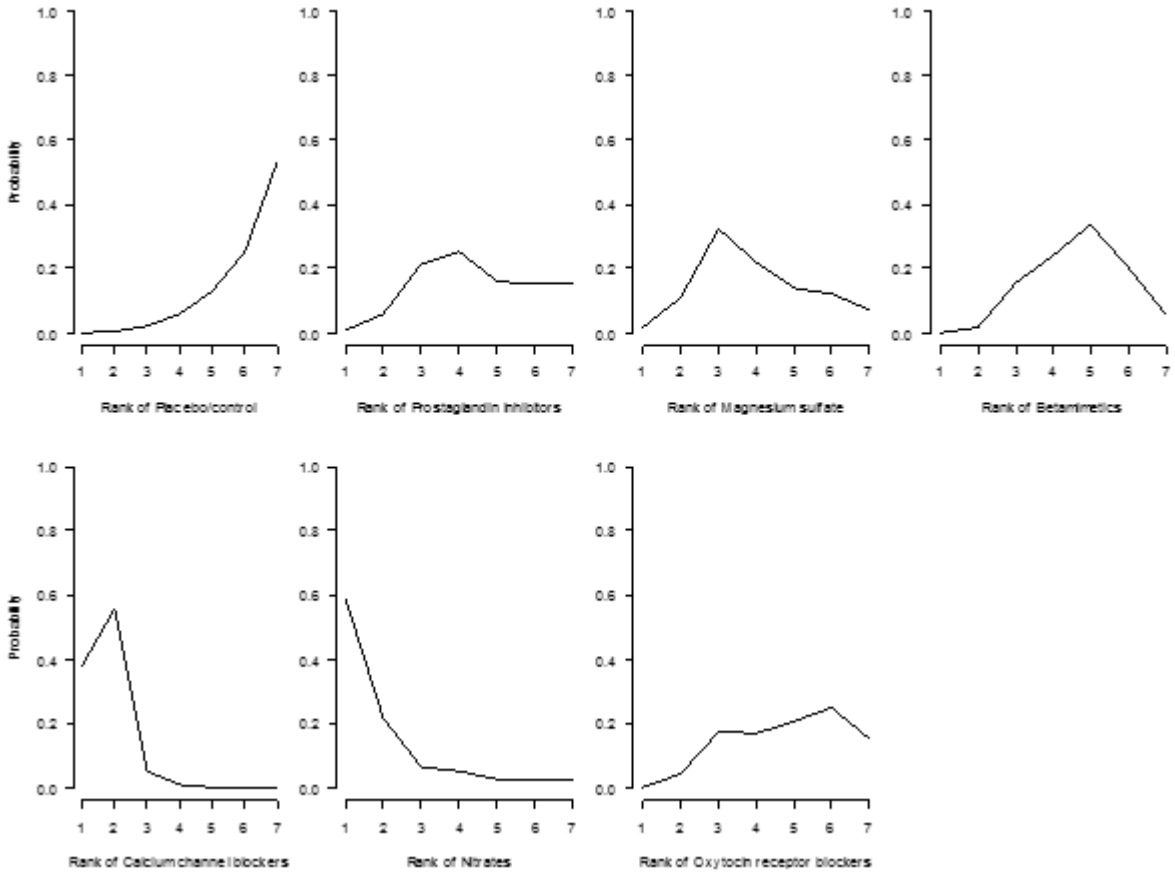
Lower diagonal: values shown are the OR for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to an adverse outcome, values higher than 1 favour the column defining treatment and values lower than 1 favour the row defining the treatment.

	Placebo/control	Prostaglandin inhibitors	Magnesium sulfate	Betamimetics	Calcium channel blockers	Nitrates	Oxytocin receptor blockers	Other treatments
Placebo/control		0.76 (0.35 to 1.59)	0.69 (0.33 to 1.43)	0.79 (0.51 to 1.22)	0.40 (0.21 to 0.74)	0.34 (0.08 to 1.13)	0.82 (0.48 to 1.37)	0.14 (0.02 to 0.77)
Prostaglandin inhibitors			0.91 (0.54 to 1.54)	1.05 (0.33 to 2.06)	0.53 (0.27 to 1.01)	0.45 (0.10 to 1.72)	1.08 (0.48 to 2.44)	0.19 (0.02 to 0.94)
Magnesium sulfate	0.76 (0.22 to 2.40)	0.91 (0.52 to 1.60)		1.15 (0.58 to 2.29)	0.58 (0.30 to 1.11)	0.49 (0.11 to 1.89)	1.19 (0.53 to 2.66)	0.21 (0.03 to 0.95)
Betamimetics	0.66 (0.39 to 1.11)	1.08 (0.32 to 3.66)	-		0.50 (0.30 to 0.83)	0.43 (0.11 to 1.39)	1.03 (0.63 to 1.71)	0.18 (0.02 to 0.96)
Calcium channel blockers	-	0.60 (0.25 to 1.34)	0.59 (0.07 to 3.95)	0.44 (0.24 to 0.79)		0.85 (0.20 to 3.05)	2.06 (1.04 to 4.08)	0.36 (0.04 to 1.88)
Nitrates	2.55 (0.19 to 81.45)	-	-	0.20 (0.03 to 0.86)	-		2.43 (0.68 to 10.28)	0.42 (0.04 to 3.66)
Oxytocin receptor blockers	0.84 (0.41 to 1.70)	-	-	1.06 (0.55 to 2.08)	-	-		0.17 (0.02 to 0.98)
Other treatments	-	-	0.21 (0.03 to 0.95)	-	-	-	-	

Table 62: Probability rankings for medicines by class to reduce IVH. Probability rankings for medicines for reducing intraventricular haemorrhage. Ranking probability of best medicine classes for improving the outcome. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Nitrates	59%	1.9	1	(1 to 7)
Calcium channel blockers	38%	1.7	2	(1 to 3)
Magnesium sulfate	2%	4.0	4	(2 to 7)
Prostaglandin inhibitor	1%	4.6	4	(2 to 7)
Betamimetics	0%	4.7	5	(3 to 7)
Oxytocin receptor blockers	0%	4.9	5	(2 to 7)
Placebo/control	0%	6.2	7	(4 to 7)

Figure 12: Graphic representation of the each medicine's effectiveness probability ranking for improving outcome



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (7). The line indicates the probability that medicine will achieve each rank.

Table 63: Quality assessment of the evidence contributing to the outcome of IVH

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
IVH							
1 NMA of 29 studies (Original data from Haas 2012)	30 RCTsa	serious1	no serious inconsistency	serious2	serious3	none	Very low

a Cotton 1984; Klauser 2012; Goodwin 1996; Panter 1999; Cox 1990; Leveno 1986; CPLIG 1992; Smith 2007; Romero 2000; Morales 1993; Parilla 1997; Morales 1989; Besinger 1991; Kurki 1991; Schorr 1998; McWhorter, 2004; Lyell 2007; Mittendorf MAGnet 2002; Bisits 2004; Holleboom 1996; Maitra 2007; VandeWater 2008; Papatsonis 1997; Papatsonis 2000; Shim 2006; Moutquin 2000; French/Australian 2001; Laohapojanart 2007; European 2001; Nassar 2009; Salim 2012.
1 Analysis was based on the class therefore different doses and co-treatment were combined together
2 Women with multiple pregnancy were included in 24/30 studies
3 Wide and very wide CrI across all comparisons

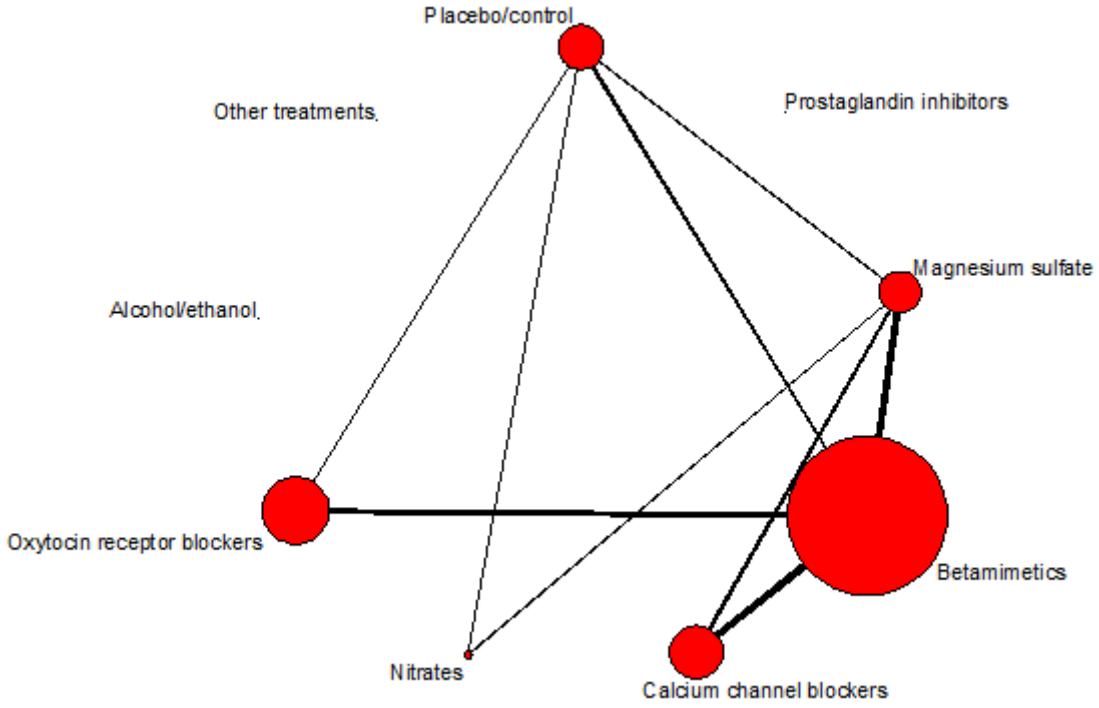
12.1.6.4 Discontinuation of treatment due to maternal adverse events

Out of the 98 studies included in the review, 62 studies reported neonatal sepsis as an outcome:

- 20 studies observed no events and were removed
- 6 studies only reported events on one arm and were also removed.

The remaining 36 studies examined 6 out of the 9 treatment classes to be assessed against each other.

Figure 13: Graphic representation of tocolytics trials for the NMA for treatment discontinued due to maternal side effect.



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the lines is proportional to the number of studies contributing to the

comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 64: Posterior median of the odds ratios (OR) and 95% credible Intervals (CrI) for discontinuation of treatment due to maternal adverse events

Upper diagonal: values shown are the OR for the columns header versus the row header and are derived from the NMA. Given that this table relates to an adverse outcome, values lower than 1 favour the column defining treatment and values higher than 1 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

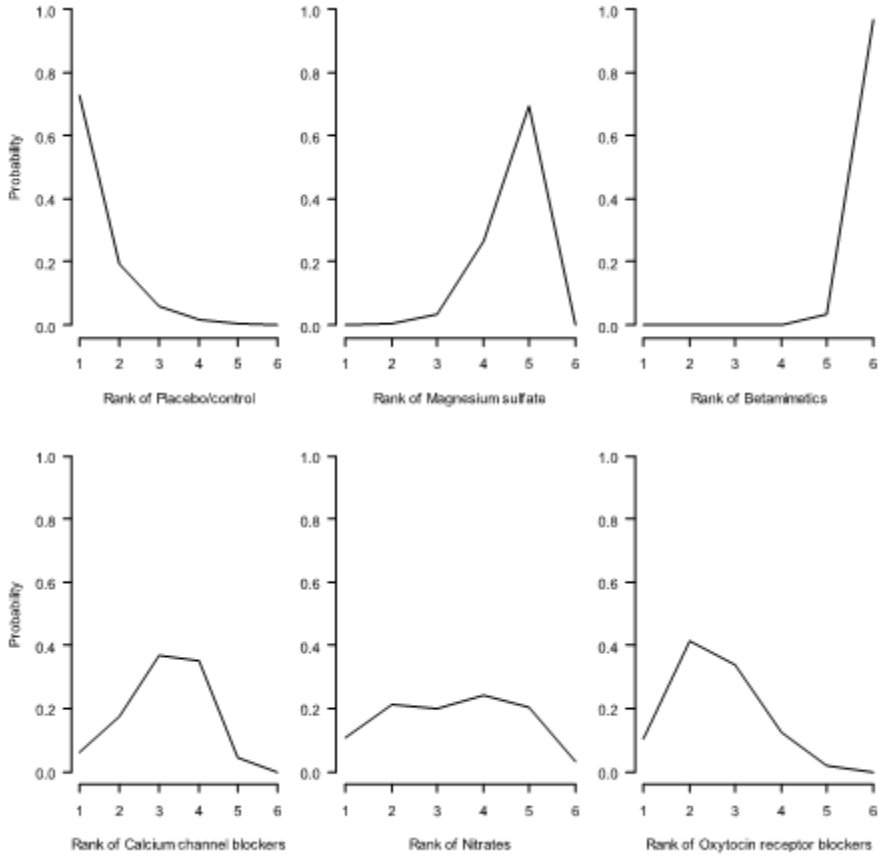
Lower diagonal: values shown are the OR for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to an adverse outcome, values higher than 1 favour the column defining treatment and values lower than 1 favour the row defining the treatment.

	Placebo/control	Magnesium sulfate	Betamimetics	Calcium channel blockers	Nitrates	Oxytocin receptor blockers
Placebo/control		16.21 (1.89 to 175.70)	132.20 (18.52 to 1284.00)	5.22 (0.35 to 56.55)	5.58 (0.26 to 165.00)	3.15 (0.31 to 23.18)
Magnesium sulfate	Zero cell		8.05 (2.23 to 34.05)	0.32 (0.04 to 1.39)	0.34 (0.01 to 9.19)	0.19 (0.02 to 1.15)
Betamimetics	109.84 (2.67 to 23623.56)	8.82 (1.01 to 90.65)		0.04 (0.01 to 0.14)	0.02 (0.00 to 1.20)	0.02 (0.00 to 0.09)
Calcium channel blockers	-	0.37 (0.01 to 12.76)	0.02 (0.00 to 0.15)		1.10 (0.04 to 48.83)	0.59 (0.07 to 5.25)
Nitrates	2.69 (0.01 to 817.29)	0.86 (0.00 to 323.11)	-	-		0.54 (0.01 to 14.60)
Oxytocin receptor blockers	4.78 (0.04 to 601.24)	-	0.01 (0.00 to 0.09)	-	-	

Table 65: Probability rankings for medicines by class for discontinuation of medicine because of maternal side effect. Ranking probability of best medicine classes for improving the outcome. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Placebo/control	73%	1.4	3	(3 to 3)
Nitrates	11%	3.3	3	(1 to 6)
Oxytocin receptor blockers	10%	2.5	2	(1 to 4)
Calcium channel blockers	6%	3.1	3	(1 to 5)
Betamimetics	0%	6.0	6	(1 to 6)
Magnesium sulfate	0%	4.7	5	(5 to 5)

Figure 14: Graphic representation of the each medicine's effectiveness probability ranking for improving outcome



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (6). The line indicates the probability that medicine will achieve each rank.

Table 66: Quality assessment of the evidence contributing to the outcome of discontinuation of treatment due to side effect

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Discontinuation due to side effect							
1 NMA of 36 studies (Original data from Haas 2012)	36 RCTsa	serious1	no serious inconsistency	serious2	serious3	none	Very low

[a] *Cotton 1984; Goodwin1996; Holleboom 1996; Papatsonis 1997/2000; Moutquin 2000; French/Australian 2001; Nassar 2009; Cox 1990; Laveno 1986; Larsen 1986; Smith 2007; Glock 1993; Essed 1978; Sirohiwal 2001; Rayamajhi, 2003; Al-Qattan 2000; Cararach 2006; Shim 2006; Trablsi 2008; Weerakul 2002; Beall 1985; Romero 2000; Hollander 1987; Wilkins 1988; Miller 1982; Surichamorn 2001; Chau 1992; Larmon 1999; Floyd 1995; El-Sayed 1999; Caritis 1984; Garcia-Velasco 1998; Van de Water 2008; Maitra 2007; Motazedian 2010; European 2001.*

[1] *Analysis was based on the class therefore different doses and co-treatment were combined together*

[2] *Women with multiple pregnancy were included in 29/36 studies*

[3] *Wide and very wide CrI across all comparisons except on (Placebo/control vs magnesium sulfate)*

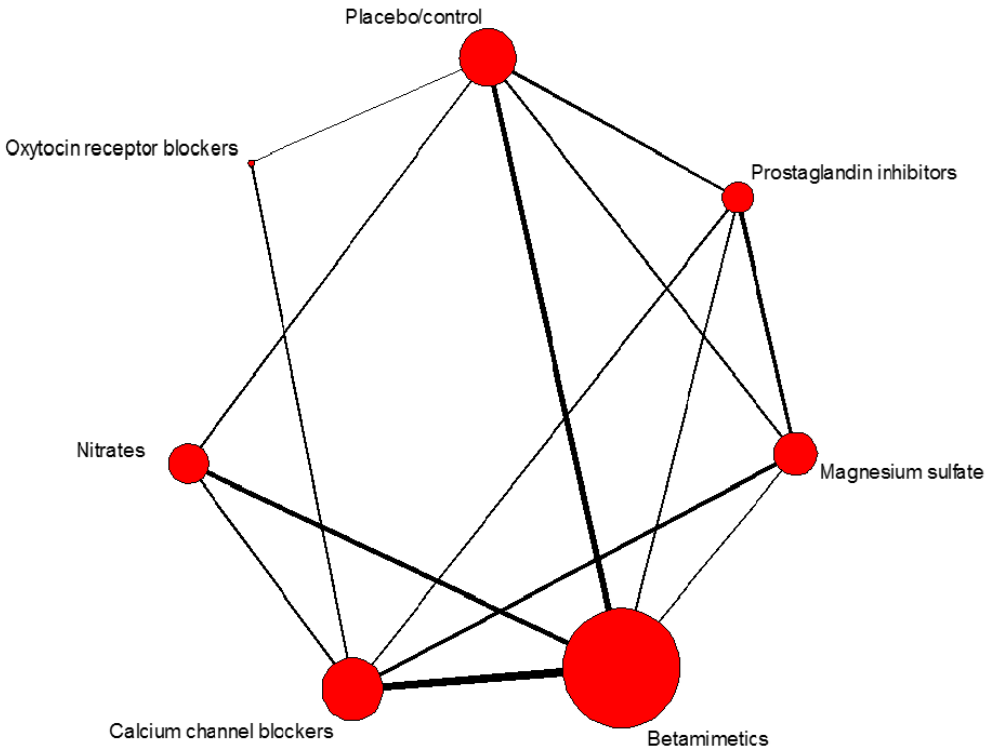
12.1.6.5 Estimated gestational age at birth

Out of the 98 studies included in the review, 53 studies reported mean gestational age as an outcome:

- 2 studies reported mean gestational age only on one arm of the study and were removed.

The remaining 51 studies examined 19 medicines allowing for 7 out of the 9 treatment classes to be assessed against each other.

Figure 15: Graphic representation of tocolytics trials for the NMA for estimated gestational age



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the linesis proportional to the number of studies contributing to the

comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 67: Posterior median of the mean difference and 95% credible Intervals (CrI) for gestational age at birth in weeks

Upper diagonal: values shown are the mean difference for the columns header versus the row header and are derived from the NMA. Given that this table relates to a positive and continuous outcome, values higher than 0 favour the column defining treatment and values lower than 0 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

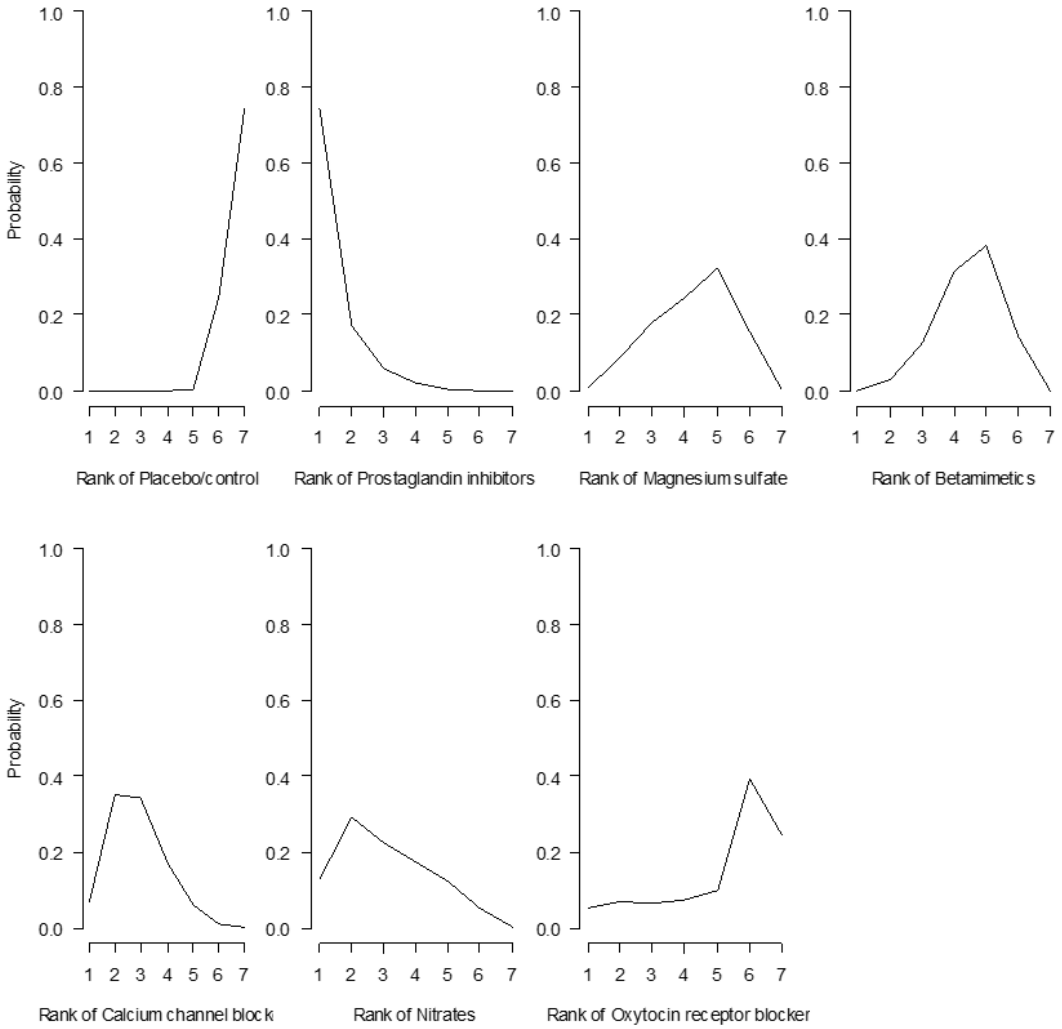
Lower diagonal: values shown are the mean difference for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to an adverse outcome, Lower diagonal: values lower than 0 favour the column defining treatment and values higher than 0 favour the row defining the treatment

	Placebo/control	Prostaglandin inhibitors	Magnesium sulfate	Betamimetics	Calcium channel blockers	Nitrates	Oxytocin receptor blockers
Placebo/control		2.32 (1.27,3.35)	1.29 (0.29,2.27)	1.25 (0.40,2.07)	1.69 (0.69,2.66)	1.65 (0.52,2.78)	0.68 (-1.32,2.67)
Prostaglandin inhibitors	3.27 (1.68,4.78)		-1.04 (-2.01,-0.04)	-1.08 (-2.08,-0.05)	-0.64 (-1.68,0.42)	-0.67 (-1.97,0.67)	-1.65 (-3.76,0.52)
Magnesium sulfate	-0.14 (-1.60,1.28)	0.92 (-1.73,3.57)		-0.04 (-0.99,0.91)	0.40 (-0.51,1.31)	0.36 (-0.88,1.63)	-0.61 (-2.69,1.50)
Betamimetics	1.91 (0.90,2.90)	-0.24 (-1.46,0.97)	-		0.44 (-0.32,1.20)	0.40 (-0.54,1.37)	-0.57 (-2.58,1.47)
Calcium channel blockers	-	-1.56 (-3.42,0.28)	-	-		-0.03 (-1.16,1.10)	-1.01 (-2.98,0.99)
Nitrates	1.09 (-1.79,4.00)	-0.53 (-2.32,1.25)	-0.19 (-2.78,2.45)	-	0.80 (-0.08,1.67)		-0.98 (-3.15,1.21)
Oxytocin receptor blockers	-0.51 (-3.00,2.01)	0.92 (-1.73,3.57)	-0.02 (-1.25,1.22)	-	-	0.58 (-0.47,1.67)	

Table 68: Ranking probability of best medicine classes for improving the outcome. Probability rankings for medicines by class for improving estimated gestational age. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Prostaglandin inhibitors	74%	1.38	1	(1 to 4)
Nitrates	13%	3.04	3	(1 to 6)
Calcium channel blockers	7%	2.84	3	(1 to 5)
Oxytocin receptor blockers	5%	5.27	6	(1 to 7)
Magnesium sulfate	1%	4.26	4	(2 to 6)
Betamimetics	0%	4.48	5	(2 to 6)
Placebo/control	0%	6.74	7	(6 to 7)

Figure 16: Graphic representation of the each medicine's effectiveness probability ranking for improving outcome



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (7). The line indicates the probability that medicine will achieve each rank.

Table 69: Quality assessment of the evidence contributing to the outcome of estimated gestational age

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Mean estimated gestational age							
1 NMA of 51 studies (Original data from Haas 2012)	51 RCTs[a]	serious[1]	serious[2]	serious[3]	serious[4]	None	Very low

[a] Cotton 1984; Klauser 2012; Goodwin1996; Niebyl 1980; Weiner 1988; Stika 2002; Al-Omari 2006; Kurki 1991; McWhorter 2004; Lyell 2007; Holleboom 1996; Papatsonis 1997/2000; Moutquin 2000; French/Australian 2001;; Nassar 2009, Sawdy 2003; Zuckerman 1984; Panter 1999; Cox 1990; How 2006; Casapo 1977; CPLIG 1992; merkatz 1980; Laveno 1986; Larsen 1986; Smith 2007; Borna 2007; Rasanen 1995; Parilla 1997; Besinger 1991; Kashanian 2011; Schorr 1998; Surichamorn 2001; Larmon 1999; Taherian 2007; Glock 1993; Essed 1978; Sirohiwal 2001; Rayamajhi 2003; Al-Qattan 2000; Cararach 2006; Fan 2003; Koks 1998; Lin 2009; Shim 2006; Neri 2009; Jannet 1997; Trablsi 2008; Weerakul 2002; Kashanian 2014; Salim 2012.
[1] Analysis was based on the class therefore different doses and co-treatment were combined together
[2] Women with multiple pregnancy were included in 36/49 studies
[3] Some evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of Placebo/control v Magnesium sulfate
Bayesian p-value= 0.015
Direct OR = 0.87 (95%CrI 0.20, 3.60) 3 small studies compare these classes directly
Indirect OR = 1.29 (95%CrI 0.29, 2.27)
The direct and indirect are contradictory to each other (both on opposite sides of one). Results from this network should be considered with caution[4] Wide and very wide CrI across all comparisons except on (placebo/control vs magnesium sulfate)

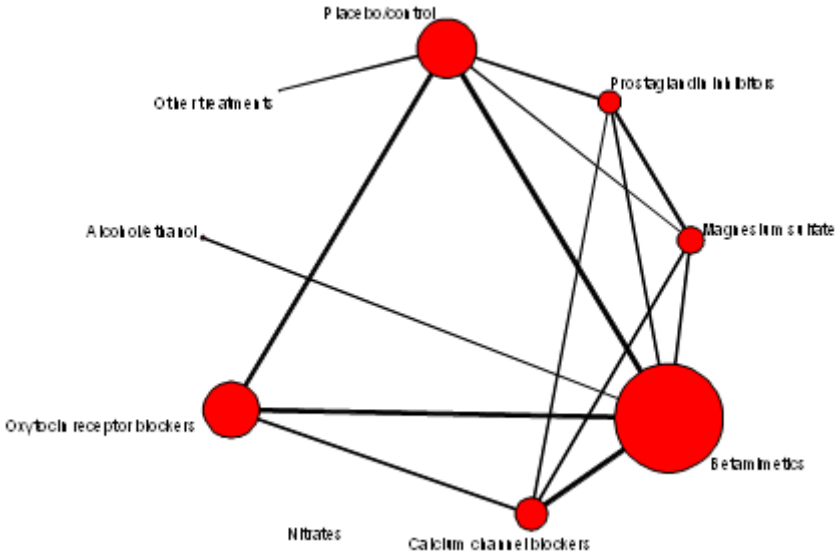
12.1.6.6 Respiratory distress syndrome (RDS)

Out of the 98 studies included in the review, 57 studies reported RDS as an outcome:

- 1 study observed no events and was removed
- 7 studies only reported events on one arm and were also removed.

The remaining 49 studies examined 22 medicines allowing for 8 out of the 9 treatment classes to be assessed against each other.

Figure 17: Graphic representation of tocolytics trials for the NMA for RDS



How to read the network diagrams: Lines represent trials comparing 2 classes of medicine. The thickness of the linesis proportional to the number of studies contributing to the

comparison. The size of the dots is proportional to the number of participants randomised to the treatment.

Table 70: Posterior median of the odds ratios (OR) and 95% credible Intervals (CrI) for respiratory distress syndrome

Upper diagonal: values shown are the OR for the columns header versus the row header and are derived from the NMA. Given that this table relates to an adverse outcome, values lower than 1 favour the column defining treatment and values higher than 1 favour the row defining the treatment. Upper diagonal data is also presented in the forest plots in Appendix I.

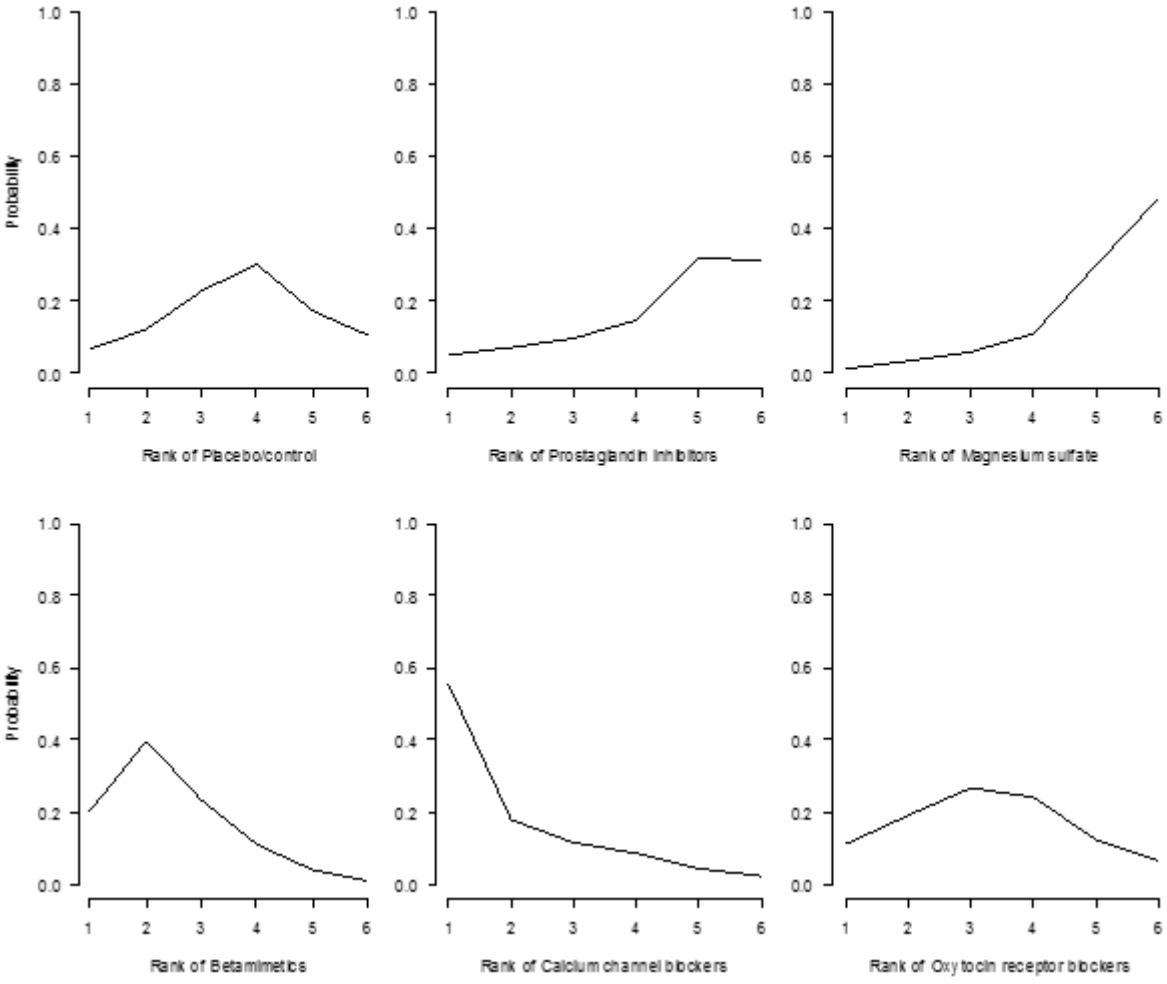
Lower diagonal: values shown are the OR for the row header versus the column header and are derived from the direct comparison analysis. Given that this table relates to an adverse outcome, values higher than 1 favour the column defining treatment and values lower than 1 favour the row defining the treatment.

	Placebo/control	Prostaglandin inhibitors	Magnesium sulfate	Betamimetics	Calcium channel blockers	Oxytocin receptor blockers	Alcohol/ethanol	Other treatments
Placebo/control		1.13 (0.68 to 1.86)	1.20 (0.76 to 1.90)	0.88 (0.65 to 1.23)	0.81 (0.50 to 1.34)	0.96 (0.66 to 1.43)	2.55 (0.78 to 9.05)	0.75 (0.26 to 2.21)
Prostaglandin inhibitors	1.01 (0.33, to 3.10)		1.06 (0.69 to 1.66)	0.78 (0.49 to 1.28)	0.71 (0.41 to 1.29)	0.85 (0.52 to 1.42)	2.25 (0.65 to 9.05)	0.75 (0.26 to 2.21)
Magnesium sulfate	1.26 (0.58 to 2.72)	1.02 (0.64 to 1.65)		0.73 (0.47 to 1.15)	0.67 (0.41 to 1.12)	0.80 (0.51 to 1.29)	2.12 (0.62 to 7.86)	0.63 (0.19 to 2.02)
Betamimetics	0.72 (0.50 to 1.04)	1.25 (0.48 to 3.25)	0.53 (0.05 to 4.31)		0.92 (0.61 to 1.39)	1.08 (0.77 to 1.54)	2.88 (0.92 to 9.75)	0.85 (0.28 to 2.59)
Calcium channel blockers	-	-	0.85 (0.41 to 1.72)	0.74 (0.45 to 1.22)		1.19 (0.73 to 1.90)	3.14 (0.93 to 11.33)	0.93 (0.28 to 3.01)
Oxytocin receptor blockers	1.50 (0.90 to 2.95)	0.60 (0.29 to 1.22)	-	1.00 (0.65 to 1.62)	0.77(0.27 to 2.14)		2.65 (0.80 to 9.46)	0.79 (0.25 to 2.41)
Alcohol/ethanol	-	-	-	2.88 (0.96 to 9.45)	-	-		0.29 (0.06 to 1.46)
Other treatments	0.75 (0.26 to 2.16)	-	-	-	-	-	-	

Table 71: Probability rankings for medicines by class to reduce respiratory distress syndrome (RDS). Ranking probability of best medicine classes for improving the outcome. Rows arranged in the decreasing order of estimate effect: best treatment at the top and the worst at the bottom of the table.

Class	Probability of being the best treatment option to improve the outcome	Mean rank	Median rank	Rank 95% credible interval
Calcium channel blockers	55%	2.	1	(1 to 5)
Betamimetics	20%	2.4	2	(1 to 5)
Oxytocin receptor blockers	11%	3.3	3	(1 to 6)
Placebo/control	7%	3.7	4	(1 to 6)
Prostaglandin inhibitors	5%	4.6	5	(1 to 6)
Magnesium sulfate	1%	5.1	5	(2 to 6)

Figure 18: Graphic representation of the each medicine's effectiveness probability ranking for improving outcome



How to read the rankograms: The numbers on the y axis indicate probability. The numbers on the x axis show the potential effectiveness ranks from most effective (1) to least effective (6). The line indicates the probability that medicine will achieve each rank.

Table 72: Quality assessment of the evidence contributing to the outcome of RDS

Quality assessment							
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
RDS							
1 NMA of 47 studies (Original data from Haas 2012)	49 RCTs[a]	serious[1]	serious[2]	serious[3]	serious[4]	none	Very low

[a] *Thornton 2009; Cotton 1984; Klauser 2012; Goodwin 1996; Niebyl 1980; Panter 1999; Zuckerman 1984; Cox 1990; Spellacy 1979; Larsen 1986; Merkatz 1980; Leveno 1986; CPLIG 1992; Goodwin 1994; Romero 2000; Weiner 1988; Stika 2002; Rasanen 1995; Morales 1993; Parilla 1997; Morales 1989; Kurki 1991; Schorr 1998; McWhorter 2004; Miller 1982; Floyd 1995; Lyell 2007; Essed 1978; Gummerus 1983; Holleboom 1996; Caritis 1984; Maitra 2007; Cararach 2006; VandeWater 2008; Al-Qattan 2000; Papatsonis 1997; Papatsonis 2000; Shim, 2006; Lin 2009; Moutquin 2000; Lauersen 1977; Trabelsi 2008; French/Australian 2001; Laohapojanart 2007; European 2001; Nassar 2009; Al-Omari 2006; Jaju 2011; Salim 2012.*

[1] *Analysis was based on the class therefore different doses and co-treatment were combined together*

[2] *There were some evidence of inconsistency (conflict between direct and indirect evidence) in comparisons of placebo/control vs. betamimetics and placebo/control vs. oxytocin receptor blockers:*

– *Placebo/control v Betamimetics*
  Bayesian p-value= 0.034
  Direct OR= 0.72 95%CrI (0.51, 1.02)
  Indirect OR=1.48 95%CrI (0.84, 2.56)
– *Placebo/control v Oxytocin receptor blockers*
  Bayesian p-value= 0.015
  Direct OR= 1.49 95%CrI (0.91, 2.72)
  Indirect OR=0.63 95%CrI (0.40, 1.02)

[3] *Women with multiple pregnancy were included in 31/47 studies*

[4] *Wide and very wide CrIs across all comparisons except 2 (magnesium sulfate vs betamimetics and magnesium sulfate vs calcium channel blockers)*

12.1.7 Evidence profiles for the pairwise comparisons

Table 73: GRADE profile for the comparison of placebo versus indomethacin

Quality assessment							Summary of findings				
							Frequency[a] (%)/ mean (SD)		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Placebo	Indomethacin	Relative (95% CI)	Absolute effect (95% CI)	Quality
Maternal infection											
2 studies (Niebyl 1980 & Panter 1999)	RCT	serious[1]	no serious inconsistency	serious[2]	Very serious[3]	None	2/33	4/31	RR 0.48 (0.09 to 2.46)	67 fewer per 1000 (from 117 fewer to 188 more)	Very low
Chronic lung disease (CLD)											
2 studies (Niebyl 1980 & Panter 1999)	RCT	serious[1]	no serious inconsistency	serious[2]	very serious[3]	None	4/35	5/35	RR 0.80 (0.23 to 2.73)	29 fewer per 1000 (from 110 fewer to 247 more)	Very low

RR risk ratio, CI confidence interval, NC not calculable
[1.] No clear inclusion/exclusion criteria hence high risk of selection bias
[2] Multiple pregnancy included
[3] Evidence was downgraded by 2 due to serious imprecision as 95% confidence interval crossed 2 default MID
[a] Denominator for the outcome of maternal infection was the number of women whereas for the outcome of chronic lung disease the number of babies.

Table 74: GRADE profile for the comparison of placebo versus nitrates

Quality assessment							Summary of findings				
							No. of babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Placebo	Nitrates	Relative/ RR (95% CI)	Absolute effect size (95% CI)	Quality
Chronic lung disease (CLD)											
1 study (Smith 2007)	RCT	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	None	7/79	1/74	RR 7.00 (0.13 to 3.44)	76 more per 1000 (from 11 fewer to 31 more)	Very low
Periventricular leucomalacia (PVL)											
1 study (Smith 2007)	RCT	no serious	no serious inconsistency	serious[1]	Very serious[2]	none	Placebo 2/79	Nitrates 0/74	RR 4.81 (0.23 to 101.79)	160 more per 1000 (from 32 fewer to 1000 more)	Very low

Quality assessment							Summary of findings				
							No. of babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Placebo	Nitrates	Relative/ RR (95% CI)	Absolute effect size (95% CI)	Quality
		risk of bias									

RR risk ratio, CI confidence interval, NC not calculable
[1] Multiple pregnancy included
[2] Evidence was downgraded by 2 due to serious imprecision as 95% confidence interval crossed 2 default MID

Table 75: GRADE profile for the comparison of betamimetics versus nitrates

Quality assessment							Summary of findings				
							No. of babies		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Beta-mimetics	Nitrates	Relative/ RR (95% CI)	Absolute (95% CI)	Quality
Chronic lung disease (CLD)											
1 study (Bisits 2004)	RCT	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	None	9/116	9/120	RR 1.03 (0.43 to 2.51)	2 more per 1000 (43 fewer to 113 more)	Very low

RR risk ratio, CI confidence interval, NC not calculable
[1] Multiple pregnancy included
[2] Evidence was downgraded by 2 due to serious imprecision as 95% confidence interval crossed 2 default MID

Table 76: GRADE profile for comparison of nifedipine versus ritodrine

Quality assessment							Summary of findings				
							Frequency (%)/ mean (SD)		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Nifedipine	Ritodrine	Relative RR/mean difference (95% CI)	Absolute effect	Quality
Periventricular leucomalacia (PVL)											
1 study (Paptsonis 2000)	RCT	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	None	1/95	5/90	RR 0.18 (0.02 to 1.58)	46 fewer per 1000 (from 54 fewer to 24 more)	Very low
Behaviour emotional functioning (follow up at age of 9 – 12 year) – measured using child behaviour checklist– higher score represent more psychosocial problem[a]											
1 study (Houtzager 2006)	RCT	serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	None	50 (11.9)	52 (11.6)	MD -2 (-6.57 to 2.57)		Low

Quality assessment							Summary of findings				
							Frequency (%)/ mean (SD)		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Nifedipine	Ritodrine	Relative RR/mean difference (95% CI)	Absolute effect	Quality
Behaviour emotional functioning (follow up at age of 9 – 12 year) – measured using teacher report form– higher score represent more psychosocial problem[a]											
1 study (Houtzager 2006)	RCT	serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	None	49 (10)	50 (9.9)	MD -1 (-4.87 to 2.87)		Low
Children's physical quality of life (QoL) (follow up at age of 9 – 12 year) – measured using teacher report form– higher score represent more favourable QoL[a]											
1 study (Houtzager 2006)	RCT	serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	None	25 (5.3)	26 (4.5)	MD -1 (-2.92 to 0.92)		Low
Children's motor quality of life (QoL) (follow up at age of 9 – 12 year) – measured using teacher report form– higher score represent more favourable QoL[a]											
1 study (Houtzager 2006)	RCT	serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	None	30 (3.1)	30 (2.5)	MD 0 (-1.10 to 1.10)		Moderate
Children's autonomy quality of life (QoL) (follow up at age of 9 – 12 year)- measured using teacher report form– higher score represent more favourable QoL[a]											
1 study (Houtzager 2006)	RCT	serious1[,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	None	31 (1.2)	31 (1.6)	MD 0 (-0.55 to 0.55)		Moderate
Children's cognitive quality of life (QoL) (follow up at age of 9 – 12 year)- measured using teacher report form– higher score represent more favourable QoL[a]											
1 study (Houtzager 2006)	RCT	serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	None	28 (4)	28 (3.8)	MD 0 (-1.52 to 1.52)		Moderate
Children's positive emotion quality of life (QoL) (follow up at age of 9 – 12 year)– measured using teacher report form– higher score represent more favourable QoL[a]											
1 study (Houtzager 2006)	RCT	serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	None	13 (2.7)	14 (2.4)	MD -1 (-2.00 to 0.00)		Low
Children's negative emotion quality of life (QoL) (follow up at age of 9 – 12 year) - measured using teacher report form– higher score represent more favourable QoL[a]											
1 study (Houtzager 2006)	RCT	serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	None	12 (2.7)	13 (2.3)	MD -1 (-1.98 to 0.02)		Low
Motor quality (follow up at age of 9 – 12 year)- movement ABC– higher score represent more motor problem[a]											
1 study (Houtzager 2006)	RCT	Very serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	None	mean 5 (6.9)	mean 9.3 (17.2)	MD -4.30 (-9.29 to 0.69)		Very low

1 *RR risk ratio, CI confidence interval, NC not calculable*

a The child long term outcomes were assessed using the Dutch version of the Child Behaviour Checklist (CBCL) completed by parents. The child' teacher completed the teacher Report Form (TRF). High score on the CBCL and TRF represent more problematic behaviour. Total score is for internalising problem such as anxiety, depression, or social behaviour, non-compliance, or hyper activity.
The child quality of life (QoL) was assessed using the Dutch TNO AZL Children's Quality of Life Questionnaire (TACQOL). The questionnaire provides score based on the 7 domains: physical functioning, motor functioning, autonomy, cognitive emotions. High score represent a more favourable QoL.
1 no clear inclusion/exclusion criteria hence high risk of selection bias
2 low response rate; 65% in nifedipine and 55% in ritodrine group
3 unclear if evaluation tools were validated
4 evidence was downgraded by 1 due to serious imprecision as 95% confidence interval crossed one default MID

Table 77: GRADE profile for the comparison of placebo versus indomethacin

Quality assessment							Summary of findings				
							No. of babies –		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Placebo	Indomethacin	Relative/ RR (95% CI)	Absolute (95% CI)	Quality
Periventricular leucomalacia (PVL)											
1 study (Panter 1999)	RCT	no serious risk of bias	no serious inconsistency	serious[1]	very serious[2]	None	0/20	1/19	RR 0.30 (0.01 to 7.85)	37 fewer per 1000 (from 52 fewer to 36 more)	Very low

RR risk ratio, CI confidence interval, NC not calculable
[1] multiple pregnancy included
[2] Evidence was downgraded by 2 due to serious imprecision as 95% confidence interval crossed 2 default MID

Table 78: GRADE profile for comparison of indomethacin versus magnesium sulfate

Quality assessment							Summary of findings				
							No. of babies –		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Indomethacin	Magnesium sulfate	Relative/ RR (95% CI)	Absolute (95% CI)	Quality
Periventricular leucomalacia (PVL)											
1 study (Klauser 2012)	RCT	no serious risk of bias	no serious inconsistency	serious[1]	No serious imprecision	none	2/103	0/95	NC	NC	Moderate

RR risk ratio, CI confidence interval, NC not calculable
1 multiple pregnancy included

Table 79: GRADE profile for comparison of indomethacin versus magnesium sulfate

Quality assessment							Summary of findings				
							No. of babies –		Effect		
No. of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Indomethacin	Nifedipine	Relative/ RR (95% CI)	Absolute (95% CI)	Quality
Periventricular leucomalacia (PVL)											
1 study (Klauser 2012)	RCT	no serious risk of bias	no serious inconsistency	serious[1]	No serious imprecision	None	2/103	0/119	NC	NC	Moderate

RR risk ratio, CI confidence interval, NC not calculable
1 multiple pregnancy included

12.1.8 Economic evidence

This question was prioritised for health economic analysis.

A de Novo health economic model was developed using the evidence from the NMA undertaken for this guideline on neonatal mortality, intraventricular haemorrhage and respiratory distress syndrome. As far as we are aware this is the first economic evaluation of tocolysis where relative treatment effects are based on the results of a NMA.

Using a cost-utility analysis approach a range of tocolytics (prostaglandin inhibitors, betamimetics, calcium channel blockers, magnesium sulfate, nitrates, oxytocin receptor blockers) were compared as well as standard care (no tocolytic) in women with between 24^{+0} and 34^{+0} weeks of pregnancy in suspected or diagnosed preterm labour.

The base case analysis found that calcium channel blockers were the most cost-effective treatment across all gestational ages considered in the model as reflected by its net mean benefit which was the highest across 10,000 Monte Carlo simulations. Oxytocin receptor blockers had the second highest net mean benefit. Nitrates actually had a slightly higher probability of being cost-effective than calcium channel blockers (e.g. 36% versus 34% at 24 weeks gestational age) but this reflects the wider confidence intervals for nitrates. A sensitivity analysis suggested that changing the assumptions with respect to the loss in Quality Adjusted Life Years (QALYs) from RDS and IVH had a negligible impact on the model results. It was also clear from the net mean benefit achieved with calcium channel blockers relative to the alternatives, that treatment costs were not an important driver of the cost-effectiveness results. Although oxytocin receptor blockers had the highest cost of all treatment options they were found to have the second highest net mean benefit across all gestational ages included in the model.

The model is described in greater detail in Chapter 16.

12.1.9 Evidence statements on NMA results

12.1.9.1 Neonatal mortality

Randomised very low quality evidence from the NMA on 36 tocolytics treatments from 9 classes (placebo, prostaglandin inhibitors, magnesium sulphate, betamimetics, calcium channel blockers, nitrates, oxytocin receptor blockers, alcohol/ethanol, other treatments) with a total sample size of almost 7000 women in diagnosed or suspected preterm labour showed that no medicine class was significantly better than placebo for reducing neonatal mortality. However, results should be interpreted with caution due to high uncertainty in the effect estimates and the inconsistency between direct and indirect analysis for a medicine comparison.

12.1.9.2 Perinatal mortality

Randomised very low quality evidence from the NMA on 35 tocolytics treatments from 9 classes (placebo, prostaglandin inhibitors, magnesium sulphate, betamimetics, calcium channel blockers, nitrates, oxytocin receptor blockers, alcohol/ethanol, other treatments) with a total sample size of over 6000 women in diagnosed or suspected preterm labour showed that nitrates were more effective than all other medicine classes (including placebo) for reducing perinatal mortality, though there was considerable uncertainty in estimates of their efficacy. They had the highest probability (89%) of being the most effective medicine class for this outcome. Prostaglandin inhibitors had the next highest probability of being best (5%). The results of this model showed some inconsistencies between direct and indirect analyses.

12.1.9.3 Delay birth by more than 48 hours

Randomised very low quality evidence on 36 treatments from 9 classes (placebo, prostaglandin inhibitors, magnesium sulphate, betamimetics, calcium channel blockers, nitrates, oxytocin receptor blockers, alcohol/ethanol, other treatments) in the NMA (with a total sample size of almost 8000 women in diagnosed or suspected preterm labour) showed that prostaglandin inhibitors were both more effective than all other medicine classes (including placebo) for delaying birth by 48 hours and had the highest probability (76%) of being the best treatment class compared to other medicine classes. Oxytocin receptors blockers and magnesium sulfate had the next highest probability of being ranked best (8 and 7%). Calcium blockers were not found to significantly improve this outcome compared to placebo.

The results from direct comparisons between medicine classes were not always consistent with those from NMA.

12.1.9.4 Neonatal sepsis

Very low quality evidence from 21 RCTs investigating 7 classes (placebo, prostaglandin inhibitors, magnesium sulphate, betamimetics, calcium channel blockers, oxytocin receptor blockers, other treatments) in the NMA with almost 2500 women in diagnosed or suspected preterm labour found:

- that calcium channel blockers were more effective than all other tocolytic medicines compared to placebo for reducing neonatal sepsis and that calcium channel blockers also had the highest probability (45%) of being the most effective medicine class compared to other medicine classes
- placebo/control and oxytocin receptor blockers had the next highest probability of being the most effective with 38% and 11% probability respectively

The result from direct comparisons was not always consistent with those from NMA.

12.1.9.5 Intraventricular haemorrhage

Very low quality evidence from 28 RCTs investigating 8 classes (placebo, prostaglandin inhibitors, magnesium sulphate, betamimetics, calcium channel blockers, nitrates, oxytocin receptor blockers, other treatments) in the NMA with over 5 thousand women in diagnosed or suspected preterm labour found:

- nitrates were more effective than all other medicine classes compared to placebo for reducing intraventricular haemorrhage and that nitrates also had the highest probability (59%) of being the most effective medicine class compared to other medicine classes
- calcium channel blockers had the next highest probability of being the most effective medicine class compared to other medicine classes (38%)

The result from direct was not always consistent with those from meta-analysis.

12.1.9.6 Discontinuation of treatment due to adverse events effect

Very low quality evidence from 15 RCTs investigating 6 classes (placebo, magnesium sulphate, betamimetics, calcium channel blockers, nitrates, oxytocin receptor blockers) in the NMA with 4000 and 410 women in diagnosed or suspected preterm labour found:

- placebo/control were more effective than all other medicine classes for not causing maternal adverse events and discontinuation of treatment and that placebo/control treatment also had the highest probability (73%) of being the most effective medicine class compared to other medicine classes.
- Nitrates and oxytocin receptor blocker had the next highest probability of being the most effective with 11% and 10% probability respectively.

The result from direct was consistent with those from meta-analysis.

12.1.9.7 Estimated gestational age at birth

Randomised very low quality evidence from 28 treatments investigating 7 classes (placebo, prostaglandin inhibitors, magnesium sulphate, betamimetics, calcium channel blockers, nitrates, oxytocin receptor blockers) in the NMA (total sample size over 5500 women in diagnosed or suspected preterm labour) concluded:

- prostaglandin inhibitors were more effective than all other medicine classes compared to placebo for increasing estimated gestational age and that prostaglandin inhibitors also had the highest probability (64%) of being the most effective medicine class compared to other medicine classes
- Nitrates and calcium channel blockers had the next highest probability of being the most effective with 21% and 9% probability respectively.

The result from direct comparisons was not always consistent with those from NMA.

12.1.9.8 Respiratory distress syndrome

Very low quality randomised evidence from 28 treatments investigating 8 classes (placebo, prostaglandin inhibitors, magnesium sulphate, betamimetics, calcium channel blockers, oxytocin receptor blockers, alcohol/ethanol, other treatments) contributed to the NMA for this outcome (total sample size over 5500 women in diagnosed or suspected pre term labour) showed that:

- calcium channel blocker were more effective than all medicine classes compared to placebo for reducing respiratory distress syndrome and that calcium channel blockers also had the highest probability (55%) of being the best medicine class compared to other medicine classes.
- Beta-mimetics and oxytocin receptor blockers had the next highest probability of being the most effective with 20% and 11% probability respectively.

The result from direct comparisons was not always consistent with those from NMA.

12.1.10 Evidence statements on pair-wise comparisons

Evidence from meta-analysis of 2 RCTs with over 60 participants found no significant difference for the outcomes of maternal infection and chronic lung disease (CLD) in women treated with indomethacin compared with those received placebo. The quality of the evidence was of very low quality.

Very low quality evidence from 10 RCT with 112 participants found no significant difference in the rate of CLD between babies whose mothers were treated with nitrates compared with those on placebo.

Evidence from 10 other RCT with 236 participants found no significant difference in the rate of CLD in babies whose mothers treated with beta-mimetics compared with those whose mothers were treated with nitrates.The quality of the evidence was of moderate quality.

Very low quality evidence from 2 studies with 192 participants found no significant difference in the rate of periventricular leucomalacia (PVL) in babies whose mothers were treated with placebo compared with those whose mothers were treated with indomethacin or nitrates.

Moderate evidence from 2 other studies with 407 participants found no difference in the rate of PVL in babies whose mothers were treated with nifedipine compared with those whose mothers were treated with ritodrine, and those whose mothers were treated with indomethacin compared with magnesium sulfate.

12.1.10.1 Long-term psychosocial functioning (follow up at age of 9 - 12 year)

Evidence from a follow up of a RCT with 102 participants found no significant difference between children whose mothers were treated with nifedipine compared with ritodrine for the behaviour emotional functioning and quality of life (physical, motor, autonomy, cognitive, positive emotion). Children whose mothers were treated with nifedipine had significantly lower negative emotion quality of life score compared with children whose mothers were treated with ritodrine. The quality of the evidence was moderate to low.

12.1.11 Evidence to recommendations

12.1.11.1 Relative value placed on the outcomes considered

The Committee considered both neonatal and maternal outcomes for this review question. In relation to neonatal outcomes, neonatal and perinatal mortality, delay of birth by at least 48 hours, neonatal sepsis, chronic lung disease, intraventricular haemorrhage, gestational age at birth, respiratory distress syndrome, periventricular leucomalacia, and neurodevelopmental disability were considered. However, neonatal and perinatal mortality and respiratory distress syndrome were considered the most critical outcomes for decision making.

Among the maternal outcomes, the Committee included maternal infection, maternal mortality and discontinuation of treatment due to maternal adverse events as the most important outcomes. Discontinuation of treatment due to maternal adverse events was the only maternal outcomes prioritised in the NMA.

12.1.11.2 Consideration of clinical benefits and harms

In relation to the most critical neonatal outcomes for decision making (neonatal and perinatal mortality and respiratory distress syndrome), the evidence reviewed in the NMA showed that calcium channel blockers had the highest probability of being the best medicine for reducing respiratory distress syndrome and were more effective for this outcome compared to the other tocolytics used in the NMA (placebo, prostaglandin inhibitors, magnesium sulfate, betamimetics, oxytocin receptor blockers, alcohol and other treatments) when given to women at suspected or diagnosed preterm labour. No one class of treatment reviewed was found more effective than placebo for improving neonatal mortality whereas nitrates were found more effective than all other drugs and with the highest probability to be the best medicine to improve perinatal mortality.

With regards to other outcomes, NMA results showed that calcium blockers were also beneficial in terms of protecting preterm babies from neonatal sepsis and were the second best treatment for improving intraventricular haemorrhage (IVH) and the third best treatment on increasing gestation age at birth. Prostaglandin inhibitors were found to be the most beneficial treatment in terms of delaying birth by more than 48 hours and for increasing estimated gestational age, and the second most effective treatment for reducing perinatal mortality. However, prostaglandin inhibitors were not found significantly better than calcium channel blockers on delaying birth by more than 48 hours, which were significantly better in achieving this outcome. In addition, prostaglandins inhibitors were not found to improve the ‘harder’ outcomes such as neonatal mortality, respiratory distress syndrome and neonatal sepsis in all of which they scored very low in the ranking of best treatments. The Committee was also aware of other harms thought to be associated with prostaglandin inhibitors such as premature closure of the ductus arteriosus therefore they didn’t consider them as a tocolytic option for women in suspected or diagnosed preterm labour.

The NMA results also found that nitrates were the most effective treatment for IVH and the second best, along with oxytocin receptors, to reduce the rate of discontinuation of treatment due to adverse events and to increase estimated gestational age, but the Committee

discussed that this benefit of nitrates needs to be balanced against the potential harm to the fetus. However, the number of trials including this treatment was small (only 6) and therefore results should be interpreted with caution. In addition, nitrates did not connect to the NMA network for the outcome of respiratory distress syndrome (10 of the selected critical outcomes), therefore the Committee was uncertain on the effectiveness of this intervention in relation to improving this outcome. Nitrates were also not found significantly better than placebo on improving the outcome of chronic lung disease in pair-wise comparisons.The use of oxytocin receptor blockers for reduction of maternal side effects and increasing gestational age has to be balanced against its poor efficacy in reducing IVH and RDS and it modest effect on perinatal mortality so the Committee decided that this should not be the first option of tocolytic treatment.

In relation to maternal outcomes, evidence from the NMA showed all reviewed treatments has an unfavourable effect on discontinuation of treatment due to adverse events effects. The evidence from the pair-wise comparison showed that indomethacin was not significantly more harmful than placebo on the outcome of maternal infection.

Betamimetics did not score highly in terms of clinical effectiveness for any of the outcomes reviewed in the NMA or pair wise meta-analysis and the Committee confirmed that their use should not be considered for tocolytic treatment for suspected or diagnosed women at preterm labour.

The Guideline Committee discussed that in their own clinical experience, the most frequent clinical case scenario of women in diagnosed or suspected preterm labour would be to administer magnesium sulfate to improve the baby's neuroprotection. The question the Committee aimed to address is whether there is any additional clinical benefit, with minimal harms for both the baby and the mother, from adding another tocolytic (in addition to magnesium sulfate). The included evidence did not provide further information on whether magnesium sulfate had been already prescribed for neuroprotection prior to a decision being made about the use of other tocolytics. However, the Committee highlighted that magnesium sulfate has only been in routine use for this reason in approximately the last 5 years, so in older studies this would not have been relevant. Based on their clinical experience, no adverse interaction was anticipated by using a combination of magnesium sulfate and another tocolytic medicine (for example increasing the frequency of adverse events).

The average gestational age profile of women included in the evidence for this section was 26 weeks but the range was wider and covered women between 24 and 36weeks of gestation. The Committee discussed the role of tocolytics by gestational age and recognised the lack of data for the effectiveness and/ or harm of tocolytics on the fetus at a gestational age below 26 weeks.

The Committee recognised that the clinical decision to start tocolytic treatment needs to take into consideration a range of maternal factors such as the woman's status in the care pathway (whether in suspected or diagnosed preterm labour) and the coexistence of other features such as bleeding and infection, for which circumstances delaying preterm labour would be contraindicated. In relation to neonatal considerations, the decision to offer tocolytic treatment should assess the likely benefit of maternal corticosteroids, the gestational age and the impact of prolonging birth, and the availability of neonatal care in the care setting or the need for transfer to another hospital unit. Women's preference on starting tocolytics should also been taken into consideration in the planning of care.

12.1.11.3 Consideration of economic benefits and harms

The Committee discussed that there is clear potential for effective tocolysis to be cost-effective as the adverse outcomes of preterm birth are associated with significant losses in health related quality of life and large subsequent health care expenditure.

The health economic model produced for this review question focused on 3 (neonatal, perinatal mortality and respiratory distress syndrome) of the outcomes prioritised in the NMA. These outcomes were also considered to have the greatest impact on health related quality of life. The decision to focus on these 3 outcomes was to avoid double counting as it would be the case for some other outcomes (for example the impact of medicines on delaying birth by 48 hours and increasing gestational age). The relative treatment effect derived from the results of NMA on neonatal mortality was used to estimate the absolute mortality risk across the entire perinatal and neonatal period with the various treatment alternatives. This model found that calcium channel blockers were the most cost-effective treatment for women of all gestational ages.

One advantage of health economic evaluation is that it allows for benefits from different outcomes to be synthesised into a single measure, in this case the QALY. This allows explicit trade-offs to be made across outcomes in a way that is conceptually difficult when comparing different outcomes in isolation. The 3 NMAs underpin the cost-effectiveness findings but it is worth highlighting that the result, whilst taking uncertainty into account through the tool of probabilistic sensitivity analysis, is not governed by the classical rules of statistical inference. So in the network analyses included in the health economic model, calcium channel blockers were either the “best” treatment or 1 of the “best”. Being the best doesn’t necessarily mean that it is possible to reject a null hypothesis of no difference against an alternative using the usual, but arbitrary, 5% statistical significance level for a particular outcome but it does mean that when averaged across the simulations this will have a very important bearing on the cost-effectiveness result. Furthermore, the implication of not recommending the most cost-effective treatment will be that a treatment (or no treatment) alternative which is less likely to be cost-effective is used instead.

Although the model was not particularly sensitive to treatment costs, calcium channel blockers are 1 of the cheapest tocolytics and may be cost saving at the lower gestational ages as “downstream” savings from averted adverse outcomes more than offset treatment cost.

The Committee felt that based on the available economic evidence,it would be reasonable to recommend calcium channel blockers as a first line tocolytic treatment. The Committee discussed that this wouldn’t deviate from current practice in many settings providing care for women at risk of preterm labour. In addition the Committee thought that oxytocin receptor blockers should be offered to those in whom calcium channel blockers were contraindicated as the model provided evidence that they were the most cost-effective treatment option after calcium channel blockers.

12.1.11.4 Quality of evidence

The quality of evidence included in the NMA was very low mainly due to indirectness as almost half of the included studies involved multiple pregnancies and imprecision around the effect size. There were considerable inconsistencies observed between direct and indirect evidence for most of the outcomes selected in the NMA.

There were limitations in the data included in the NMA; most comparisons were only made in 1 or 2 trials. Furthermore, not all trials report all outcomes so some networks were very sparse in terms of patient numbers contributing to each loop. In addition, some trials had zero events in all arms and could not contribute to the estimation of treatment effects and were removed. The variation in results between different NMAs also make overall assessment of treatment options difficult for clinical interpretation.

Because some studies in the NMA included multiple births, allowing more than 1 infant per mother, it was not always clear which was the most appropriate number of individuals to consider for neonatal/ infant outcomes. Where available we used the number of infants as the denominator. Although this does not account for the expected correlation in outcomes of

infants from the same mother, it prevents double counting of infants from the same mother who may both (twins) have had an event.

The evidence included in the pair-wise comparisons (outcomes in the protocol that were not prioritised in the NMA) was also of moderate to very low quality given the high risk of selection bias, imprecision and indirectness (a proportion of women in studies had multiple pregnancies).

12.1.11.5 Other considerations

The Committee stressed the importance of information provision to women in suspected or diagnosed pre-term labour about the effect of different tocolytics on child and mother outcomes. It was discussed how different women may place different importance on the different maternal and neonatal outcomes and this may be affected by any previous similar pre-term delivery experience. Therefore, the Committee emphasised the role of provision of information on the role of tocolytics on delaying birth and improving neonatal outcomes.

12.1.11.6 Key conclusions

The guideline developers concluded that:

- calcium blockers were found to be the most clinical and cost-effective tocolytic medicine for women in suspected or diagnosed preterm labour with intact membranes
- oxytocin receptor blockers were also found effective for some other outcomes but not the most effective option overall
- Prostaglandin inhibitors may produce a protective effect for delaying birth by more than 48 hours
- There is limited data on the long term consequences of tocolytics for both babies and their mothers

12.1.12 Recommendations

39. Take the following factors into account when making a decision about whether to start tocolysis:

- whether the woman is in suspected or diagnosed preterm labour
- other clinical features (for example, bleeding or infection) which might suggest that stopping labour is contraindicated
- gestational age at presentation
- likely benefit of maternal corticosteroids (see Chapter 10)
- availability of neonatal care (need for transfer to another unit)
- the preference of the woman.

40. Offer calcium channel blockers for tocolysis[f] to women between 24^{+1} and 34^{+0} weeks of pregnancy who have intact membranes and are in suspected or diagnosed preterm labour.

41. If calcium channel blockers are contraindicated, offer oxytocin receptor antagonists for tocolysis.

f Although this use is common in UK clinical practice, at the time of consultation (June 2015), calcium channel blockers did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Prescribing guidance: prescribing unlicensed medicines for further information.

42. **Be aware that there is an absence of evidence about all tocolytic medicines before 26^{+0} weeks of pregnancy.**

43. **Do not offer betamimetics for tocolysis.**

13 Fetal monitoring

13.1 Introduction

Babies in the uterus derive oxygen from the mother via the placenta and umbilical cord. During contractions of the uterus in labour this oxygen exchange can be intermittently interrupted. Well babies at term during normal labour are not adversely affected by this. However, this is not always the case, and fetal hypoxia and then acidosis can occur. In theory, the preterm fetus is more vulnerable to this risk of intrapartum hypoxia-acidosis.

Surveillance for fetal hypoxia in labour is undertaken by fetal heart rate monitoring. The fetal heart rate can be monitored using either intermittent auscultation (listening in to the baby's heart using a hand-held device) or by a continuous electronic recording. Continuous electronic recording can be undertaken using either an external ultrasound transducer, positioned on the mother's abdomen to pick up the fetal heart rate, or a fetal scalp electrode (a small clip introduced through the mother's vagina and attached to the baby's head).

The outputs of both electronic methods are displayed as a cardiotocograph (CTG) trace. The tocograph is a simultaneous recording of the uterine contractions, so the CTG trace provides a visual continuous record of fetal heart rate and uterine contractions. There are features that can indicate the baby is well and responding normally to the events of labour (for example, slowing of the fetal heart rate during a contraction). There are other features that may indicate a serious emergency (for example, development of a persistent bradycardia following cord prolapse or placental abruption).

The search strategies for this chapter can be found in Appendix E, the excluded studies for this chapter can be found in Appendix G, the evidence tables for this chapter can be found in Appendix H, and the forest plots for this chapter can found in Appendix I.

13.2 Interpretation of the fetal heart rate (FHR) pattern

13.2.1 Introduction

The 4 features of the fetal heart rate that are scrutinised in the CTG trace at term are: the baseline heart rate, the baseline variability, the presence or absence of decelerations, and the presence of accelerations – all of which are used to indicate whether fetal hypoxia-acidosis is developing. The physiological changes of hypoxia-acidosis result in alterations in these parameters. Whether the recognised features of the fetal heart rate at term can be extrapolated to the preterm fetus is uncertain.

13.2.2 Review question

What are the criteria for best interpreting the preterm fetal heart rate trace at different gestational ages for unborn babies whose mothers are in suspected or diagnosed preterm labour?

13.2.3 Description of included studies

Twelve studies were included in this review (Althaus 2005; Aina-Mumuney 2007; Kariniemi 1984; Bowes 1980; Matsuda 2003; Holmes 2001; Martin 1974; Rayburn 1987; Burrus 1994; Braithwaite 1986; Douvas 1984; Nisenblat 2006).

Seven included studies are from the USA (Burrus 1994; Bowes 1980; Althaus 2005; Martin 1974; Aina-Mumuney 2007; Douvas 1984; Rayburn 1987); 2 studies from Canada (Braithwaite 1986; Holmes 2001), with the further 3 studies from Finland (Kariniemi 1984), Israel (Nisenblat 2006) and Japan (Matsuda 2003).

This review question was designed to test the predictive value of fetal heart rate features for neonatal adverse outcomes including, neonatal acidemia, intraventricular haemorrhage and neonatal death. A more inclusive approach was adopted and different types of observational studies (either retrospective or prospective cohort studies, case control studies or consecutive or non-consecutive case series) were included. Two main types of analyses were conducted:

- observational studies looking at the predictive value of features of fetal heart rate for neonatal adverse outcomes; tachycardia and bradycardia were assessed in 2 studies (Althaus 2005; Aina-Mumuney 2007), accelerations and decelerations was assessed in 5 studies (Kariniemi 1984; Matsuda 2003; Bowes 1980; Martin 1974; Holmes 2001), defined CTG classification systems was assessed in 6 studies (Rayburn 1987; Kariniem 1991; Douvas 1984; Nisenblat 2006; Burrus 1994; Braithwaite 1986)
- observational studies testing the association between fetal heart rate baseline variability and neonatal respiratory distress syndrome, neonatal death and/or metabolic acidosis in 4 studies (Kariniemi 1984; Althaus 2005; Bowes 1980; Aina-Mumuney 2007)

The mean gestational age of babies in 7 included studies ranged from 26 to 30 weeks (Althaus 2005; Kariniemi 1984; Nisenblat 2006; Rayburn 1987; Holmes 2001; Bowes 1980; Burrus 1994). Two studies included women giving birth at less than 36 weeks gestation (Aina-Mumuney 2007, Matsuda 2003). Two studies included women with gestational age of less than 35 and 30 weeks (Martin 1974; Braithwaite 1986) and a further 1 study included babies with birth weight < 1800 grams and did not report the specific gestational age (Douvas 1984).

The use of tocolytics was not reported in 8 studies. In 4 studies women received tocolysis but the proportion of women receiving these medicines was not reported.

13.2.4 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Data is reported in modified GRADE profiles below for the following CTG parameters:

- Fetal heart rate
  - Table 80: GRADE profile for predictive value of bradycardia and tachycardia for adverse neonatal outcomes
  - Table 81: GRADE profile for association between tachycardia and systemic fetal inflammation
- Baseline variability
  - Table 82: GRADE profile for predictive value of fetal heart rate baseline variability for neonatal adverse outcomes
  - Table 83: GRADE profile for association between fetal heart rate baseline variability and neonatal adverse outcomes or umbilical artery blood gas values
- Accelerations
  - Table 84: GRADE profile for predictive value of absence of fetal heart rate accelerations (non-reactive CTG) for adverse neonatal outcomes
- Decelerations
  - Table 85: GRADE profile for predictive value of fetal heart rate late, “prolonged” and “severe variable” decelerations for adverse neonatal outcomes

  - Table 86: GRADE profile for association between variable fetal heart rate decelerations and adverse neonatal outcome
- Categorization/classification of CTGs
  - Table 87: GRADE profile for predictive value of published categorisation of CTGs for adverse neonatal outcomes
  - Table 88: GRADE profile for association between categorisation of CTGs and adverse neonatal outcomes

The grading of evidence from prospective comparative observational studies or prospective consecutive case series started at high quality and was then downgraded if there were any issues identified that would undermine the trustworthiness of the findings. Evidence from retrospective comparative observational studies or retrospective consecutive case series started at moderate quality and was then downgraded if there were any issues. Evidence from non-consecutive case series started at low quality and was then downgraded if there were any issues.

The classifications of CTGs used and reported in the 6 studies (Rayburn 1987; Kariniem 1991; Douvas 1984; Nisenblat 2006; Burrus 1994; Braithwaite 1986) are detailed in the Evidence Tables of Appendix H.

The stage of labour was considered an important additional information for the interpretation of results and was included in the GRADE table along with the study's sample size. Although the most appropriate measures of assessing the predictive ability of the criteria for interpreting preterm fetal hearth rate are the positive and negative likelihood ratios, supplementary information were given for sensitivity and specificity. For the studies that tested associations of features of fetal heart trace with neonatal outcomes, odd ratios (with 95% confidence intervals) were considered as best measures of these associations.

Table 80: GRADE profile for predictive value of bradycardia and tachycardia for adverse neonatal outcomes

Quality assessment						Definition of outcome	Stage of labour	Total number of women & baby pairs	Measure of diagnostic accuracy (95% CI)a				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Tachycardia (> 160 bpm) (mean duration 35.2 min SD 22.8)													
1 study (Althaus 2005)	Case control	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	cerebral white matter injury a	1 hour before birth	246	17.24% (11.48 to 24.39)	78.5% (70.1 to 85.4)	0.80 (0.49 to 1.31) Not useful	1.05 (0.94 to 1.19) Not useful	Low
Bradycardia (< 110 bpm) (NICHD classification) (duration > 2 min)													
1 study (Althaus 2005)	Case control	serious[3]	no serious inconsistency	serious[1]	serious[2]	cerebral white matter injury a	1 hour before birth	246	4.8% (1.79 to 10.16)	95.8% (90.6 to 98.6)	1.16 (0.36 to 3.71) Not useful	0.99 (0.94 to 1.05) Not useful	Very low
Bradycardia (< 110 bpm) episodes[bc]													
1 study (Althaus 2005)	Case control	serious[1]	no serious inconsistency	serious[1]	serious[2]	cerebral white matter injury a	1 hour before birth	246	4.80% (1.79 to 10.16)	92.56% (86.3 to 96.5)	0.65 (0.24 to 1.76) Not useful	1.03 (0.96 to 1.10) Not useful	Very low

CI Confidence interval; bpm beats per minute
a Diagnosed by neonatal head ultrasound; first at 24 – 72 hours after birth, second at 10 -14 days of life and third at 6 weeks of life.
b The number of bradycardia episodes lasting > 2 min reported. There were 6 bradycardia episodes in cases and 9 in controls.
c Bradycardia mean nadir (bpm): cases 87.3 (SD 4.1), control: 83.3 (SD 23.4). Bradycardia mean duration (min): cases: 5.88 (SD 4.1), controls: 5.02 (SD 2.20)
1 n = 27 (29.8%) of cases and n = 10 (15.8%) of controls had multiple gestations
2 Wide CI (LRs)

Table 81: GRADE profile for association between tachycardia and systemic fetal inflammation

Quality assessment						Definition of outcome	Stage of labour	Number of babies with tachycardia > 160 bpm	OR (95% CI)	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision					
Tachycardia (> 160 bpm) (NICHD classification)										
1 study (Aina-Mumuney 2007)	Case control	no serious risk of bias	no serious inconsistency	serious1	no serious imprecision	systemic fetal inflammation[a]	2 hours before birth	150	OR 1.38 (0.30 to 6.42)	Moderate

bpm beats per minute; OR odds ratio; CI confidence interval; NICHD National Institute of Child Health and Human Development
a systemic fetal inflammation was diagnosed by histologically confirmed chorioamnionitis and funisitis
1 n = 3 (2.7%) of cases and n = 23 (30.7%) of controls had multiple gestations

Table 82: GRADE profile for predictive value of fetal heart rate baseline variability for neonatal adverse outcomes

Quality assessment						Definition of outcome	Stage of labour	Total number of women & baby pairs	Measure of diagnostic accuracy (95% CI)*				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Reduced variability ("silent" pattern: FHR variability < 5 bpm > 5 min)													
1 study (Kariniemi 1991)	Case control	serious[1,2,3]	no serious inconsistency	serious[4]	serious[5]	neonatal death	NR	74	.42.3% (23.4 to 63.0)	29.2% (16.9 to 44)	0.60 (0.37 to 0.97) Not useful	0.77 (0.14 to 3.43) Not useful	Very low
1 study (Kariniemi 1991)	Case control	serious1[,2,3]	no serious inconsistency	serious[4]	serious[5]	respiratory distress syndrome[a]	NR	74	50.0% (31.9 to 68.1)	18.9% (8.0 to 35.1)	0.73 (0.53 to 1.01) Not useful	2.15 (0.98 to 4.72) Not useful	Very low
Baseline variability < 5 bpm (NICHD classification)													
1 study (Althaus 2005)	Case control	no serious risk of bias	no serious inconsistency	serious[6]	serious[5]	cerebral white matter injury[b]	1 hour before birth	246	19.2% (12.7 to 27.2)	75.2% (66.5 to 82.6)	0.77 (0.48 to 1.25) Not useful	1.07 (0.94 to 1.23) Not useful	Low
Baseline variability < 5 bpm (duration 20 min)													
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[5]	neonatal death	1 hour before birth	61	10.0% (1.66 to 44.5)	82.3% (69.1 to 91.5)	6.57 (0.08 to 3.99) Moderately useful	1.09 (0.86 to 1.39) Not useful	Very low
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[5]	respiratory distress syndrome[c]	1 hour before birth	61	12.0% (2.69 to 31.25)	85.3% (86.9 to 94.9)	0.82 (0.21 to 3.10) Not useful	1.03 (0.84 to 1.26) Not useful	Very low
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	central nervous system haemorrhaged	1 hour before birth	61	0.00%	81.8% (69.1 to 90.9)	0.00 Not useful	1.22 (1.08 to 1.38) Not useful	Low
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[5]	umbilical cord pH < 7.20	1 hour before birth	61	50.0% (18.9 to 81.1)	92.3% (74.8 to 98.3)	6.50 (1.50 to 28.23) Moderately useful	0.54 (0.29 to 1.02) Not useful	Very low

CI confidence interval; FHR fetal heart rate; bpm beats per minute; NICHD National Institute of Child Health and Human Development; NR not reported
a Respiratory distress syndrome was defined as in the presence of tachypnoea, retraction and granting, hypoxaemia in room air and air bronchogram and reticulogranular pattern in X-ray when symptoms appears 6 hours after birth and lasted 24 hours.
b The diagnosis of white matter injury was made by neonatal head ultrasonograms at 24- 72 hours after birth, 10 -14 days of life and at 6 weeks.
c Respiratory distress syndrome (RDS) was diagnosed if the all following were present: arterial Po2 was < 50 mm Hg in room air, increased ambient oxygen, continuous positive airway pressure or ventilation required > 24 hours to support respiration, chest x-ray evidence, no evidence of other disease caused RDS
d central nervous system CNS haemorrhage was diagnosed in babies who exhibited: seizures, fullness of anterior fontanelle, a decreased in the haematocrit, and blood in the cerebral spinal fluid
1 The traces were evaluated by only 1 of the study's authors
2 No clear inclusion/exclusion criteria, hence high risk of selection bias

3 No clear definition of FHR pattern. Unclear in what stage of labour the traces were obtained and evaluated
4 Most babies delivered by caesarean section before labour started
5 Confidence interval crossed one default MIDs (LRs)
6 n = 27 (29.8%) of cases and n = 10 (15.8%) of controls had multiple gestations

Table 83: GRADE profile for association between fetal heart rate baseline variability and neonatal adverse outcomes or umbilical artery blood gas values

Quality assessment						Definition of outcome	Stage of labour	Number of babies with defined CTG pattern	OR (95% CI) or mean (SD)	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision					
Decreased short term variability (<5 bpm NICHD classification)										
1 study (Aina-Mumuney 2007)	Cohort	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	systemic fetal inflammation	1st stage	150	OR 0.71 (0.34 to 1.50)	Moderate
Reduced reactivity (NICHD classification)										
1 study (Aina-Mumuney 2007)	Cohort	no serious risk of bias	no serious inconsistency	serious[1]	no serious imprecision	systemic fetal inflammation	1st stage	150	OR 0.96 (0.49 to 1.87)	Moderate
Increase reactivity (NICHD classification)										
1 study (Althaus 2005)	Case control	no serious risk of bias	no serious inconsistency	serious[2]	no serious imprecision	Umbilical cord pH and base excess	1 hour before birth	246	Non-reactive: pH: 7.29 ± 0.10 Base excess: -2.7 (3.8) Reactive: pH: 7.31 ± 0.08 Base excess: -2.9 (3.4) Both p = NS	Moderate

FHR fetal heart rate; BPM beats per minute; NICHD National Institute of Child Health and Human Development; BD base deficit; OR odds ratio; CI confidence interval; p = NS no significant difference
1 n = 3 (2.7%) of cases and n = 23 (30.7%) of controls had multiple gestations
2 17% of cases and 6% of controls had multiple gestations

Table 84: GRADE profile for predictive value of absence of fetal heart rate accelerations (non-reactive CTG) for adverse neonatal outcomes

Quality assessment						Definition of outcome	Stage of labour	Total number of women & baby pairs	Measure of diagnostic accuracy (95% CI)a				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Non-reactive CTG (< 2 accelerations >15bpm in 30 min)													
1 study (Kariniem 1984)	Case control	serious[1,2,3]	no serious inconsistency	serious[4]	no serious imprecision	neonatal death	NR	74	50.0% (29.9 to 70.0)	14.6% (6.10 to 27.7)	0.59 (0.39 to 0.87) Not useful	0.77 (1.56 to 7.52) Not useful	Very low
1 study (Kariniem 1984)	Case control	serious[1,2,3]	no serious inconsistency	serious[4]	no serious imprecision	respiratory distress syndrome	NR	74	68.7% (49.9 to 83.8)	24.3% (11.8 to 41.2)	0.91 (0.68 to 1.22) Not useful	1.28 (0.60 to 2.76) Not useful	Very low

CI confidence interval; NR not reported
a Respiratory distress syndrome was defined as the presence of tachypnoea, retraction and grunting, hypoxaemia in room air and air bronchogram and reticulogranular pattern in X-ray when symptoms appeared 6 hours after birth and lasted 24 hours.
1 The traces were evaluated by only 1 of the study's authors
2 No clear inclusion/exclusion criteria, hence high risk of selection bias
3 No clear definition of FHR pattern. Unclear in what stage of labour the traces were obtained and evaluated
4 Most babies delivered by caesarean section before labour started
5 Confidence interval crossed 1 default MIDs (LRs)

Table 85: GRADE profile for predictive value of fetal heart rate late, "prolonged" and "severe variable" decelerations for adverse neonatal outcomes

Quality assessment						Definition of outcome	Stage of labour	Number of women & baby pairs	Measure of diagnostic accuracy (95% CI)a				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
Late decelerations (not defined)													
1 study (Kariniem 1984)	Case control	serious[1,2,3]	no serious inconsistency	serious[4]	serious[5]	neonatal death	NR	74	53.8% (33.4 to 73.4)	16.67% (7.50 to 30.2)	0.65 (0.44 to 0.94) Not useful	0.77 (0.30 to 5.60) Not useful	Very low
1 study (Kariniem 1984)	Case control	serious[1,2,3]	no serious inconsistency	serious[4]	serious[5]	respiratory distress syndrome[a]	NR	74	59.3% (40.6 to 76.2)	18.9% (8.0 to 35.1)	0.73 (0.53 to 1.01) Not useful	2.15 (0.98 to 4.72) Not useful	Very low
"Combined distress" pattern (decelerations and "silent" pattern in 30 min)													
1 study (Kariniem 1984)	Case control	serious[1,2,3]	no serious inconsistency	serious[4]	no serious imprecision	neonatal death	NR	74	19.2% (6.63 to 39.4)	35.4% (22.2 to 50.4)	0.30 (0.13 to 0.67) Not useful	2.28 (1.49 to 3.49) Not useful	Very low
1 study (Kariniem 1984)	Case control	serious[1,2,3]	no serious inconsistency	serious[4]	serious[5]	respiratory distress syndrome[b]	NR	74	37.5% (21.1 to 56.3)	40.5% (24.7 to 57.9)	0.63 (0.37 to 1.06) Not useful	2.54 (0.96 to 2.48) Not useful	Very low
Late decelerations with loss of variability for < 30 min													

Quality assessment						Definition of outcome	Stage of labour	Number of women & baby pairs	Measure of diagnostic accuracy (95% CI)a				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
1 study (Matsuda 2003)	Cohort	serious[2,6,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	umbilical artery pH < 7.1	2 hours before birth	772	28.6%[b]	86.4% [b]	2.10 b Not useful	0.82 [b] Not useful	Low
Late decelerations with loss of variability for < 60 min													
1 study (Matsuda 2003)	Cohort	serious[2,6,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	umbilical artery pH < 7.1	2 hours before birth	772	85.7% [b]	68.2% [b]	2.69 [b] Not useful	0.20 [b] Moderatel y useful	Low
Late decelerations with loss of variability for < 90 min													
1 study (Matsuda 2003)	Cohort	serious[2,6,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	umbilical artery pH < 7.1	2 hours before birth	772	100% [b]	45.5% [b]	1.83 [b] Not useful	0.0 [b] very useful	Low
"Prolonged" decelerations with loss of variability < 30 min (prolonged decelerations not defined)													
1 study (Matsuda 2003)	Cohort	serious[2,6,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	umbilical artery pH < 7.1	2 hours before birth	772	81.8% [b]	56.8% [b]	1.9 [b] Not useful	0.32 [b] Moderatel y useful	Low
"Prolonged" decelerations with loss of variability < 60 min (prolonged decelerations not defined)													
1 study (Matsuda 2003)	Cohort	serious[2,6,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	umbilical artery pH < 7.1	2 hours before birth	772	90.9% [b]	37.8% [b]	1.49 [b] Not useful	0.24 [b] Moderatel y useful	Low
"Prolonged" decelerations with loss of variability < 90 min (prolonged decelerations not defined)													
1 study (Matsuda 2003)	Cohort	serious[2,6,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	umbilical artery pH < 7.1	2 hours before birth	772	100% [b]	16.2% [b]	1.19 [b] Not useful	0.0 [b] Very useful	Low
"Severe variable late" decelerations (ominous periodic changes not defined)													
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	Neonatal death	1 hour before birth	61	12.0 % (2.69 to 31.2)	86.11% (70.04 to 95.2)	0.86 (0.23 to 3.29) Not useful	1.02 (0.84to 1.24) Not useful	Low
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[5]	central nervous system haemorrhage [c]	1 hour before birth	61	16.7% (2.76 to 63.90)	12.73% (5.30 to 24.5)	0.19 (0.03 to 1.15) Not useful	6.55 (3.00 to 14.27) Not useful	Very low
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[5]	respiratory distress syndrome[d]	1 hour before birth	61	0.00	84.3% (71.4 to 93)	0.00 Not useful	1.19 (1.05 to 1.34) Not useful	Very low

Quality assessment								Number of women & baby pairs	Measure of diagnostic accuracy (95% CI)a				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Definition of outcome	Stage of labour		Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
1 study (Bowes 1980)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[5]	umbilical cord pH < 7.20	1 hour before birth	61	60.0% (26.3 to 87.7)	100% (86.6 to 98.3)	0.00 Not useful	0.40 (0.19 to 0.85) Moderately useful	Very low

CI confidence interval; NC not calculable; NR not reported; NICHD National Institute of Child Health and Human Development
a Respiratory distress syndrome was defined as in the presence of tachypnoea, retraction and granting, hypoxaemia in room air and air bronchogram and reticulogranular pattern in X-ray when symptoms appears 6 hours after birth and lasted 24 hours.
cb Insufficient data reported to calculate CI
c central nervous system (CNS) haemorrhage was diagnosed in babies who exhibited: seizures, fullness of anterior fontanelle, a decreased in the haematocrit, and blood in the cerebral spinal fluid
d Respiratory distress syndrome (RDS) was diagnosed if the all following were present: arterial Po2 was < 50 mm Hg in room air, increased ambient oxygen, continuous positive airway pressure or ventilation required > 24 hours to support respiration, chest x-ray evidence, no evidence of other disease caused RDS
1 The traces were evaluated by only 1 of the study's authors
2 No clear inclusion/exclusion criteria, hence high risk of selection bias
3 No clear definition of CTG pattern. Unclear in what stage of labour the traces were obtained and evaluated
4 Most babies delivered by caesarean section before labour start
5 Confidence interval crossed 1 default MIDs (LRs)
6 Unclear how and by whom data were analysed
7 Unclear if the assessors were blinded to outcomes
8 Women's characteristics not reported

Table 86: GRADE profile for association between variable fetal heart rate decelerations and adverse neonatal outcome

Quality assessment						Definition of outcome	Stage of labour	Number of babies with defined CTG patterns (gestation)	number of babies with defined outcome or mean outcome value or p value	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision					
Variable decelerations[a]										
1 study (Holmes 2001)	Case control	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	neonatal death	1 hour before birth	82	Cases:2/41 Controls: 0/41 p = NS	Moderate
1 study (Holmes 2001)	Case control	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	umbilical cord artery pH < 7.1	1 hour before birth	79	Cases:0/38 Controls: 2/41 p = NS	Moderate
1 study (Holmes 2001)	Case control	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	resuscitation (cardiac massage or drug therapy)	1 hour before birth	82	Cases:1/41 Controls: 2/41 p = NS	Moderate
1 study (Holmes 2001)	Case control	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	intraventricular haemorrhage grade III or IV	1 hour before birth	82	Cases:4/41 Controls: 0/41 p = 0.04	Moderate

Quality assessment						Definition of outcome	Stage of labour	Number of babies with defined CTG patterns (gestation)	number of babies with defined outcome or mean outcome value or p value	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision					
1 study (Holmes 2001)	Case control	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	periventricular leukomalacia	1 hour before birth	82	Cases:1/41 Controls: 0/41 p = NS	Moderate
"Severe variable" decelerations with late component										
1 study (Martin 1974)	Case series	serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	neonatal death	1st stage	73	Severe variable deceleration with late component: 10/11 Mild/moderate variable decelerations without late component: 1/11 p = 0.05	Moderate

FHR fetal heart rate; NR not reported; NICU neonatal intensive care unit; NICHD National Institute of Child Health and Human Development; BD base deficit; OR odds ratio; CI confidence interval; p=NS no significant difference
a Cases consisted of traces with ≥ 3 variable decelerations in the hour prior to birth and controls consisted of traces with ≤ 2 variable decelerations. Variable deceleration defined as an abrupt decrease in FHR of at least 15 bpm lasting for between 15 seconds and 2 minutes according to the National Institutes of Child Health and Human Development (NICHD)
1 Unclear if the assessors were blinded to the outcomes
2 No clear exclusion criteria, hence high risk of selection bias
3 Women's characteristics not reported

Table 87: GRADE profile for predictive value of published categorisation of CTGs for adverse neonatal outcomes

Quality assessment						Definition of outcome	Stage of labour	Total number of women & baby pairs	Measure of diagnostic accuracy (95% CI)[a]				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
"Reassuring" CTG (normal pattern with/without occasional mild/moderate variable decelerations)													
1 study (Rayburn 1987)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[3]	intraventricular haemorrhage	minimum of 20 min of tracing during the first stage of labour	72	55.2% (38.3 to 71.3)	47.3% (31 to 61.1)	1.05 Not useful (0.69 to 1.59) Not useful	0.94 Not useful (0.58 to 1.54) Not useful	Very low
Fischer classification (abnormal vs. normal)[a]													
1 study (Braithwaite 1986)	Case control	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	neonatal death	last 30 min of tracing during the first	383	86.5% (69.8 to 77.7)	48.39% (30.17 to 66.9)	1.71 (0.19 to 2.48) Not useful	0.24 (0.08 to 0.73) Not useful	Moderate

Quality assessment						Definition of outcome	Stage of labour	Total number of women & baby pairs	Measure of diagnostic accuracy (95% CI)[a]				Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	
							stage of labour						
"Abnormal" CTG (not defined)													
1 study (Kariniemi 1991)	Case control	serious[5,6,7]	no serious inconsistency	serious[8]	serious[3]	neonatal death	NR	74	80.7% (60.6 to 93.3)	8.33% (2.37 to 20.2)	0.88 (0.72 to 1.08) Not useful	2.31 (0.68 to 7.86) Not useful	Very low
1 study (Kariniemi 1991)	Case control	serious[5,6,7]	no serious inconsistency	serious[8]	serious[3]	respiratory distress syndrome	NR	74	81.2% (63.5 to 92.7)	8.11% (1.80 to 21.9)	0.88 (0.73 to 1.07) Not useful	2.31 (0.63 to 8.51) Not useful	Very low
"Abnormal" CTG[b]													
1 study (Douvas 1984)	Case series	serious[9,10]	no serious inconsistency	no serious indirectness	serious[3]	birth asphyxia[d]	NR	89	72.7% (54.4 to 86.7)	94.6% (85.1 to 98.9)	13.5 (4.43 to 41.6) Very useful	0.29 (0.16 to 0.50) Moderately useful	Very low
1 study (Douvas 1984)	Case series	serious[9,10]	no serious inconsistency	no serious indirectness	serious[3]	respiratory distress syndrome[c]	NR	89	66.7% (47.2 to 82.7)	88.1% (77 to 95)	5.62 (2.68 to 11.78) Moderately useful	0.38 (0.23 to 0.63) Moderately useful	Very low
"Pathological" CTG[d]													
1 study (Nisenblat 2006)	Cohort	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	neurodevelopmental abnormality	1 hour before birth	111	27%	74%	1.03 Not useful	0.98 Not useful	Low
"Suspicious" CTG[e]													
1 study (Rayburn 1987)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[3]	intraventricular haemorrhage	Minimum of 20 min of tracing during the first stage of labour	72	29.1% (12.6 to 51)	69.2% (48.2 to 85.6)	0.95 (0.41 to 2.22) Not useful	1.02 (0.71 to 1.47) Not useful	Very low
"Ominous" CTG[f]													
1 study (Rayburn 1987)	Case control	serious[1,2]	no serious inconsistency	no serious indirectness	serious[3]	intraventricular haemorrhage	Minimum of 20 min of tracing during the first stage of labour	72	45.1% (27.3 to 63.9)	66.67% (40.6 to 77.3)	1.13 (0.63 to 2.03) Not useful	0.91 (0.59 to 1.41) Not useful	Very low

CI confidence interval; FIGO International Federation of Obstetrics and Gynaecology; CTG cardiotocography; NR not reported; NICHD National Institute of Child Health and Human Development; NICU neonatal intensive care unit; FHR fetal heart rate; BE base excess;

a Normal CTG in 30 min period was defined as: baseline 120 – 160 bpm, variability > 6 bpm, accelerations present, and no decelerations (extracted from Fischer 1976, translated from German)
b CTG considered as abnormal if any of the following were seen late decelerations defined as persistent decelerations following 50% of the contractions over a 30 minute perio, severe variable decelerations defined as decelerations < 70 bpm asting for > 60 second, absent or minimal beat to beat variability, defined as < 5 bpm over a 30 minute perio, prolonged bradycardia defined as FHR < 100 bpm persistently over a period of > 3 minutes
c The measure of asphyxia was based on the 1 of the following: Apgar score < 3 at 1 minute or < 6 at 5 minute, immediate resuscitation requiring positive pressure oxygen for > 1 minute, pH < 7.25 on arrival in the neonatal intensive care unit
d "Pathological" CTG was defined as: baseline fetal heart rate >160 bpm or < 110 bpm, absence of FHR variability (amplitude range undetectable), either recurrent late decelerations (deceleration is associated with the uterine contraction, with nadir of the deceleration occurring after peak of the contraction) or recurrent severe variable decelerations (decrease in FHR below 70 bpm lasting longer than 60 seconds or other decelerations with slow return to baseline, associated with the uterine contractions, the onset, depth, and duration vary with successive uterine contractions) classified as mild, moderate, or severe on the basis of umbilical artery base deficit (cut off >12 mmol/l) and neonatal encephalopathy and other organ system complications
e "Suspicious" CTG: intermittent late decelerations, decreased variability or tachycardia present
f "Ominous" CTG: consistent with repetitive severe variable or late decelerations or repetitive prolonged decelerations (>2 min)
"Suspicious" or "ominous" patterns that were continuous and repetitive for > 30 min were considered indicative of "fetal distress"
1 No clear exclusion criteria, hence high risk of selection bias
2 No clear definition of CTG features
3 Confidence interval crossed 1 default MIDs (LRs)
5 The CTGs were evaluated by only 1 of the study's authors
6 No clear inclusion/exclusion criteria, hence high risk of selection bias
7 No clear definition of CTG pattern. Unclear in what stage of labour the traces were obtained and evaluated
8 Most babies delivered by caesarean section before labour start
9 No clear inclusion and exclusion criteria, hence high risk of selection bias
10 Women's characteristics not reported

Table 88: GRADE profile for association between categorisation of CTGs and adverse neonatal outcomes

Quality assessment						Definition of outcome	Stage of labour	Number of woman and baby pairs	Number of babies with defined outcome and p value	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision					
"Normal" versus "abnormal" CTG[b]										
1 study (Burrus 1994)	Case series	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	neonatal death	1 hour before birth	41	n = 13/41 normal pattern n = 3/22 abnormal pattern n = 10/19 p = 0.007	Low
1 study (Burrus 1994)	Case series	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	intraventricular haemorrhage	1 hour before birth	41	n = 8/41 normal pattern n = 2/22 abnormal pattern n = 6/19 p = NS	Low
1 study (Burrus 1994)	Case series	serious1[,2]	no serious inconsistency	no serious indirectness	no serious imprecision	> 42 days on assisted ventilation	1 hour before birth	41	n = 4/41 normal pattern n = 2/22 abnormal pattern n = 2/19 p = NS	Low
1 study (Burrus 1994)	Case series	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	> 90 days of hospitalisation	1 hour before birth	41	n = 6/41 normal pattern n = 5/22 abnormal pattern n = 1/19 p = NS	Low
1 study (Burrus 1994)	Case series	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	cerebral palsy at 1 year	1 hour before birth	41	n = 3/41 normal pattern n = 1/22 abnormal pattern n = 2/19 p = NS	Low

FHR fetal heart rate; NICHD National Institute of Child Health and Human Development; BD base deficit; OR odds ratio; CI confidence interval; NR not reported; SD standard deviation; p = NS no significant difference

a "Ominous" CTG defined as repetitive pattern of late deceleration and pronounced variable decelerations (> 40 seconds duration and/or > 60 beats loss)

b "Normal" and "abnormal" CTG defined by Kubli 1969 as: normal baseline (FHR 120 – 160 bpm), bradycardia (FHR 100 – 120 bpm) and severe bradycardia (FHR < 100 bpm) Variability was defined as normal variability (amplitude range > 5 bpm), moderately reduced variability (2 – 5 bpm), severely reduced variability (< 2 bpm), a salutatory or hypervariable pattern was diagnosed if amplitude range exceeded 25 bpm Decelerations defined as mild variable deceleration (lasted <30 sec irrespective of level, if the nadir was >80 bpm irrespective of duration, or if their nadir was 70 -80 bpm if lasting <60 sec), moderate variable deceleration (lasted 30 to 60 sec with the nadir < 60 bpm, or lasted > 60 sec but with a nadir between 70 -80 bpm), severe variable deceleration (lasted > 60 sec with a nadir. < 70 bpm, occasional (2 or fewer in a 10 min window) or frequent (3 or more)

1 Small study with low statistical power

2 Poor level of agreement between 2 CTG evaluators

13.2.5 Evidence statements

Fetal heart rate: bradycardia, tachycardia

Evidence from 2 observational studies with approximately 400 participants found that fetal baseline tachycardia or bradycardia had a not useful positive and negative likelihood ratios for predicting neonatal cerebral white matter injury at birth. There was no evidence of a difference in the risk of systemic fetal inflammation between babies with intrapartum tachycardia compared to those without intrapartum tachycardia. The evidence was of moderate to very low quality.

Baseline variability

Evidence from 3 observational studies with over 400 participants indicated that reduced baseline variability had low to moderate positive likelihood ratios for predicting poor neonatal outcomes. There was no evidence of a difference in the risk of systemic fetal inflammation between babies with reduced baseline variability compared to those without reduced baseline variability. The evidence was of moderate quality.

Absence/presence of accelerations (reactivity)

Very low quality evidence from 1 observational study with 74 participants showed no difference in the risk of systemic fetal inflammation or poor umbilical cord blood gases for babies with non-reactive traces compared to those with reactive traces.

The diagnostic value of absence of accelerations was low for respiratory distress syndrome and neonatal death (across all diagnostic parameters).

Late, “prolonged” and “severe variable late” decelerations

Evidence from 6 observational studies (number of participants ranged from 61 to 772) for all diagnostic values of late decelerations for poor neonatal outcomes showed that alone they were not useful for predicting poor neonatal outcome (very low quality evidence). Late decelerations with loss of variability lasting < 30 min, < 60 min and < 90 min had not useful positive and negative likelihood ratios (low quality evidence). “Prolonged” decelerations with loss of variability lasting < 30 min, < 60 min and < 90 min had generally not useful positive and negative likelihood ratios. ”Severe variable late” decelerations had also not useful positive and negative likelihood ratios for predicting poor neonatal outcomes although the study was very old with no clinical relevance. The evidence was of low and very low quality.

Variable decelerations

Evidence from 2 observational studies with over 150 participants showed that variable decelerations (defined as ‘not-severe’) had no significant association with neonatal death, umbilical cord artery pH < 7.1, or the need for neonatal resuscitation.

There was some evidence of an association between variable decelerations and grade III or IV intraventricular haemorrhage. There was some evidence that variable decelerations with a depth of 30–60 bpm had a higher degree of association with fetal acidosis when compared to variable decelerations with a depth of less than 30 bpm. However this association was only seen for babies under 34 weeks gestation and the numbers of cases in the studies was very small. There was some evidence that variable decelerations classified as severe with a late component had a high degree of association with neonatal death whilst mild/moderate variable decelerations without a late component had a low degree of association with neonatal death, though, again, the numbers of cases in the study was very small. The evidence was of low quality.

Categorisation/classification of fetal heart rate traces

Evidence from 5 observational studies (number of participants ranged from 72 to 383) for the diagnostic values of a range of different categorisations of CTGs were conflicting, ranging from high to low across all diagnostic parameters. Studies describing CTGs as “reassuring”, “pathological”, “suspicious” and “ominous” all found low and not useful diagnostic accuracy across all diagnostic parameters for adverse neonatal outcomes. Two studies describing the CTG as “abnormal” found moderate to low sensitivity and high to low specificity, high to low positive likelihood ratios and low negative likelihood ratios for adverse neonatal outcomes. The evidence was predominately of low and very low quality.

There was some evidence of a high association between CTGs categorised as abnormal (however defined) and neonatal death but there was no association with other poor neonatal outcomes. The evidence was of low quality.

13.2.6 Health economics profile

No search for health economic evidence was undertaken for this question as it was thought to be concerned with how to interpret a trace rather than making decisions between alternative courses of actions.

Therefore this question was not identified as a priority for health economic analysis.

13.2.7 Evidence to recommendations

13.2.7.1 Relative value placed on the outcomes considered

The Committee prioritised neonatal death and fetal acidosis as the most important outcomes for this question. The Committee felt it was important to assess how effective the cardiotocograph is at identifying babies with fetal hypoxia that may lead to acidosis and other adverse outcomes both in terms of identifying true positives and ruling out false negatives.

13.2.7.2 Consideration of clinical benefits and harms

The Committee discussed that the physiological control of fetal heart rate differs in preterm and term fetuses. The additional impact of immaturity of control of fetal heart rate in preterm fetuses makes the CTG interpretation different from term. They noted that some characteristics and patterns of the FHR are dependent on gestational age as they reflect the development and maturity of the central nervous system as well as the cardiovascular system. In the term fetus certain FHR features may be pathological but in the preterm setting they could be physiological. The Committee noted that the mean gestational age of women in most of the included studies ranged from 26 to 30 weeks. It was disappointing that no subgroup analysis was performed in the included studies for the significance of fetal heart rate patterns at different gestations.

The Committee found the definition of a normal CTG for preterm fetuses challenging as the evidence was limited. Although the category of abnormal CTG seems to be the best predictor of birth asphyxia, this result relates to the least useful outcome in the results because its incidence is very high in preterm babies.

The Committee noted that in very premature labour (24 to 26 weeks) there is a high risk of neonatal morbidity and mortality, and survival is dependant more on fetal weight and maturity rather than intrapartum hypoxia and mode of delivery.

It was also further recognised that it is well established that the baseline fetal heart rate in preterm fetuses is at the higher end of the normal range for the term fetus for physiological reasons, but that this reverts to the range more consistent with term fetuses as gestation advances. However, they felt that any rate more than 160bpm should be defined as

tachycardia across all preterm gestational ages. The Committee were concerned that women in preterm labour were at increased risk of infection (e.g. chorioamnionitis) and that this may present as a persistent fetal tachycardia, giving rise to a risk of misinterpretation of the CTG.

The Committee discussed that the baseline variability may be reduced at preterm gestations for physiological reasons. However, at term, fetal heart rate variability is an important clinical indicator of fetal acid base balance and oxygenation of the autonomic nerve centres within the brain. In the term setting, sustained absent variability is predictive of cerebral asphyxia. In this preterm review, there was some evidence of this in 1 study with moderate specificity and moderate useful likelihood ratio for adverse neonatal outcomes. However, given a possible physiological explanation for a higher baseline and reduced variability, these features should not be considered alone as indications for operative interventions in the preterm setting.

The Committee discussed that accelerations in the fetal heart rate of very preterm babies may not be present or their height and frequency may be significantly reduced. The Committee discussed that fetal heart rate decelerations are common and normal at very early preterm gestations (26 weeks and less) reflecting immature development of cardioregulatory mechanisms. They discussed that the presence of shallow or short-lived decelerations in very preterm babies should not be considered necessarily as an indicative of hypoxia when all other CTG features are reassuring. From 27 weeks onwards the frequency and height of accelerations increases and decelerations are normally not physiological. Importantly anticipated survival following preterm delivery also improves.

The Committee believed that although electronic fetal monitoring guidelines for term fetuses (see the NICE guideline on intrapartum care) cannot be always applied during labour to preterm fetuses, they can be considered as relevant after 32 weeks, as physiological maturity of the cardiovascular and neurological systems from this gestational age is comparable to that of term fetuses. Thus, from 32 weeks, baseline fetal heart rate and variability should be similar to that in term fetuses and accelerations with an amplitude of more than 15 beats from the baseline should be present as an indicator of fetal well-being. Decelerations can be interpreted as for the term fetus. The Committee discussed that theoretically, compared to term fetuses, preterm fetuses tend to have lower reserves and may deteriorate more quickly than term fetuses. Thus earlier and/or more prompt intervention may be required compared to term fetuses.

The Committee believed that a normal CTG is reassuring that the fetus is in good condition. An abnormal CTG does not always indicate that the outcome for the fetus will be poor. There is considerable variation between individuals in what is considered normal and abnormal CTG. The Committee were aware that there are risks to the mother and fetus if an abnormal CTG is used as the sole indication for intervention.

The Committee commented that clinical staff should not focus only on the CTG when caring for the woman in preterm labour, but should take the full clinical picture into account.

The Committee noted that the evidence from this review showed that the use of CTG is only moderately useful at best in predicting poor fetal/neonatal outcomes, with the majority of studies showing it to be not useful (not useful positive likelihood ratios).

The data showed that only a few CTG features have some limited evidence supporting their usefulness in predicting fetal outcome:

- fetal heart rate (bradycardia, tachycardia) did not seem to be of value.(see Table 80 and Table 81)
- abnormal baseline variability) of < 5bpm for 20 min in 1 hour before birth was moderately useful positive likelihood ratio in 1 study for identifying cord pH < 7.20 and neonatal death. (see Table 82 and Table 83)
- accelerations did not seem to be of value (Table 84)
- decelerations

- late, “prolonged” and “severe variable late” decelerations seemed to predict low cord pH values and absence of this feature was associated with normal pH. (see Table 85)
- variable decelerations had a positive association with intraventricular haemorrhage however the evidence was from a single very old study.(see Table 86)

Different categorizations/classifications of CTGs were associated with conflicting evidence of their value in predicting adverse outcomes or reassurance (see Table 87 and Table 88).The Committee felt that individual parameters in CTG could not be viewed and interpreted alone. The available evidence does not support the assumption that the CTG tracing can be viewed precisely. The evidence presented in the studies takes no account of the gestational age and the degree of prematurity and the potential for physiological changes of the fetal heart rate to be considered as pathological. It is for these reasons that the Committee felt that the classification should be less complex and less rigid. They felt that there is a need to consider the CTG as part of a bigger picture and CTG alone should not be the basis for intervention decisions. The Committee emphasised that the potential for harm arising from a false positive result in preterm labour is higher than in labour at term.

The Committee agreed that women should be fully consulted before performing continuous fetal heart rate monitoring.

The Committee commented that clinical staff should not focus only on the CTG when caring for the woman in preterm labour, but should take the full clinical picture into account.

13.2.7.3 Consideration of health benefits and resource uses

As this question looked at the diagnostic accuracy of different features of fetal heart rate traces, there were no resource use issues to consider.

13.2.7.4 Quality of evidence

The quality of the evidence reviewed varied from moderate to very low. The Committee noted many limitations of the research findings.

The Committee noted the small sample size in most studies; much larger numbers of cases would be needed to show a significant effect, particularly in terms of long term neurodevelopment.

Clinicians were not, and could not be blinded in these observational studies, so would have offered treatment based on the cardiotocograph. This ‘treatment effect’ is a serious problem: it would have undermined the validity of the estimates of diagnostic accuracy, an abnormal CTG is likely to prompt steps to intervene to delivery to improve neonatal outcome. Therefore when a case of adverse outcome was avoided by intervention before any harm occurred, it may well have erroneously been counted as a ‘false positive’.

The Committee also noted that fetal hypoxia was not likely to be the only or the principal risk for many preterm babies in labour. The interaction of these factors and their effect on the CTG could be complex.

Much of the evidence reviewed came from old studies. Clinical practice has changed to such an extent that the findings might not be relevant to current clinical practice. In particular, in 1 study reported in 1978 the assessment of ‘central nervous system haemorrhage’ was not robust as no cranial ultrasound was performed.

13.2.7.5 Other considerations

The Committee considered that the interpretation of CTG might be different between health care professionals, therefore care should be taken when interpreting CTGs so that appropriate action is taken where there are concerning signs.

The Committee discussed that it would be more appropriate to establish principles rather than precise parameters by which to assess the CTG.

13.2.8 Recommendations

The recommendations on interpreting fetal heart rate are in Section 13.5.5

13.3 Monitoring options: cardiotocography and intermittent auscultation

13.3.1 Introduction

Intermittent auscultation involves the healthcare professional listening to the fetal heart rate at regular intervals. At term it is recommended that intermittent auscultation is undertaken every 15 minutes in the first stage of labour and every 5 minutes in the second stage of labour and for 1 minute after a contraction.

Continuous electronic recording can be undertaken using either an external ultrasound transducer, positioned on the mother's abdomen to pick up the fetal heart rate, or a fetal scalp electrode (a small clip introduced through the mother's vagina and attached to the baby's head). The outputs of both electronic methods are displayed as a cardiotocograph (CTG) trace. The tocograph is a simultaneous recording of the uterine contractions, so the CTG trace provides a visual continuous record of fetal heart rate and uterine contractions.

This chapter will review the evidence comparing the effectiveness of these 2 monitoring techniques to improve outcomes for babies born preterm.

13.3.2 Review question

What is the effectiveness of electronic fetal monitoring compared with intermittent auscultation at different gestational ages for unborn babies whose mothers are in suspected or diagnosed preterm labour?

13.3.3 Description of included studies

As this question was set out to assess the comparative effectiveness of 2 interventions, RCTs were selected as the best study design to answer this review question. Comparative observational studies were considered when there were limited RCTs data. Two studies were included in this review (Luthy 1987; Shy 1988). Both included studies were from the USA. One included study was a multicentre RCT with 246 participants (Luthy 1987) and the other was a retrospective cohort study with 304 participants carried out in 14 hospitals that provided obstetric care (Shy 1988)

Both included studies examined the association between use of electronic fetal heart rate monitoring (EFM) and intermittent fetal heart rate auscultation with neonatal outcomes in women with suspected or diagnosed preterm labour. The term cardiotocography (CTG) is used throughout this review since this more accurately describes the monitoring carried out during labour which simultaneously records the fetal heart rate and uterine contractions.

The mean pregnancy gestation of babies in 1 included study (Luthy 1987) was 26 weeks in both study arms. No major differences were observed in this study between the CTG group and intermittent auscultation group for maternal age, marital status, race, postnatal care and birthweight. The other study (Shy 1988) included babies with birthweight from 700 to 1500 grams and did not analyse data based on the gestational age. Although women's characteristics were not reported in this study, the analysis was adjusted for birthweight, community hospital birth, rupture of membranes, and non-cephalic presentation. The use of

tocolytics was reported in both studies. In 1 study (Shy 1988) over 50% of women in both intermittent auscultation and electronic fetal monitoring groups were exposed to tocolytics, while in other study tocolytics were only given to women with intact membranes, although the number of women exposed to tocolytics is not reported (Luthy 1987). Fetal distress was the most common indication for caesarean section in 1 study (Luthy 1987). Caesarean section for fetal distress was performed for 8.2% of women with electronic fetal heart monitoring compared with 5.6% of women with intermittent auscultation (Luthy 1987).

13.3.4 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Data is reported in 2 GRADE profiles.

- Table 89: GRADE profile for comparison of cardiotocography versus intermittent auscultation – RCTs
- Table 90: GRADE profile for comparison of cardiotocography versus intermittent auscultation – Observational studies

The grading of evidence from the RCT was started at high quality and then downgraded if there were any issues identified that would undermine the trustworthiness of the findings. Evidence from the retrospective comparative observational study started at low quality and was then downgraded if there were any issues identified.

Although the protocol was set up to investigate the effect of different gestational ages on outcomes as subgroup analyses, 1 of the included studies looked at the role of babies birth weight and results are presented for this factor in a subgroup analysis as a surrogate of gestational age.

1 **Table 89: GRADE profile for comparison of cardiotocography versus intermittent auscultation – RCTs**

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cardiotocography	Intermittent auscultation	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal mortality											
Perinatal mortality - birthweight 500 – 1500											
1 study (Luthy 1987)	randomised trial	very serious[1]	no serious inconsistency	serious[2]	serious[3]	None	14/122 (11.5%)	14/124 (11.3%)	RR 1.02 (0.51 to 2.04)	2 more per 1000 (from 40 fewer to 84 more)	Very low
Intracranial haemorrhage grade III/IV[d]											
Total intracranial haemorrhage - birthweight 501 – 1750 grams											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious[2]	very serious[4]	None	20/122 (16.1%)	16/124 (12.9%)	RR 1.25 (0.68 to 2.30)	32 more per 1000 (from 41 fewer to 168 more)	Very low
Intracranial haemorrhage - birthweight 1101 - 1750 grams											
1 study (Luthy 1987)	randomised trial	serious1	no serious inconsistency	serious[2]	very serious[4]	None	4/122 (3.2%)	6/124 (4.8%)	RR 0.67 (0.19 to 2.3)	16 fewer per 1000 (from 39 fewer to 63 more)	Very low
Intracranial haemorrhage - birthweight 501 - 1100 grams											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious[2]	serious[3]	None	16/122 (12.9%)	10/124 (8.1%)	RR 1.6 (0.76 to 3.39)	48 more per 1000 (from 19 fewer to 193 more)	Very Low
Intracranial haemorrhage grade I/II[a]											
Total Intracranial haemorrhage - birthweight 501 – 1750 grams											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious[2]	serious[3]	None	19/122 (15.6%)	27/124 (22.1%)	RR 0.7 (0.41 to 1.2)	66 fewer per 1000 (from 131 fewer to 44 more)	Very Low
Intracranial haemorrhage - birthweight 1101 - 1750 grams											
1 study (Luthy 1987)	randomised trial	serious1	no serious inconsistency	serious[2]	very serious[4]	None	12/122 (9.8%)	16/124 (13.1%)	RR 0.75 (0.37 to 1.52)	33 fewer per 1000 (from 83 fewer to 68 more)	Very low
Intracranial haemorrhage - birthweight 501 - 1100 grams											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious2	very serious[4]	None	7/122 (5.7%)	11/124 (9%)	RR 0.64 (0.26 to 1.59)	32 fewer per 1000 (from 67 fewer to 53 more)	Very low
Severe respiratory distress syndrome											
Severe respiratory distress syndrome - birthweight 501 – 1750 grams											
1 study (Luthy 1987)	randomised trial	Serious[1]	no serious inconsistency	serious2	very serious[4]	None	33/122 (27%)	35/124 (28.2%)	RR 0.96 (0.64 to 1.44)	11 fewer per 1000 (from 102 fewer to 124 more)	Very low
Severe respiratory distress syndrome - birthweight 1101 - 1750 grams											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious2	very serious[4]	None	12/122 (9.8%)	17/124 (13.7%)	RR 0.72 (0.36 to 1.44)	38 fewer per 1000 (from 88 fewer to 60 more)	Very low
Severe respiratory distress syndrome - birthweight 501 - 1100 grams											

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cardiotocography	Intermittent auscultation	Relative (95% CI)	Absolute (95% CI)	Quality
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious2	very serious[4]	None	21/122 (17.2%)	18/124 (14.5%)	RR 1.19 (0.67 to 2.11)	28 more per 1000 (from 48 fewer to 161 more)	Very low
Umbilical cord arterial pH < 7.20											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious2	very serious[4]	None	6/122 (4.9%)	9/124 (7.3%)	RR 0.68 (0.25 to 1.85)	23 fewer per 1000 (from 54 fewer to 62 more)	Very low
Umbilical cord arterial pH ≥ 7.20											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious2	serious[3]	None	74/122 (60.7%)	72/124 (58.1%)	RR 1.04 (0.85 to 1.28)	23 more per 1000 (from 87 fewer to 163 more)	Low
Maternal outcomes											
Mode of birth – Caesarean birth											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious2	very serious[4]	None	19/122 (15.6%)	18/124 (14.5%)	RR 1.07 (0.59 to 1.94)	10 more per 1000 (from 60 fewer to 136 more)	Very low
Mode of birth – Spontaneous birth											
1 study (Luthy 1987)	randomised trial	serious[1]	no serious inconsistency	serious2	no serious imprecision	None	88/122 (72%)	97/124 (78%)	RR 0.92 (0.80 to 1.07)	63 fewer per 1000 (from 156 fewer to 55 more)	Low

CI Confidence interval; RR risk ratio
a Outcome not defined in the paper
1 Unclear how outcomes were ascertained, diagnosed or verified
2 Data was not analysed based on gestational age
3 Confidence interval crossed 1 default MIDs
4 Confidence interval crossed 2 default MIDs

Table 90: GRADE profile for comparison of cardiotocography versus intermittent auscultation – Observational studies

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cardiotocography	Intermittent auscultation	Relative (95% CI)	Absolute (95% CI)	Quality
Perinatal mortality											
Perinatal mortality - birthweight 700 – 1500 grams											
1 study (Shy 1988)	cohort	very serious[1,2,3]	no serious inconsistency	no serious indirectness	very serious[4]	None	82/213(39%)[a]	49/91(54%)	Adjusted[b] RR 0.91 (0.65 to 1.3)	48 fewer per 1000 (from 188 fewer to 162 more)	Very low
Perinatal mortality - birthweight 1100 – 1500 grams											
1 study (Shy 1988	cohort	very serious[5,6,7]	no serious inconsistency	no serious indirectness	no serious imprecision	None	30/136(22%)	10/37(27%)	Adjusted[b] RR 0.82 (0.39 to 1.7)	49 fewer per 1000 (from 165 fewer to 189 more)	Very low
Perinatal mortality - birthweight 700 – 1099 grams											

Quality assessment							Number of women		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cardiotocography	Intermittent auscultation	Relative (95% CI)	Absolute (95% CI)	Quality
1 study (Shy 1988	cohort	very serious[5,6,7]	no serious inconsistency	no serious indirectness	no serious imprecision	None	52/77 (68%)	39/54(72%)	Adjusted[b] RR 0.94 (0.63 to 1.4)	43 fewer per 1000 (from 267 fewer to 289 more)	Very low

CI Confidence interval; RR risk ratio
a Reported wrongly as 31% in the published paper.
b Adjusted for birth-weight, community hospital birth, rupture of membranes, and non-cephalic presentation
1 Unclear on what basis women allocated to have IA or EFM
2 Women's characteristics not reported
3 No standard protocol for intermittent auscultation used in 14 participating hospitals
4 Confidence interval crossed 2 default MIDs
5 Unclear how outcomes were ascertained, diagnosed or verified
6 Data was not analysed based on gestational age
7 Confidence interval crossed 1 default MIDs

13.3.5 Evidence statements

13.3.5.1 Maternal and neonatal outcomes

Evidence from 1 RCT (n = 246) showed no significant difference in the risk of perinatal mortality, intracranial haemorrhage, severe respiratory distress syndrome, seizure, and umbilical artery pH at birth in preterm babies born to women monitored with CTG compared with those receiving intermittent auscultation. The evidence across all outcomes was of very low quality.

Evidence from the same RCT (n = 246) showed no significant difference in rates of caesarean section or spontaneous vaginal birth between women who received intrapartum CTG and women who received intermittent auscultation. The evidence was of very low and low quality.

Very low quality evidence from 1 retrospective observational study (n = 304) also showed no significant difference in the risk of perinatal mortality in preterm babies born to women monitored with CTG compared with those receiving intermittent auscultation.

No evidence was found for the outcomes of; trauma/injury to the baby, periventricular leucomalacia (PVL) or white matter injury, neonatal sepsis, need for mechanical ventilation, and length of stay in neonatal intensive care unit or neonatal unit.

13.3.6 Health economics profile

A search was undertaken for health economic evidence on electronic fetal monitoring compared with intermittent auscultation at different gestational ages for unborn babies whose mothers are in suspected or diagnosed preterm labour. A total of 54 articles were identified by the search. After reviewing titles and abstracts, no papers were obtained. Therefore, no relevant economic evidence was identified for this question.

This question was not identified as a priority for health economic analysis as it was thought by the Committee that any recommendations would have a relatively low cost, especially as the equipment is readily available on labour wards.

13.3.7 Evidence to recommendations

13.3.7.1 Relative value placed on the outcomes considered

In this review, the Committee aimed to investigate to find whether the use of continuous CTG in labour was any more effective than intermittent auscultation in identifying fetuses who are at greater risk of poor outcomes arising as a complication of hypoxia-acidosis during preterm birth. The key outcomes of interest were fetal and neonatal death and the rates of more serious morbidities such as intracranial haemorrhage and respiratory distress syndrome.

13.3.7.2 Consideration of clinical benefits and harms

There was limited evidence in this area with only 2 studies (1 randomised, 1 comparative retrospective) of 550 preterm labours contributing to this review. The evidence showed that were no significant differences in any of the identified clinical outcomes in the CTG group compared with the group monitored by IA. Further subgroup analysis by birthweight showed no significant outcomes for intracranial haemorrhage, respiratory distress syndrome and perinatal mortality when analysed between the 2 groups. However, they felt that there was likely insufficient power to detect such outcomes, as most of the informative cases (event rates) would be among babies with the lowest gestational ages. Furthermore, the Committee acknowledged that although the caesarean section rates were not different between the 2

groups, they could not be sure how much emphasis should be placed on this finding, given the limitations of the studies. They also noted that the caesarean section rates were relatively low in both groups compared to the current rates but speculated that this probably was a reflection of changes in practice rather than anything to do with the actual monitoring strategy.

The Committee therefore concluded that the evidence from the 2 studies failed to demonstrate any benefit or harm from the use of CTG compared to IA in preterm labour. They argued that, in contrast to the term setting, the outcome in preterm babies was more likely to be determined by factors such as gestational age, birthweight and whether steroids were administered rather than intrapartum hypxoxia-acidosis (which is what intrapartum fetal monitoring is intended to detect). The Committee did not feel they could make a strong recommendation about the use of 1 method over the other. The Committee was aware that many women in preterm labour would have additional risks that would prompt the use of CTG and were also aware that even in the absence of risk factors use of CTG is in common practice for most women in preterm labour.

The Committee felt that women's views should be taken into account when the decision of fetal monitored is made and that it was important to provide ongoing information and support for mothers when using either CTG or IA. Some women will have a preference for IA because the intervention means that there will be greater interaction with a midwife (who would be present continuously though her labour) undertaking IA correctly every 15min during labour. However, some women may get more reassurance by not only the presence of the midwife but also having the more detailed information from a CTG.

The Committee also recognised that although in the majority of preterm labours monitoring the fetal heart rate would be considered standard practice, in certain circumstances an active decision may be taken not to monitor (for example, with extreme prematurity if the results of monitoring would not inform the obstetric or neonatal management). The Committee noted that in very premature labour (< 26 weeks) there is a high risk of neonatal morbidity and mortality, and survival is dependant more on fetal weight and maturity rather than intrapartum hypoxia and mode of delivery. The Committee did not discuss this in detail as they felt that this should be approached on a case-by-case basis taking account of the woman's views (and her partner's or family's preferences as appropriate) and the individual circumstances surrounding the pregnancy. Overall the Committee felt that, especially at low gestations, the decision-making process regarding the decision to monitor and the method by which to monitor was a difficult 1 and therefore a senior clinician needs to be involved in the discussion. .

With regards to continuous CTG monitoring, the Committee recognised the importance of a good quality recording, and that in circumstances where an external transducer may not provide this (e.g. high BMI) there would be the potential need for use of a fetal scalp electrode (FSE). The Committee acknowledged that the formal review of FSE had not been able to identify any evidence that met the protocol. The Committee had a consensus view that FSE had the potential to cause harm in preterm babies, and concluded that in the absence of evidence for or against its use, the potential risks and benefits of FSE in the preterm fetus should be considered and discussed with women on a case-by-case basis (see Section 13.4 on use of FSE).

With regards to monitoring by IA, the Committee agreed that in the absence of any evidence to the contrary, this should be carried out in accordance with the guidelines on monitoring the term fetus. The Committee acknowledged that performing IA is intensive and emphasised that its effectiveness as a method of monitoring would be highly dependent on availability and competence of staff to auscultate the fetal heart for 1 minute every 15 minutes in the first stage and every 5 minutes in the second stage. The Committee noted the importance of auscultation for a minute immediately after a contraction in order to confirm the absence of

late decelerations as this is reassuring (in IA and CTG) and is highlighted in the recommendations.

Consideration of health benefits and resource uses

The clinical evidence showed no difference in the outcomes when CTG was compared with IA when used in preterm labour. Both technologies are readily available in most settings already and thus it is unlikely that this intervention would have major cost implications.

Quality of evidence

The quality of the evidence was predominately very low for the different outcomes. The Committee noted that the proportion of high-risk women in each arm of the cohort study might not have been comparable and both studies were old and underpowered to detect important outcomes like perinatal mortality. However, in principle, the adverse outcomes that fetal monitoring is designed to prevent are commoner in preterm babies than in normally grown singleton babies at term.

Women in the IA group with abnormal fetal heart monitoring were assigned to have Caesarean section, and this does not reflect current clinical practice in that it is unlikely that women would go straight from IA to caesarean section without EFM in between.

Overall, the Committee did not feel sufficiently confident in the findings to make strong recommendations in favour of either CTG or IA.

13.3.7.3 Other considerations

Despite the paucity of research evidence relating to the method of fetal heart monitoring in preterm labour, the Committee felt it was nevertheless important to monitor fetal heart rate by some means during preterm labour. The Committee discussed the fact that physiological reserves available to combat hypoxia are less than those available to term babies. Hence, a preterm baby may become hypoxic sooner than its term counterpart. The Committee felt that it was important to inform women of the lack of evidence of benefit of CTG versus IA prior to offering monitoring.

13.3.8 Recommendations

The recommendations on monitoring options are in Section 13.5.5

13.4 Fetal scalp electrode

13.4.1 Introduction

Fetal scalp electrode (FSE) placement is used in electronic fetal heart monitoring (EFM) to assess the fetal heart rate (FHR) pattern when external monitoring is unable to be used, or when the signal quality is poor. There is uncertainty about risks and benefits of using an FSE to perform EFM in preterm labour. There is concern that application of FSE is associated with an increase in the risk of trauma and infection in preterm babies. Evidence is needed to determine if the benefits of using an FSE to perform EFM outweigh any risks and if there are gestational ages which the risks of FSE outweigh any potential benefits.

13.4.2 Review question

At what gestational age can a fetal scalp electrode be used for unborn babies whose mothers are in diagnosed preterm labour?

13.4.3 Evidence statements

No evidence was identified that addressed this question.

13.4.4 Evidence to recommendations

13.4.4.1 Relative value placed on the outcomes considered

The aim of fetal heart rate monitoring, whether it is by intermittent auscultation or continuous electronic fetal monitoring, performed either externally using an ultrasound transducer or by a fetal scalp electrode, is to detect 'fetal distress' and inform the use of interventions (primarily birth by caesarean section) before hypoxia-acidosis occur and cause harm to the unborn baby.

In order to establish whether the use of a FSE was of benefit in preterm labour, the Committee identified poor neonatal outcomes pertaining to hypoxia and acidosis (including mortality and intraventricular haemorrhage/periventricular leucomalacia) as priority outcomes, as well as cord blood gas values at birth as another measure of acidosis.

Balanced against this was the Committee's wish to understand whether FSE is associated with iatrogenic harm to the fetus and/or the woman. The Committee felt that the attachment of an electrode to the preterm fetus's soft scalp could theoretically cause trauma as well as local or more widespread infection in the fetus. These fetal outcomes were therefore prioritised. In terms of the woman, the Committee were interested to know whether the use of FSE was associated with an increase or a reduction in delivery by caesarean section, and hence mode of birth was prioritised. Maternal mortality was also measured as it was acknowledged that this would always be of concern as a potential harm.

Length of stay in neonatal intensive care unit or neonatal unit and need for mechanical intervention were also prioritised as these were considered to be potentially relevant to both poor outcomes arising from the presence of acidosis or iatrogenic harm caused by the use of FSE.

13.4.4.2 Consideration of clinical benefits and harms

There was no evidence that met the protocol which could demonstrate whether FSE usage was of benefit or harm to preterm babies, but there was consensus in the Committee (based

on their knowledge of preterm fetal anatomy) that attaching a scalp electrode had the potential to cause complications.

The Committee also noted that it is not current practice to routinely use FSE in preterm fetuses less than 34 weeks.

There was agreement that it is not always possible to obtain a good quality fetal heart rate trace from continuous electronic fetal monitoring performed externally using an ultrasound transducer (for example it can be difficult in women with high BMI).

The Committee were conscious of the fact that when a fetal heart rate cannot be detected adequately, the likely result is that a caesarean section will be offered, and if this was the case it might be preferable to apply a FSE rather than deliver the fetus when potentially there is nothing wrong. The Committee were also aware, however, of the limited evidence base for both the effectiveness of any form of continuous fetal monitoring (see Section 13.3) and the predictive value of the CTG trace (see Section 13.2) so felt that it was only appropriate to make a weak recommendation that FSE should be considered in such circumstances.

They also agreed that if FSE was considered then it was important to explain to the woman that the risks and benefits of FSE were unknown and that there were possible alternatives to EFM with FSE including the option not to monitoring at all.

13.4.4.3 Consideration of health benefits and resource uses

The Committee noted that the use of FSE is potentially costly given the level of skill required to deliver the intervention. The fact that it is not current practice below 34 weeks gestational age and the fact that if harm was caused through its use, this harm could potentially be serious and associated with high treatment costs. However, the Committee were also aware that any decision to use FSE would be taken with a view to preventing other serious harms to the fetus or woman which would themselves be associated with high QALY losses, so overall they felt that balance between the potential benefits and resource use supported a weak recommendation.

13.4.4.4 Quality of evidence

No evidence was identified for inclusion in the review.

13.4.4.5 Other considerations

There were no other considerations.

13.4.4.6 Key conclusions

FSE should only be considered when continuous electronic fetal monitoring using an external ultrasound transducer cannot be performed. The decision to use a FSE should be taken in discussion with the woman and a senior obstetrician. The Committee felt that in particular, it should be explained to the woman that the risks of FSE are unknown but that in the absence of being able to monitor externally, these must be balanced against the alternatives of no monitoring or expedited delivery.

13.5 Fetal blood sampling

13.5.1 Introduction

Fetal blood sampling is an invasive procedure used to obtain fetal blood for measurement of either pH or lactate. In preterm labour, as in term labour, it can be used as an adjunct to electronic fetal monitoring to establish whether an abnormal heart rate pattern is due to

hypoxaemia/acidosis. It is currently unclear in the preterm setting whether fetal blood sampling may confer additional benefit by reducing the risk of a false positive result from electronic fetal monitoring. This must be balanced against the potential adverse effects to the mother and fetus. A SR is needed to address the question of efficacy of fetal blood sampling in preterm labour, and to consider how this may vary for different gestational ages.

13.5.2 Review question

What is the utility of fetal blood sampling as an adjunct to intrapartum fetal heart rate monitoring at different gestational ages?

13.5.3 Evidence statements

No evidence was identified that addressed this question.

13.5.4 Evidence to recommendations

13.5.4.1 Relative value placed on the outcomes considered

The aim of fetal heart rate monitoring is to detect 'fetal distress' and inform the use of interventions (primarily birth by caesarean section) before hypoxia or acidosis occur and cause harm to the unborn baby. FBS is intended as an adjunct to continuous electronic fetal monitoring with the aim of confirming or refuting the presence of hypoxia oracidosis in those fetuses with an abnormal fetal heart rate pattern.

In order to establish whether FBS was efficacious, the Committee therefore identified poor neonatal outcomes pertaining to hypoxia and acidosis (including mortality and intraventricular haemorrhage/periventricular leucomalacia) as priority outcomes, as well as cord blood gas values at birth.

Balanced against this was the Committee's wish to understand whether FBS is associated with iatrogenic harm to the baby and/or the woman. The Committee felt the necessity to pierce the baby's scalp on an immature skull with large fontanelles to take a blood sample meant that the use of FBS theoretically could cause trauma and neonatal sepsis; these fetal outcomes were therefore prioritised. In terms of the woman, their main concern was whether the use of FBS was associated with unnecessary caesarean section so mode of birth was prioritised. Maternal mortality was also measured as it was acknowledged that this would always be of concern as a potential harm.

Length of stay in neonatal intensive care unit or neonatal unit and need for mechanical intervention were also prioritised as these were considered to be potentially relevant to both poor outcomes arising from the presence of acidosis or iatrogenic harm caused by the use of FBS.

13.5.4.2 Consideration of clinical benefits and harms

There was no evidence that met the protocol that showed FBS was of benefit to preterm fetuses but there was consensus in the Committee (based on their knowledge of preterm fetal anatomy) that piercing the fetus's scalp in order to take a blood sample was potentially associated with risks such as bleeding, infection and dural puncture.

The Committee also noted that it is not current practice to routinely use FBS in preterm fetuses less than 34 weeks.

The Committee were aware of some observational studies (that did not conform to the protocol and thus were not included in the review) in which women were recruited from as early as 26 weeks gestational age and FBS was used. However, these studies did not

provide any information about adverse outcomes and were unhelpful to addressing the issues of risk. However, they did confirm that FBS is used in some circumstances at these low gestational ages.

The Committee were aware that the Intrapartum Care Guideline (CG 190) recommends the use of FBS as an adjunct to electronic fetal monitoring in term babies but did not feel that they could extrapolate this to the entire preterm population because of the theoretical effect of anatomical and physiological differences between term and preterm babies, especially at lower gestational ages.

The Committee also noted that FBS would only be used as an adjunct to electronic fetal monitoring and so if electronic fetal monitoring is not being used (for example if intermittent auscultation is being carried out instead), then there is no justification for using FBS.

In light of all these considerations the Committee felt that FBS should not be used in fetuses below 34 weeks gestational age on the basis of theoretical concerns about complications below this gestation. Their main concerns related to the risk of excessive bleeding and of cerebro-spinal fluid leakage resulting from accidental dural puncture via the anterior fontanelle. These risks are likely to be mitigated by advancing gestation and therefore the Committee felt that, after consultation with a senior obstetrician and the woman, FBS could be considered after 34 weeks gestation if the likely benefits outweigh the risks. The Committee felt that it was reasonable to recommend that if a FBS is performed it should be done in accordance with the advice given in the NICE intrapartum care guideline (CG 190).

13.5.4.3 Consideration of health benefits and resource uses

The Committee noted that the use of FBS is potentially costly given the level of skill required to deliver the intervention, the fact that it is not current practice below 34 weeks gestational age and the fact that if harm was caused through its use, this harm could potentially be serious and associated with high treatment costs. Overall they felt that balance between the potential benefits and resource use only supported a weak recommendation for the use of FBS in the specific subgroup of fetuses between 34^{+0}-36^{+6} gestational weeks.

13.5.4.4 Quality of evidence

No evidence was identified for inclusion in the review.

13.5.4.5 Key conclusions

FBS should not be performed before 34 gestational weeks. The use of FBS in fetuses between 34^{+0} to 36^{+6} gestational weeks should be discussed with a consultant obstetrician and with the woman. If FBS is performed it should be done in accordance with the advice given in the NICE guideline on intrapartum care (CG 190).

13.5.5 Recommendations

Monitoring options: cardiotocography and intermittent auscultation

44. Discuss with women in suspected, diagnosed or established preterm labour (and their family members or carers as appropriate):

- the purpose of fetal monitoring
- the clinical decisions it informs at different gestational ages
- if appropriate, the option not to monitor the fetal heart rate (for example, at the threshold of viability).

45. Involve a senior obstetrician in discussions about whether and how to monitor the fetal heart rate in women between 23^{+0} and 24^{+6} weeks of pregnancy.

46. Explain the different fetal monitoring options to the woman (and her family members or carers as appropriate), being aware that:

- there is limited evidence about the usefulness of specific cardiotocography features suggestive of hypoxia or acidosis in preterm babies
- the available evidence is broadly consistent with that for babies born at term (see section 1.10 in the NICE guideline on intrapartum care)
- a normal cardiotocography trace is reassuring and indicates that the baby is coping well with labour, but an abnormal trace does not necessarily indicate that fetal hypoxia or acidosis is present.

47. Explain to the woman (and her family members or carers as appropriate) that there is an absence of evidence that using cardiotocography improves the outcomes of preterm labour for the woman or the baby compared with intermittent auscultation.

48. Offer women in established preterm labour but with no other risk factors (see section 1.10 in the NICE guideline on intrapartum care) fetal heart rate monitoring using either:

- cardiotocography using external ultrasound or
- intermittent auscultation.

Take the woman's preferences into account when deciding on choice of monitoring option.

49. For guidance on using intermittent auscultation for fetal heart rate monitoring, see recommendation 1.10.1 in the NICE guideline on intrapartum care.

Fetal scalp electrode

50. Do not use a fetal scalp electrode for fetal heart rate monitoring if the woman is less than 34^{+0} weeks pregnant unless all of the following apply:

- it is not possible to monitor the fetal heart rate using either external cardiotocography or intermittent auscultation
- it has been discussed with a senior obstetrician
- the benefits are likely to outweigh the potential risks
- the alternatives (immediate birth, intermittent ultrasound and no monitoring) have been discussed with the woman and are unacceptable to her.

51. Discuss with the woman (and her family members or carers as appropriate) the possible use of a fetal scalp electrode between 34^{+0} and 36^{+6} weeks of pregnancy if it is not possible to monitor the fetal heart rate using either external cardiotocography or intermittent auscultation.

Fetal blood sampling

52. **Do not carry out fetal blood sampling if the woman is less than 34^{+0} weeks pregnant.**

53. **Discuss with the woman the possible use of fetal blood sampling between 34+0 and 36^{+6} weeks of pregnancy if the benefits are likely to outweigh the potential risks.**

54. **When offering fetal blood sampling, discuss this with the woman as described in recommendation 1.10.41 in the NICE guideline on intrapartum care, and advise her that if a blood sample cannot be obtained a caesarean section is likely.**

13.5.6 Research recommendations

Research question	6. Is intermittent auscultation or electronic fetal monitoring effective in the preterm fetus?
Why this is needed	
Importance to 'patients' or the population	Intermittent auscultation of the fetal heart creates a very different patient experience to continuous electronic recording, but the relative impacts of these modes of fetal monitoring in terms of clinical decision-making and infant outcome has not been tested in a population of women in preterm labour. Fetal monitoring is ingrained as part of term deliveries and there is a lack of evidence to allow extrapolation from term birth to pre-term birth, both in terms of which mode to use and how to interpret the results. Identification of the fetus at risk of hypoxia-acidosis would reduce unnecessary intervention whilst ensuring delivery of the compromised fetus before neurological damage or death ensues.
Relevance to NICE guidance	High priority, as there is currently very poor quality evidence on which to base GDG recommendations.
Relevance to the NHS	The experience of women in labour at any gestation is an important metric of the quality of maternity services. Improved detection of the fetus at risk of hypoxia-acidosis would reduce mortality and morbidity from preterm birth.
National priorities	NHS outcomes framework 2015-6, #4, Ensuring that people have a positive experience of care. NHS Outcomes Framework #1: Preventing people from dying prematurely
Current evidence base	Existing trials reviewed by the Guideline Committee have small numbers and serious methodological flaws.
Equality	There are no obvious equality issues; the population is defined by gestational age.
Feasibility	There is no reason in principle why more adequately designed and powered trials should not be carried out. The ethical issues are not in principle different from those affecting other perinatal trials.
Other comments	None

14 Mode of birth

14.1 Introduction

The potential health risks facing babies born preterm may be compounded by complications occurring at the time of delivery. In clinical practice, caesarean section (CS) delivery is performed if there are fetal indications that this would be safer than vaginal birth (for example, evidence of hypoxia-acidosis in labour when vaginal delivery is not imminent and rapid delivery may prevent permanent neurological damage). CS would also be preferable to a traumatic vaginal delivery (for example, with a preterm footling breech presentation, most clinicians would offer CS rather than vaginal delivery). However, the value of CS in comparison to vaginal birth, in the absence of clinical indications, is uncertain for women in suspected or diagnosed preterm labour.

14.1.1 Review question

For women who present in suspected or diagnosed preterm labour (who have not planned antenatally to give birth by caesarean section (CS) and for whom there are no other known indications for CS birth), what is the clinical effectiveness of deciding to carry out a CS compared with deciding to allow vaginal birth?

This review question aims to assess whether there is any difference in maternal and neonatal outcomes when CS (which has not been planned before the onset of labour for other indications) is compared with vaginal birth for women in suspected or diagnosed preterm labour. As this question was set out to assess the comparative effectiveness of 2 interventions, RCTs were selected as the best study design. The Committee also preselected subgroup analysis at the protocol stage based on the following factors: breech presentation, cephalic presentation, instrumental birth and gestational age.

The Committee had extensive discussion when the protocol was developed about the description of the type of CS to be captured. The Committee agreed that this question is not about planned mode of birth decided antenatally but seeks to answer what is the optimum mode of birth for women who present in preterm labour. It does not examine the emergency use of CS for acute fetal or maternal compromise in women attempting vaginal delivery, because they felt that there was overwhelming *a priori* belief that emergency CS in this scenario was beneficial. The differences between the focus of this review and much of the data in the literature (which refer to emergency CS) was taken into consideration in the interpretation of evidence and its influence in the decision making.

14.1.2 Description of included studies

One Cochrane SR and meta-analysis (Alfirevic 2013) with 4 component RCTs (Viegas 1985; Penn 1996; Zlatnik 1993; Wallace 1984) is included in this review. Studies which investigated a pre-planned mode of birth for women in suspected or diagnosed preterm labour were excluded for the purposes of this review.

The studies included in the Cochrane SR were conducted in various locations including 2 studies in the USA, 1 in the UK and 1 in Singapore.

All included trials examined the impact of the mode of birth (immediate CS or vaginal birth) on neonatal outcomes in preterm and very low birth weight babies (gestational age across all studies ranged from 26 weeks to 33 weeks) with cephalic or breech presentation. Three studies included preterm babies with only breech presentation (Penn 1996; Zlatnik 1993; Viegas 1985) and 1 study included only babies with cephalic presentation (Wallace 1984).

The quality assessment of the included trials was downgraded due to study design (outcome assessors were unblinded), incomplete outcome data and small sample size. In addition, recruitment in all 4 included trials was stopped early.

The main issue with the interpretation of results in this SR is that a large number of women randomised to 1 type of mode of birth or another actually gave birth by the other method (in other words a cross-over effect with women ending in a different group from the 1 to which they were randomised). More specifically, 3 trials (Penn 1996; Zlatnik 1993; Wallace 1984) in the SR included 20% (9/46) of women who were allocated to caesarean section group but subsequently gave birth vaginally because birth was too rapid to allow a caesarean to be performed, and 21% (9/43) of women allocated to vaginal birth group who actually gave birth by caesarean section for fetal or maternal indications.

14.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

Data is reported in 2 GRADE profiles separately for neonatal and maternal outcomes.

- Table 91: GRADE profile for comparison of CS (which has not been planned before the onset of labour for other indications) versus vaginal birth - neonatal outcomes
- Table 92: GRADE profile for comparison of CS (which has not been planned before the onset of labour for other indications) versus vaginal birth - maternal outcomes

The grading of evidence from the SR was assessed at high quality and then downgraded if there were any issues identified that would undermine the confidence in the findings.

Table 91: GRADE profile for comparison of CS (which has not been planned before the onset of labour for other indications) versus vaginal birth - neonatal outcomes

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CS	Vaginal Birth	Relative (95% CI)	Absolute	Quality
Perinatal death											
1 meta-analysis of 3 studies (Alfirevic 2013)	randomised trials	serious1	no serious inconsistency	serious[2]	serious[3]	None	2/46 (4.3%)	8/43 (18.6%)	RR 0.29 (0.07 to 1.14)	132 fewer per 1000 (from 173 fewer to 26 more)	Very low
Perinatal death – Breech (subgroup analysis)											
1 meta-analysis of 2 studies (Alfirevic 2013)	randomised trials	serious[1]	no serious inconsistency	serious[4]	serious[5]	None	1/23 (4.3%)	6/28 (21.4%)	RR 0.28 (0.05 to 1.49)	154 fewer per 1000 (from 204 fewer to 105 more)	Very low
Perinatal death – Cephalic (subgroup analysis)											
1 study (Alfirevic 2013)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	None	1/23 (4.3%)	2/15 (13.3%)	RR 0.33 (0.03 to 3.29)	89 fewer per 1000 (from 129 fewer to 305 more)	Low
Brain injury- Hypoxic ischemic encephalopathy											
1 study (Alfirevic 2013)	randomised trials	very serious[1,7]	no serious inconsistency	very serious[4]	very serious[6]	None	1/5 (20%)	0/7 (0%)	RR 4 (0.2 to 82.01)	NC	Very low
Intracranial haemorrhage											
1 meta-analysis of 4 studies (Alfirevic 2013)	randomised trials	very serious[1,5,7]	no serious inconsistency	serious[8]	very serious[6]	None	4/56 (7.1%)	4/54 (7.4%)	RR 0.92 (0.27 to 3.14)	6 fewer per 1000 (from 54 fewer to 159 more)	Very low
Intracranial haemorrhage – Breech (subgroup analysis)											
1 meta-analysis of 3 studies (Alfirevic 2013)	randomised trials	very serious[1,5,8]	serious[9]	serious[4]	very serious[6]	None	1/33 (3%)	3/39 (7.7%)	RR 0.58 (0.12 to 2.86)	32 fewer per 1000 (from 68 fewer to 143 more)	Very low
Intracranial haemorrhage – Cephalic (subgroup analysis)											
1 study (Alfirevic 2013)	randomised trials	serious[1,5,7]	no serious inconsistency	no serious indirectness	very serious[6]	None	3/23 (13%)	1/15 (6.7%)	RR 1.96 (0.22 to 17.1)	64 more per 1000 (from 52 fewer to 1000 more)	Very low
Abnormal follow-up in childhood (outcome not defined) – Cephalic											
1 study (Alfirevic 2013)	randomised trials	very serious[1,5,7]	no serious inconsistency	no serious indirectness	very serious[6]	None	4/23 (17.4%)	4/15 (26.7%)	RR 0.65 (0.19 to 2.22)	93 fewer per 1000 (from 216 fewer to 325 more)	Very low
Respiratory distress syndrome											

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	CS	Vaginal Birth	Relative (95% CI)	Absolute	Quality
1 meta-analysis of 3 studies (Alfirevic 2013)	randomised trials	serious[1]	no serious inconsistency	serious[2]	serious[3]	None	9/53 (17%)	16/50 (32%)	RR 0.55 (0.27 to 1.1)	144 fewer per 1000 (from 234 fewer to 32 more)	Very low
Respiratory distress syndrome – Breech (subgroup analysis)											
1 meta-analysis of 2 studies (Alfirevic 2013)	randomised trials	serious[1,5]	no serious inconsistency	serious[4]	very serious[6]	None	6/30 (20%)	12/35 (34.3%)	RR 0.57 (0.25 to 1.3)	147 fewer per 1000 (from 257 fewer to 103 more)	Very low
Respiratory distress syndrome – Cephalic (subgroup analysis)											
1 study (Alfirevic 2013)	randomised trials	very serious1[.5,7]	no serious inconsistency	no serious indirectness	very serious[6]	None	3/23 (13%)	4/15 (26.7%)	RR 0.49 (0.13 to 1.88)	136 fewer per 1000 (from 232 fewer to 235 more)	Very low
Need for mechanical ventilation											
1 study (Alfirevic 2013)	randomised trials	very serious[1,5,7]	no serious inconsistency	serious4	very serious[6]	None	4/5 (80%)	3/7 (42.9%)	RR 1.87 (0.71 to 4.88)	373 more per 1000 (from 124 fewer to 1000 more)	Very low

RR risk ratio, CI confidence interval
1 Outcomes assessors were not blinded to the group allocation
2 Participants in 2 studies had breech presentation
3 Evidence was downgraded by 2 due to very serious imprecision as 95%CI crossed 2 default MIDs
4 All participants had breech presentation
5 Detail of allocation concealment not reported
6 Evidence was downgraded by 2 due to very serious imprecision as 95%CI crossed 2 default MIDs
7 Unclear detail of randomisation
8 Participants in 3 studies had breech presentation
9 Evidence was downgraded by 1 due to serious heterogeneity (chi-squared p<0.1, I-squared inconsistency statistic of 50%-74.99%) and no plausible explanation was found with subgroup analysis

Table 92: GRADE profile for comparison of CS (which has not been planned before the onset of labour for other indications) versus vaginal birth - maternal outcomes

Quality assessment							No of patients		Effect		
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	Other considerations	**CS**	**Vaginal birth**	**Relative (95% CI)**	**Absolute**	**Quality**
Postpartum haemorrhage											
1 meta-analysis of 3 studies (Alfirevic 2013)	random ised trials	serious[3]	no serious inconsistency	serious[1]	very serious[2]	None	1/54 (1.9%)	0/51 (0%)	RR 3.69 (0.16 to 83.27)	NC	Very low
Postpartum haemorrhage – Breech (subgroup analysis)											
1 meta-analysis of 2 studies (Alfirevic 2013)	random ised trials	serious[3]	no serious inconsistency	serious[4]	serious[2]	None	1/30 (3.3%)	0/35 (0%)	RR 3.69 (0.16 to 83.27)	NC	Very low
Postpartum haemorrhage – Cephalic (subgroup analysis)											
1 study (Alfirevic 2013)	random ised trials	seriou[s3]	no serious inconsistency	no serious indirectness	very serious[2]	None	0/23 (0%)	0/15 (0%)	NC	NC	Very low
Maternal wound infection											
1 meta-analysis of 3 studies (Alfirevic 2013)	random ised trials	serious[3]	no serious inconsistency	serious[1]	very serious[2]	None	1/53 (1.9%)	1/50 (2%)	RR 1.16 (0.18 to 7.7)	3 more per 1000 (from 16 fewer to 134 more)	Very low
Maternal wound infection – Breech (subgroup analysis)											
1 meta-analysis of 2 studies (Alfirevic 2013)	random ised trials	serious[3]	no serious inconsistency	serious[4]	very serious[2]	None	1/30 (3.3%)	1/35 (2.9%)	RR 1.16 (0.18 to 7.7)	5 more per 1000 (from 23 fewer to 191 more)	Very low
Maternal wound infection – Cephalic (subgroup analysis)											
1 study (Alfirevic 2013)	random ised trial	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	None	0/23 (0%)	0/15 (0%)	RR 1.0	0 more per 1000 (from 0 more to 0more)	Moderate
Other maternal infection (outcome not defined)											
1 meta-analysis of 3 studies (Alfirevic 2013)	random ised trials	serious[3]	no serious inconsistency	serious[1]	very serious[5]	None	10/53 (18.9%)	4/50 (8%)	RR 2.63 (1.02 to 6.78)	130 more per 1000 (from 2 more to 462 more)	Very low
Other maternal infection (outcome not defined) – Breech (subgroup analysis)											

Quality assessment							No of patients		Effect		
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	Other considerations	**CS**	**Vaginal birth**	**Relative (95% CI)**	**Absolute**	**Quality**
1 meta-analysis of 2 studies (Alfirevic 2013)	random ised trials	serious[3]	no serious inconsistency	serious1	serious[5]	None	10/30 (33.3%)	4/35 (11.4%)	RR 2.63 (1.02 to 6.78)	186 more per 1000 (from 2 more to 661 more)	Very low
Other maternal infection (outcome not defined) – Cephalic (subgroup analysis)											
1 study (Alfirevic 2013)	random ised trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	None	0/23 (0%)	0/15 (0%)	RR 1.0	0 more per 1000 (from 0 more to 0more)	Moderate

NC not calculable, RR risk ratio, CI confidence interval
1 Participants in 2 studies had breech presentation
2 Evidence was downgraded by 2 due to very serious imprecision as 95%CI crossed 2 default MIDs
3 Outcomes assessors were not blinded to the group allocation
4 All participants had breech presentation
5 Evidence was downgraded by 1 due to serious imprecision as 95%CI crossed 1 default MID

14.1.4 Evidence statements

Neonatal outcomes

Perinatal mortality

Evidence from 1 SR of 3 RCTs (n=89) showed no significant difference in perinatal death in preterm babies born by caesarean section (not planned before the onset of labour) compared with those born by vaginal birth. A subgroup analysis by breech for cephalic presentation confirmed the direction of results from the overall analysis. The evidence was of low or very low quality.

Brain injury

Evidence from 1 small RCT (n=13) showed no significant difference in the proportion of preterm babies with brain injury (hypoxic ischemic encephalopathy) when born by caesarean section (not planned before the onset of labour) compared with those born by vaginal birth. A subgroup analysis by babies presentation found that the risk of brain injury was higher with CS that with vaginal birth but not significantly.

Evidence from 1 SR of 4 studies (n=110) also showed no significant difference in the incidence of intracranial pathology in preterm babies born by caesarean section (not planned before the onset of labour) compared with those born by vaginal birth.

The evidence across all studies for these outcomes was of very low quality.

Neurodevelopment follow-up in childhood

There was very low quality evidence from 1 study (n=38) that showed no difference in the incidence of abnormal follow-up in childhood in preterm babies born by caesarean section compared with those born by vaginal birth. The evidence was from a small study and abnormal follow-up was not defined by the study so results should be interpreted with caution.

Respiratory distress syndrome

The evidence from 1 SR of 3 RCTs (n=103) showed no significant difference in the outcome of respiratory distress syndrome in preterm babies born by caesarean section compared with those born by vaginal birth. Evidence from 1 study (n=15) also showed no significant difference in the need for mechanical ventilation in preterm babies born by caesarean section compared with those born by vaginal birth. The evidence across all studies was of low and very low quality.

Maternal outcomes

Postpartum haemorrhage

The evidence from 1 SR of 3 RCTs (n=105) showed no significant difference in the incidence of postpartum haemorrhage in women whose babies were born by caesarean section compared with those whose babies were born by vaginal birth. The evidence across all studies was of very low quality.

Maternal infection

The evidence from 1 SR of 3 RCTs (n = 103) showed no significant difference in the incidence of maternal wound infection in women whose babies were born by caesarean

section compared with those whose babies were born by vaginal birth. However evidence from the same RCTs (n = 103) showed significantly higher incidence of other maternal infection in women whose babies were born by caesarean section compared with those whose babies were born by vaginal birth. The evidence across all studies was of very low quality.

14.1.5 Health economics profile

A search was undertaken for health economic evidence women who present in suspected or diagnosed preterm labour (who have not planned antenatally to give birth by caesarean section (CS) and for whom there are no other known indications for CS birth), what is the clinical effectiveness of deciding to carry out a CS compared with deciding to allow vaginal birth . A total of 82 articles were identified by the search. After reviewing titles and abstracts, a single paper was obtained and then excluded as it was based on a fetal indication for caesarean section. Therefore, no relevant economic evidence was identified for this question.

This question was identified as a medium priority for health economic analysis as there are potentially large resource implications. However, it was thought that the cost-effectiveness would simply reinforce the clinical evidence if that suggested the superiority of 1 mode of birth. Conversely, if the clinical evidence is ambiguous then the Committee thought that economic analysis might provide little added value to aid decision making. Ultimately, no new health economic analysis was undertaken and the clinical evidence review did not demonstrate the superiority of a particular mode of birth.

14.1.6 Evidence to recommendations

14.1.6.1 Relative value placed on the outcomes considered

The Committee considered neonatal outcomes of critical importance in answering the question whether CS which was not planned before the onset of labour would be more effective than vaginal birth for women in suspected or diagnosed preterm labour.

In relation to neonatal outcomes, neonatal mortality, long term neurodevelopmental delay and respiratory distress were considered the most critical outcomes. Caesarean section is known to be associated with an increased risk of respiratory morbidity in term neonates, probably because the process of vaginal delivery is associated with a more effective transition to postnatal life.

With regards to maternal outcomes, infections acquired during delivery, although not considered critical for this review question, were still considered to be important. The risk of postnatal infection (pyrexia, endometritis, puerperal sepsis), thrombosis and pulmonary embolism, and excessive blood loss are higher after caesarean section than vaginal delivery. However, some of these complications could also be due to the underlying causes that lead to the need for caesarean section, and not necessarily as CS-induced complications.

Although the Committee considered that CS is a relatively safe procedure, maternal mortality was still an important outcome to be considered.

The Committee noted that CS for extremely preterm babies may pose technical difficulties and require incision in the upper uterine segment. The Committee were aware that a history of CS birth, especially those that involve an incision of the upper uterine segment, will have implications for the selection of mode of birth for future pregnancies,

14.1.6.2 Consideration of clinical benefits and harms

The included evidence on neonatal outcomes in women in whom CS had not already been planned before the onset of labour concluded that CS was neither beneficial nor harmful compared to vaginal birth. Although there was an indication that the risk of perinatal mortality for babies may be lower for those who were delivered by CS compared to vaginal birth, this evidence was inconclusive.

In addition, there was no clear direction of effect for the neonatal outcomes from investigating which is the optimal mode of delivery of preterm babies in subgroup analysis by cephalic or breech presentation of the baby.

However, there was some evidence showing that there may be an adverse effect of CS in increasing “other” maternal infection.

The Guideline Committee was aware that in current clinical practice the selection of the mode of birth for preterm babies is often extrapolated from full-term babies. For example if the baby has a breech presentation then CS would be the most favoured mode of birth. There was no evidence in this review to suggest that following current practice for term babies (e.g. delivery by CS for breech presentation) would be harmful to the baby. Indeed the Committee noted that the point estimate for reduction in perinatal death in babies with breech presentation was RR 0.28 (0.05 to 1.49) - a considerable reduction in risk but with low precision given the wide confidence intervals. Given that the majority of women included in the studies were of 26-36 weeks of gestational age, the Committee did not feel that there was sufficient evidence to make any recommendations about the optimal mode of birth for women in pregnancy below 26 weeks.

14.1.6.3 Consideration of health benefits and resource uses

A planned caesarean section is usually a more expensive procedure than a planned vaginal birth but an unplanned emergency caesarean section following a planned vaginal birth is more expensive still. However, the Committee did not think that the costs of a particular mode of birth would be an important driver of cost-effectiveness if a particular mode of birth produced better maternal and neonatal outcomes as the cost of adverse outcomes and complication would more than offset any differential in the cost of birth itself. However, the evidence did not demonstrate the superiority of any 1 mode of birth and therefore the implications for resource uses and health benefits, if any, is uncertain.

14.1.6.4 Quality of evidence

The evidence across all studies was of very low quality mainly due to serious limitations on risk of bias and imprecision which gives less confidence to the direction of effects.

The main methodological challenge for this review was the very limited data from randomised trials (due to low numbers of women recruited) and the high proportion of babies (20%) who were not delivered by the planned (randomised) mode of birth. This can restrict the generalization of results as the cross over effect of moving from 1 randomised arm to another can introduce bias. Cross over may reflect rapid progress of preterm labour which ends in vaginal birth despite caesarean section being planned and conversely, problems developing during labour may require caesarean section despite aiming for vaginal birth. Three out of 4 studies in the meta-analysis included preterm babies with only breech presentation which are also in higher risk of developing complications than preterm babies with cephalic presentation.

14.1.6.5 Other considerations

The Committee were aware of the evidence on CS at term, as reviewed in the NICE CS guideline. They felt that the significant maternal effects (such as perineal and abdominal pain during birth, and 3 days post-partum, injury to vagina, early postpartum haemorrhage and obstetric shock) of CS would be similar at term and preterm, although preterm CS would be more likely to require a vertical uterine incision, after which most clinicians would advise caesarean delivery in the next pregnancy. They noted the adverse effects of increased blood loss and risk of wound infection and venous thromboembolism following surgery. The Committee had less confidence about extrapolating the neonatal effects of CS from term to preterm except that upper segment incision has implications for future delivery compared with standard lower segment CS. Nevertheless, the Committee noted that babies born following CS at term had an increased chance of admission for respiratory distress compared with babies born vaginally.

14.1.6.6 Key conclusions

There was inconclusive evidence about the difference in neonatal and maternal outcomes for either CS or vaginal birth for women in suspected or diagnosed preterm labour.

14.1.7 Recommendations

55. Discuss the general benefits and risks of caesarean section and vaginal birth with women in suspected or diagnosed preterm labour and women with P-PROM (and their family members or carers as appropriate) - see recommendation 1.1.2.1 in the NICE guideline on caesarean section.

56. Explain to women in suspected or diagnosed preterm labour and women with P-PROM about the benefits and risks of caesarean section that are specific to gestational age. In particular, highlight the difficulties associated with performing a caesarean section for a preterm birth, especially the increased likelihood of a vertical uterine incision and the implications of this for future pregnancies.

57. Explain to women in suspected or diagnosed preterm labour that there are no known benefits or harms for the baby from caesarean section, but the evidence is very limited.

58. Consider caesarean section for women presenting in suspected or diagnosed preterm labour between 26^{+0} and 36^{+6} weeks of pregnancy with breech presentation, and explain to the woman that:

- caesarean section for breech presentation for preterm babies is common but not universal practice
- this practice is based on an extrapolation of evidence of best management for breech presentation for babies born at term
- there is some evidence that there may be a large reduction in perinatal mortality associated with caesarean section for preterm babies with breech presentation, but overall the evidence is inconclusive.

15 Timing of cord clamping for preterm babies

15.1 Introduction

There has been a recent change of practice in management of the third stage of labour in term deliveries, away from immediate clamping of the umbilical cord to deferred cord clamping. In healthy term babies the evidence supports deferred clamping (RCOG Scientific Impact Paper No. 1 2015). Immediate clamping of the cord reduces blood flow from the placenta to the baby, and this could impact upon the transition from fetal to neonatal circulation. The loss of blood volume also lessens the baby's iron stores and thus increases the risk of anaemia after birth. Both these effects could be particularly important in preterm babies, but need to be set against the possible risks of delayed clamping, such as increased risk of jaundice or delay in resuscitation at a critical time for both mother and baby. This chapter examines the evidence relating to preterm babies and their mothers.

15.1.1 Review question

In preterm birth, does later or delayed cord clamping in active management of third stage improve maternal and neonatal outcomes compared to earlier or immediate cord clamping?

15.1.2 Description of included studies

Four studies were included in this review (Rabe 2012; March 2011; Elimian 2014, Ranjit 2014).

The included studies consisted of 1 SR with 15 component trials from a variety of locations in developed countries (Rabe 2012), 2 RCTs from the USA (March 2011; Elimian 2014) and 1 RCT from India (Ranjit 2014).

All included trials evaluated the effect of the timing of umbilical cord clamping of preterm infants on neonatal outcomes. Only 1 included study reported results for maternal outcomes (Ranjit 2014). The timing of cord clamping varied between the studies. "Earlier" cord clamping was defined as immediate clamping of the cord ranging from 5 seconds to 30 seconds after the birth of the baby. "Later" cord clamping was defined as clamping the cord from 30 seconds to 3 minutes after the birth of the baby. One study in the SR(Rabe 2012) and 1 of the individual RCTs (March 2011) compared earlier cord clamping plus cord milking with earlier cord clamping without cord milking.

Four RCTs in the SR included only women giving birth at less than 30 weeks gestation, 7 trials included women giving birth at less than 33 weeks gestation, 4 trials included women giving birth at less than 35 weeks gestation and 2 trials only included low birth weight babies under 2000 grams and on trial included babies less than 36 weeks. Whilst some of the trials reported sub-group analyses by gestation at randomisation, they did not report outcomes analysed by gestation of babies at birth therefore this information could not be further explored.

Not all the included studies in the SR followed the active management of third stage in the cord clamping regarding the administration of a uterotonic. Five out of 15 studies in the SR specified that a uterotonic was used intravenously (IV) at birth. The type of uterotonic, the dose used, and the timing of administration varied between the studies. The use of an uterotonic was not reported in the other 10 studies included in the SR, nor was it reported in the 2 separately reviewed clinical trials. For details of study interventions and comparisons please see table Table 94 and Table 95 below and the evidence tables in Appendix H.

Seven of the studies in the SR provided information on the choice of mode of birth: in 1 trial all babies were born by caesarean section, in 3 all babies were born vaginally, and in 3 RCTs there was a mixed population in which an approximately equal number of babies were born by vaginal or caesarean delivery. Mode of birth was not specified in the remaining 3 RCTs of the SR nor in the 2 separate trials.

Babies born by vaginal birth were held 10 to 15 inches below the level of the introitus in the later cord clamping groups in most studies, except in 1 study where babies were held at the level of the uterus. Babies born by caesarean section were also held below the level of incision, apart from in 3 studies where the babies were held above the uterus either beside the woman's legs or on her thighs.

A summary of the main characteristics of each study in the SR and the individual trials is given in below Table 93 as these characteristics varied considerably.

Table 93: Summary of baseline characteristics of included studies

Study author and date	Definition of early cord clamping (ECC)	Definition of delayed cord clamping (DCC) or details of cord milking	Position of the baby at time of clamping in Intervention group (DCC)	Uterotonic (type/route/time)	Country	Gestational age or birth weight	Mode of birth	Additional comments
Aladagandy 2006 (included in Rabe 2012)	Immediately after birth.	30–90 seconds after birth	Infant held as low as the cord's length permitted	Syntocinon 5 IU IV at birth of baby's head	UK (Glasgow)	24^{+0} – 32^{+6}	Vaginal/caesarean Caesarean section: ECC: n = 12/23 DCC: n = 14/23	Intention to treat analysis. n = 3/23 allocated to DCC had early clamping (1 due to short cord, 2 asked for by neonatologist).
Baezinger 2007 (included in Rabe 2012)	Immediately after birth (< 20 seconds)	60–90 seconds after birth	Infant held as low as possible for vaginal births, and 15 cm below the placenta at caesarean section.	Syntocinon IV immediately after birth in DCC group	Switzerland	24^{+0}–32^{+6} weeks	Vaginal/caesarean Caesarean section: ECC: n = 16/24 DCC: n = 11/15	N =3 infants in ECC group died within 72 hours after birth (n =1 from sepsis, n =2 from hyaline membrane disease)
Elimian 2014	Within 5 seconds of birth	30–35 seconds after birth	Not reported	Not reported	USA	24–34 weeks	Not reported	
Hofmeyr 1988 (included in Rabe 2012)	Immediately after birth	Intervention 1: 60 seconds after birth Intervention 2: 60 seconds after birth	Not reported	Ergometrine at birth in second intervention group (route not reported)	South Africa	< 35 weeks	Not reported	
Hofmeyr 1993 (included in Rabe 2012)	Shortly after birth, according to usual practice	60–120 seconds after birth	At level of uterus at vaginal births or above level of uterus on the woman's thighs at caesarean section	Not reported		< 2000 g (gestational age not reported)	Not reported	N = 8 in DCC had cord clamped early because of cord around the neck or need for resuscitation
Hosono 2008 (included in Rabe 2012)	Immediately after birth	Umbilical cord milked vigorously towards umbilicus 2–3 times (estimated speed 20 cm/second)	Below or at the level of placenta and about 20cmof the placenta	Not reported	Japan	24–28 weeks	Vaginal/caesarean Caesarean section: ECC: n = 14/20 (70%) DCC: n = 14/20 (70%)	N = 2 infants in the milking group died; 1 at the 26 day due to intestinal perforation and 1 at 42

Study author and date	Definition of early cord clamping (ECC)	Definition of delayed cord clamping (DCC) or details of cord milking	Position of the baby at time of clamping in Intervention group (DCC)	Uterotonic (type/route/time)	Country	Gestational age or birth weight	Mode of birth	Additional comments
								days owing to sepsis N = 3 deaths in control group; the reason not reported
Kinmond 1993 (included in Rabe 2012)	Mean time to cord clamping 10 seconds	30 seconds after birth	20cm below the introitus	Not reported	UK (Cardiff)	27-33 weeks	All vaginal birth: n = 36	In ECC group clamping performed within 20 seconds for n = 18/19 and at 25 seconds for n = 1.
Kugelman 2007 (included in Rabe 2012)	< 10 seconds after birth	30–45 seconds after birth	20–30 cm below level of introitus at vaginal births or below level of the incision at caesarean section	No uterotonic used	Israel	< 35 weeks	Vaginal/caesarean Caesarean section: ECC: n = 23/35 DCC: n = 20/30	Intention to treat analysis. High risk pregnancy: ECC: n = 13/35 DCC: n = 9/30 Multiple pregnancy: ECC: n = 8/35 DCC: n = 9/30
March 2011	Immediately after birth		Approximately 20cm of umbilical cord was milked toward the baby immediately following birth.	Not reported	USA	24^{+0} 28^{+6} weeks	Not reported	
McDonnell 1997 (included in Rabe 2012)	Immediately after birth	30 seconds after birth	Between woman's legs	Syntocinon IV at birth	Australia	26–33 weeks.	Not reported	Intention to treat analysis (on 3 occasions the protocol was broken to allow resuscitation) 4 twin pregnancies were included, each twin

Study author and date	Definition of early cord clamping (ECC)	Definition of delayed cord clamping (DCC) or details of cord milking	Position of the baby at time of clamping in Intervention group (DCC)	Uterotonic (type/route/time)	Country	Gestational age or birth weight	Mode of birth	Additional comments
								randomised separately.
Mercer 2003 (included in Rabe 2012)	5 –10 seconds after birth	30–45 seconds after birth	10–15 inches below level of the placenta at vaginal births or below the level of the incision at caesarean section	No uterotonic used before cord clamping	USA	24^{+0}–31^{+6} weeks	Vaginal/caesarean Caesarean section: ECC: n = 6/16 DCC: n = 9/16	Intention to treat analysis (n = 2 ECC had DCC)
Mercer 2006 (included in Rabe 2012)	5-10 seconds after birth	30–45 seconds after birth	10–15 inches below level of placenta at vaginal births or below the incision at caesarean section.	No uterotonic used before cord clamping	USA	24^{+0}–31^{+6} weeks	Vaginal/caesarean	Intent to treat analysis No infant needed resuscitation
Nelle 1998 (included in Rabe 2012)	Immediately after birth	30 seconds after birth	30cm below placenta	Not reported	Germany	< 1500 g (gestational age not reported)	All caesarean section (n = 19)	
Oh 2002 (included in Rabe 2012)	< 5 seconds after birth	30–45 seconds after birth	Not reported	Not reported	USA	24 –28 weeks		Pilot study Powered to detect change in human chorionic gonadotrophin levels
Rabe 2000 (included in Rabe 2012)	20 seconds after birth	45 seconds after birth	Below the level of placenta if possible	Oxytocin IV at birth of the first shoulder	Germany	< 33 weeks Mean gestational age: ECC: n = 29.48 ± 1.96 DCC: n = 30.01 ± 1.57	Vaginal/caesarean Caesarean section: ECC: n = 19/20 DCC: n = 15/20	N = 1 baby in ECC died on 3rd day of life because of severe necrotising enterocolitis
Ranjit 2014	Immediately after birth	After 2 minutes following birth	Mothers abdomen in vaginal birth and on mother's thigh in caesarean birth	Not specified	India	30^{+0}–36^{+6} weeks	Caesarean section: ECC: n = 25/50 DCC: n = 20/45	No intention to treat analysis
Strauss 2008 (included in Rabe 2012)	"Within 2–5 seconds of birth (not exceeding 15 seconds)"	60 seconds after birth	10–12 inches below introitus at vaginal birth or beside the woman's thigh at caesarean section	Not specified	USA	30-36 weeks	Not reported	Intention to treat analysis

Study author and date	Definition of early cord clamping (ECC)	Definition of delayed cord clamping (DCC) or details of cord milking	Position of the baby at time of clamping in Intervention group (DCC)	Uterotonic (type/route/time)	Country	Gestational age or birth weight	Mode of birth	Additional comments
Ultee 2008 (included in Rabe 2012)	Within 30 seconds of birth (mean 13.4 seconds - (SD 5.6 seconds)	180 seconds after birth	On woman's abdomen in both groups	Not reported	Netherland	34^{+0} weeks to 36^{+6} weeks	All vaginal birth	All Caucasian mothers n = 2 with protocol violation were excluded, 1 in each arm

1

2

15.1.3 Evidence profile

The search strategies for this chapter can be found in Appendix E, the excluded studies in Appendix G, the evidence tables in Appendix H, and the forest plots in Appendix I.

The findings for the effect of "later" versus "earlier" cord clamping on neonatal outcomes are reported in 2 GRADE profiles. The first includes also the subgroup analysis for use of uterotonic and the second compares different strategies for increasing placental transfusion (later cord clamping vs cord milking)

Table 94: GRADE profile for comparison of later cord clamping versus earlier cord clamping - neonatal and maternal outcomes: overall and with sub-group analysis

Quality assessment							Number of babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Later cord clamping	Early cord clamping	Relative (95% CI)	Absolute	Quality
Infant death (up to discharge/variable)											
1 meta-analysis of 13 studies (Rabe 2012) and 1 study (Ranjit, 2014)	randomised trials	very serious [1]	no serious inconsistency	serious[2]	serious[3]	none	10/363 (2.8%)	22/399 (5.5%)	RR 0.51 (0.26 to 1.01)	27 fewer per 1000 (from 41 fewer to 1 more)	Very low
Infant death (up to discharge/variable) – Uterotonic used at birth (subgroup analysis)											
1 meta-analysis of 4 studies (Rabe 2012)	randomised trials	very serious [4]	no serious inconsistency	no serious indirectness	very serious[5]	none	5/77 (6.5%)	5/78 (6.4%)	RR 0.92 (0.29 to 2.95)	5 fewer per 1000 (from 46 fewer to 125 more)	Very low
Severe intraventricular haemorrhage											
1 meta-analysis of 6 studies (Rabe 2012)	randomised trials	very serious [6]	no serious inconsistency	serious[2]	very serious[5]	none	5/154 (3.2%)	7/151 (4.6%)	RR 0.68 (0.23 to 1.96)	15 fewer per 1000 (from 36 fewer to 45 more)	Very low
Severe intraventricular haemorrhage – Uterotonic used at birth (subgroup analysis)											
1 meta-analysis of 2 studies (Rabe 2012)	randomised trials	very serious [7]	no serious inconsistency	no serious indirectness	very serious[5]	none	2/42 (4.8%)	0/33 (0%)	RR 2.92 (0.15 to 56.51)	NC	Very low
Intraventricular haemorrhage (all grades)											
1 meta-analysis of 10 studies (Rabe 2012) and 1 study (Ranjit 2014)	randomised trials	very serious [8]	no serious inconsistency	serious[2]	serious[3]	none	35/304 (11.5%)	57/329 (17.3%)	RR 0.59 (0.41 to 0.84)	47 fewer per 1000 (from 18 fewer to 104 fewer)	Very low
Intraventricular haemorrhage (all grades) – Uterotonic used at birth (subgroup[e] analysis)											

Quality assessment							Number of babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Later cord clamping	Early cord clamping	Relative (95% CI)	Absolute	Quality
1 meta-analysis of 3 studies (Rabe 2012)	randomised trials	very serious [9]	no serious inconsistency	no serious indirectness	serious[3]	none	12/54 (22.2%)	14/46 (30.4%)	RR 0.61 (0.34 to 1.08)	119 fewer per 1000 (from 201 fewer to 24 more)	Very low
Ventilated for respiratory distress syndrome											
1 meta-analysis of 5 studies (Rabe 2012) and 1 study (Ranjit 2014)	randomised trials	very serious 10	no serious inconsistency	no serious indirectness	serious3	none	54/163 (27.6%)	57/196 (29.1%)	RR 0.93 (0.695 to 1.25)	123 fewer per 1000 (from 230 fewer to 230 more)	Very low
Hypoxic ischemic encephalopathy											
1 study (Ranjit 2014)	randomised trials	serious [2]	no serious inconsistency	no serious indirectness	Very serious[5]	none	0/44 (0%)	1/50 (2%)	RR 0.57 (0.05 to 6.05)	17 fewer per 1000 (from 38 fewer to 201 more)	Very low
Ventilated for respiratory distress syndrome – Uterotonic used at birth (subgroup analysis)											
1 meta-analysis of 3 studies (Rabe 2012)	randomised trials	very serious [11]	no serious inconsistency	no serious indirectness	very serious[5]	none	18/83 (21.7%)	28/107 (26.2%)	RR 0.82 (0.5 to 1.33)	47 fewer per 1000 (from 131 fewer to 86 more)	Very low
Hyperbilirubinemia (treated)											
1 meta-analysis of 3 studies (Rabe 2012)	randomised trials	very serious [12]	no serious inconsistency	no serious indirectness	serious[3]	none	51/82 (62.2%)	51/98 (52%)	RR 1.21 (0.94 to 1.55)	109 more per 1000 (from 31 fewer to 286 more)	Very low
Hyperbilirubinemia (treated) – Uterotonic used at birth (subgroup analysis)											
1 study (Rabe 2012)	randomised trials	very serious [13]	no serious inconsistency	no serious indirectness	very serious[5]	none	12/19 (63.2%)	12/20 (60%)	RR 1.05 (0.64 to 1.73)	30 more per 1000 (from 216 fewer to 438 more)	Very low
Transfused for anaemia											
1 meta-analysis of 8 studies (Rabe 2012) and 1 study (March 2011)	randomised trials	very serious [14]	no serious inconsistency	serious[2]	serious[3]	none	61/207 (29.5%)	91/223 (40.8%)	RR 0.66 (0.52 to 0.82)	139 fewer per 1000 (from 75 fewer to 196 fewer)	Very low
Transfused for anaemia – Uterotonic used at birth (subgroup analysis)											
1 meta-analysis of 2 studies (Rabe 2012)	randomised trials	very serious [15]	no serious inconsistency	no serious indirectness	serious[3]	none	13/42 (31%)	22/43 (51.2%)	RR 0.61 (0.37 to 1)	200 fewer per 1000 (from 322 fewer to 0 more)	Very low
Anaemia of prematurity											

Quality assessment							Number of babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Later cord clamping	Early cord clamping	Relative (95% CI)	Absolute	Quality
1 study (Elimian 2014)	randomised trials	very serious [16]	no serious inconsistency	no serious indirectness	serious[3]	none	36/99 (36.4%)	48/101 (47.5%)	RR 0.77 (0.55 to 1.07)	109 fewer per 1000 (from 214 fewer to 33 more)	Very low
Apgar score at 5 minutes < 8											
1 meta-analysis of 3 studies (Rabe 2012)	randomised trials	very serious [17]	no serious inconsistency	no serious indirectness	very serious[5]	none	13/72 (18.1%)	18/89 (20.2%)	RR 0.86 (0.45 to 1.62)	28 fewer per 1000 (from 111 fewer to 125 more)	Very low
Apgar score at 5 minutes < 8 – Uterotonic used at birth (subgroup analysis)											
1 meta-analysis of 2 studies (Rabe 2012)	randomised trials	very serious [18]	no serious inconsistency	no serious indirectness	very serious[5]	none	5/33 (15%)	12/44 (27%)	RR 0.75 (0.29 to 1.96)	62 fewer per 1000 (from 178 fewer to 240 more)	Very low
Haematocrit at 4 hours of life (%)											
1 meta-analysis of 5 studies (Rabe 2012)	randomised trials	very serious [19]	no serious inconsistency	no serious indirectness	no serious imprecision	none	82	91	NC	MD 5.40 higher (3.62 higher to 7.17 higher)	Low
Haematocrit at 4 hours of life (%) – Uterotonic used at birth (subgroup analysis)											
1 meta-analysis of 2 studies (Rabe 2012)	randomised trials	very serious [10]	no serious inconsistency	no serious indirectness	no serious imprecision	none	38	47	NC	MD 3.56 higher (0.35 higher to 6.77 higher)	Low
Haematocrit at 24 hours of life (%)											
1 meta-analysis of 3 studies (Rabe 2012) and 1 study (Ranjit 2014)	randomised trials	very serious [20]	no serious inconsistency	no serious indirectness	no serious imprecision	none	130	163	NC	MD 5.33 higher (3.91 higher to 6.76 higher)	Low
Haematocrit at 24 hours of life (%) – Uterotonic used at birth (subgroup analysis)											
1 study (Rabe 2012)	randomised trials	very serious [21]	no serious inconsistency	no serious indirectness	no serious imprecision	none	15	23	NC	MD 6.19 higher (1.2 higher to 11.18 higher)	Low
Maternal outcomes											
Postpartum haemorrhage											
1 study (Ranjit 2012)	randomised trials	serious [22]	no serious inconsistency	no serious indirectness	Very serious[5]	none	1/44 (2%)	1/50 (2%)	RR 1.14 (0.0 to 17.6)	3 more per 1000 (from 19 fewer to 333 more)	Very low

CI confidence interval, RR relative risk, MD mean difference, NC not calculable
1 High risk of bias for blinding in n = 4 studies, high risk of bias for incomplete outcome data in n = 3 studies and high risk of bias for allocation concealment and selective reporting in n = 2 studies. Unclear risk of bias for randomisation and allocation concealment in n = 10 studies, unclear risk of bias for blinding in n = 8 studies, unclear risk of bias for incomplete outcome data in n = 3 studies and unclear risk of bias for selective reporting data n = 9 studies
2 Immediate cord clamping was compared with milking the umbilical cord in n = 1 study
3 Confidence interval crossed 1 default MIDs
4 High risk of bias for incomplete outcome data reporting and selective reporting in n = 1 study, for blinding in n = 1 study
5 Confidence interval crossed 2 default MIDs
6 High risk of bias for blinding in n = 1 study. Unclear risk of bias for randomisation in n = 4 studies. Immediate cord clamping was compared with milking the umbilical cord in n = 1 study
7 Unclear risk of bias for randomisation, allocation concealment and blinding in n = 2 studies
8 High risk of bias for blinding in n = 4 studies. High risk of bias for selective reporting in n = 2 studies. Unclear risk of bias for randomisation in n = 7 studies. Immediate cord clamping was compared with milking the umbilical cord in n = 1 study
9 High risk of bias in blinding in n=1 study and unclear risk of bias for randomisation, allocation concealment in n = 2 studies
10 High risk of bias for blinding in n = 2 studies. Unclear risk of bias for allocation concealment and randomisation in all 5 studies, uneven number of participants present in the earlier and later cord clamping groups in n = 2 studies. 1 study was part of a multicentre trial, and the outcome reported was collected just for this subject.
11 High risk of bias for incomplete outcome data reporting and selective reporting in n = 1 study, for blinding in n = 1 study
12 High risk of bias for allocation concealment and blinding in n = 1 study. High risk of bias for incomplete data in n = 2 studies. Unclear risk of bias for randomisation in n = 2 studies.
13 Unclear risk of bias for randomisation, allocation concealment and blinding
14 High risk of bias for blinding in n = 3 studies, high risk of bias for incomplete outcome data and selective reporting in n = 1 study. Unclear risk of bias for randomisation and allocation concealment in n = 6 studies. In 1 study n = 3/36 early deaths reported in the immediate cord clamping group and this group was then excluded from the analysis as they were no longer eligible to experience outcomes.
15 Unclear risk of bias for randomisation, allocation concealment and blinding in n = 1 study and high risk of bias for blinding in n = 1 study
16 A published conference abstract with very limited data reported
17 Unclear risk of bias for blinding in n = 2 studies, unclear risk of bias for randomisation, allocation concealment and selective reporting in n = 2 studies. In 1 study 57% (n = 8) of babies allocated to delayed cord clamping had the cord clamped early, either due to cord round the neck, or need for resuscitation.
18 Unclear risk of bias for randomisation, allocation concealment and blinding in n = 2 studies
19 High risk of bias for incomplete outcome data and selective reporting in n = 1 study, for blinding n = 2 studies
20 High risk of bias for incomplete outcome data reporting and selective reporting in n = 2 studies
21 High risk of bias for incomplete outcome data reporting and selective reporting
22 No intention to treat analysis performed

Table 95: GRADE finding for comparison of later cord clamping versus earlier cord clamping on neonatal outcomes (cord milking)

Quality assessment							Number of babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Later cord clamping/more placental transfusion	Earlier cord clamping	Relative (95% CI)	Absolute	Quality
Infant death (up to discharge/variable)											
1 meta-analysis of 12 studies (Rabe 2012)	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	8/299 (2.7%)	14/329 (4.3%)	RR 0.62 (0.28 to 1.36)	16 fewer per 1000 (from 31 fewer to 15 more)	Very low

Quality assessment							Number of babies		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Later cord clamping/more placental transfusion	Earlier cord clamping	Relative (95% CI)	Absolute	Quality
Infant death (up to discharge/variable) – Cord milking[a] (subgroup analysis) [g]											
1 study (Rabe 2012)	randomised trial	serious[3]	no serious inconsistency	serious[4]	very serious[2]	none	2/20 (10%)[b]	3/20 (15%)[c]	RR 0.67 (0.12 to 3.57)	49 fewer per 1000 (from 132 fewer to 386 more)	Very low
Severe intraventricular haemorrhage											
1 meta-analysis of 5 studies (Rabe 2012)	randomised trials	very serious5	no serious inconsistency	no serious indirectness	very serious[2]	none	3/134 (2.2%)	3/131 (2.2%)	RR 0.85 (0.20 to 3.66)	18 fewer per 1000 (from 18 fewer to 61 more)	Very low
Severe intraventricular haemorrhage – Cord milking[a] -(subgroup analysis)											
1 study (Rabe 2012)	randomised trial	serious[3]	no serious inconsistency	serious[4]	very serious[2]	none	2/20 (10%)	4/20 (20%)	RR 0.50 (0.10 to 2.43)	100 fewer per 1000 (from 180 fewer to 286 more)	Very low
Transfused for anaemia											
1 meta-analysis of 6 studies (Rabe 2012)	randomised trials	serious6	no serious inconsistency	no serious indirectness	serious[7]	none	37/166 (22.2 %)	61/186 (32.7%)	RR 0.63 (0.46 to 0.87)	199 fewer per 1000 (from 70 fewer to 291 fewer)	Very low
Transfused for anaemia – Cord milking[a] (subgroup analysis)											
1 meta-analysis of 2 studies (March 2011 and Rabe 2012)	randomised trials	serious[6]	no serious inconsistency	serious[4]	serious[7]	none	24/47 (51%)	30/37 (81%)	RR 0.70 (0.53 to 0.94)	162 fewer per 1000 (from 32 fewer to 253 fewer)	Very low

CI confidence interval, RR relative risk
a Umbilical cord milked vigorously towards umbilicus 2-3 times
b One baby died at 26 days after birth due to intestinal perforation and 1 baby died at 42 days owing to sepsis
c The reason for death is not reported
1 High risk of bias for blinding in n = 5 studies, high risk of bias for incomplete outcome data in n = 2 studies and high risk of bias for allocation concealment and selective reporting n = 2 studies. Unclear risk of bias for randomisation and selective reporting in n = 9 studies, unclear risk of bias for blinding in n = 6 studies, unclear risk of bias for incomplete outcome data in n = 3 studies
2 Confidence interval crossed 1 default MIDs
3 High risk of bias in blinding
4 Immediate cord clamping was compared with milking the umbilical cord
5 Unclear risk of bias for randomisation and selection bias in n = 4 studies.

6 High risk of bias for blinding in n = 3 studies, high risk of bias for incomplete outcome data and selective reporting in n = 1 study. Unclear risk of bias for randomisation and allocation concealment in n = 6 studies. In 1 study n = 3/36 early deaths reported in the immediate cord clamping group and this group was then excluded from the analysis as they were no longer eligible to experience outcomes.
7 Confidence interval crossed 1 default MIDs
8 March 2011 is a published conference abstract with very limited data reported

15.1.4 Evidence statements

15.1.4.1 Overall analysis

Very low quality evidence from 1 SR of 15 RCTs and 1 trial (n=976) found that there was no significant difference in the risk of infant death, mechanical ventilation for respiratory distress syndrome, hyperbilirubinemia, severe intraventricular haemorrhage and Apgar score (less than 8) at 5 minutes in preterm neonates allocated to receive later cord clamping compared with those allocated to receive earlier cord clamping. However, the rate of intraventricular haemorrhage (all grades) and the numbers transfused for anaemia were significantly lower in neonates allocated to receive later cord clamping compared with those allocated to receive earlier cord clamping. Low quality evidence found that the levels of haematocrit at 4 hours and 24 first hours were significantly higher in neonates allocated to receive later cord clamping compared with those allocated to receive earlier cord clamping.

Regarding maternal outcomes only 1 study reported postpartum haemorrhage which found no significant difference in the incidence of postpartum haemorrhage in women allocated to the later cord clamping group compared with those allocated to the earlier cord clamping group. The evidence was of very low quality. No other maternal outcomes were reported.

Subgroup analysis by use of uterotonic

Findings from a meta-analysis of the 15 RCTs (n=976) showed that the administration of a uterotonic agent intravenously at birth did not change the direction of estimates of effects in most of the reported outcomes reported in the main meta-analysis. However, 2 significant results from the main analysis on the outcomes (reduction in intraventricular haemorrhage [all grades] and reduction in the number of babies needing transfusion for anaemia) lost significance at the subgroup analysis although the direction of the main effect is still beneficial for the later clamping group. The evidence was of low and very low quality.

Subgroup analysis by strategy for increasing placental transfusion (cord milking)

Very low quality evidence from a sub-group analysis by strategy for increasing placental transfusion from 2 RCTs (n=84) showed no difference in the 2 reported outcomes of infant death and severe intraventricular haemorrhage between earlier cord clamping compared with later cord clamping or earlier cord clamping compared with earlier cord clamping following cord milking. Further meta-analyses found fewer babies who had either later cord clamping or cord milking received a blood transfusion for anaemia compared with babies who had earlier cord clamping.

15.1.5 Health economics profile

The decision of when to clamp the cord has no or negligible resource implications, unless there are secondary costs associated with adverse outcomes which are affected by the timing of cord clamping. If a decision on timing improves outcomes then that decision is likely to be cost neutral or cost saving and therefore if a certain timing is clinically optimal then it will almost certainly be cost-effective too.

No search for health economic evidence was undertaken for this question as the decision on the timing of cord clamping was thought to have negligible resources implications.

Therefore this question was not identified as a priority for health economic analysis.

15.1.6 Evidence to recommendations

15.1.6.1 Relative value placed on the outcomes considered

The Committee prioritised both neonatal and maternal outcomes to base the draft of their recommendations. The maternal outcomes agreed at the protocol stage were mortality, clinical indicators (such as primary postpartum haemorrhage [PPH]) and the need for further intervention (such as blood transfusion, emergency anaesthesia), length of hospital stay and women's personal experience in terms of attachment and breast-feeding.

In terms of the neonatal outcomes, neonatal mortality, respiratory disease, brain injury, treatment for hyperbilirubinaemia with exchange transfusion, blood counts at 6 and 12 hours (haemoglobin or haematocrit) were prioritised as the most important.

The Committee considered that severe IVH (grade III/IV) may be linked with significant long term neuro-developmental impairment, including cerebral palsy (CP) in preterm infants, whereas many babies who have a mild IVH go on to develop normally or might have only minimal disabilities associated with learning, such as reduced IQ or poor concentration or mild behavioural difficulty. The Committee noted that few studies included in this review had looked at severe IVH (grade III/IV), while most of the studies reported the rate of all grades of IVH. As a result, the Committee focussed on both severe IVH and all grades of IVH as proxies for poor neonatal neurodevelopment outcomes. 'Transfused for anaemia' was considered important in relation to reducing the risks associated with blood products and donor exposure a beneficial effect of later cord clamping. It was felt that hyperbilirubinaemia may not be particularly relevant because hyperbilirubinaemia is common, usually mild, and only rarely associated with serious outcomes such as encephalopathy. Thus greater emphasis was placed on findings for 'transfused for anaemia' and IVH than hyperbilirubinaemia.

The Committee felt that infant death was a less informative outcome here because in the populations studied it was so rare that trials were unlikely to have adequate statistical power to detect any difference. In the presence of very limited evidence for maternal outcomes, Committee decision making was mainly driven by neonatal outcomes.

15.1.6.2 Consideration of clinical benefits and harms

Very low quality evidence demonstrated that there was a significantly lower rate of all grades of IVH and 'transfused for anaemia' and significantly higher haematocrit at 4 hours and 24 hours of age in the delayed cord clamping group compared to the early group. Furthermore these benefits were confirmed in the subgroup analysis based on the use of any uterotonic at birth. This was an important consideration for the Committee when drafting the recommendations, because in their experience uterotonic use is very common in clinical practice and is part of the active management of labour recommended in the NICE guideline on intrapartum care.

However, no significant difference was found between the early and late cord clamping groups for a number of neonatal outcomes such as hyperbilirubinaemia (treated), infant death, anaemia of prematurity and respiratory distress syndrome.

The Committee felt that the fact that there was no difference in Apgar scores at 5 minutes in relation to the timing of cord clamping was reassuring; indeed there was no evidence of any other harm being associated with later cord clamping.

The Committee noted that the majority of the studies defined delayed cord clamping as being between 30 to 60 seconds after birth. In some studies the cord was clamped after a longer interval (up to 180 seconds after birth). The Committee felt that in clinical practice, delayed cord clamping is generally conducted within the 30-60 second time limit and although they

felt the same benefits might be seen at other timings, decided that the recommendations should reflect the 30-60 second interval.

The Committee noted that in nearly all studies the baby was kept below the level of the placenta in order to facilitate blood flow. This practice was presumed to be beneficial, but the Committee noted that in a recent study in term babies it was found that the transfer of placental blood when cord clamping was delayed was not reduced when the baby was placed on the mother's abdomen, above her uterus. This was also evident from 3 included studies in this review where babies were held above the uterus

The Committee was aware that 2 out of 17 included studies examined the effect of 'cord milking' compared to early cord clamping. Cord milking is a technique used to increase the passage of blood along the cord to the baby so that the baby can be removed more quickly for resuscitation or respiratory support. Results from these studies showed that cord milking seems to provide a benefit similar to that of delayed cord clamping in terms of the rate of 'transfused for anaemia' and all grades of IVH.

The Committee noted that all studies were carried out in resource rich countries where the level of haemoglobin is generally higher than babies born in resource poor countries. Hyperbilirubinaemia was more prevalent among babies where cord clamping was delayed but this difference was not significant. The Committee did not feel that there was evidence to show that a slightly higher level of hyperbilirubinaemia in early life would lead to worse long term outcomes.

15.1.6.3 Consideration of health benefits and resource uses

The decision of when to clamp the cord has no or negligible resource implications, unless there are secondary costs associated with adverse outcomes which are affected by the timing of cord clamping. If a decision on timing improves outcomes then that decision is likely to be cost neutral or cost saving and therefore if a certain timing is clinically optimal then it will almost certainly be cost-effective too.

15.1.6.4 Quality of evidence

The quality of the evidence included in this section was low and very low. Two of the main reasons for the studies being downgraded were risk of bias due to lack of blinding and imprecision.

In particular, the Committee recognised some variations in the methods and populations in the included trials in terms of gestational age, different definitions of "early cord clamping" and "late cord clamping", and the nature of the active management of third stage, including the administration of a uterotonic. Timings of cord clamping also varied between the studies and this variation was taken into consideration during the Committee's discussion and interpretation of results.

15.1.6.5 Other considerations

The Committee noted that gestational age might make a difference to the care strategy, and if baby is very premature and needing immediate resuscitation early cord clamping might take precedence, yet these might be the babies with the most to gain from a larger placental transfusion. The Committee felt this decision would need to be made on an individual case-by-case basis following the clinician's judgement on the balance between benefits and harms. Similarly there might be other maternal reasons for separation between the woman and baby straight away after birth, for example in the case of severe haemorrhage. In addition, the Committee was aware that a recent trial in term babies has cast doubt on the assumption that the position of the baby in relation to the uterus is important, but noted that this has not been tested in preterm babies, so no further conclusions can be made.

15.1.7 Key conclusions

Given that there was limited evidence available in this area, the Committee did not feel confident about making strong recommendations for practice regarding the timing of cord clamping. They noted there is some evidence in favour of delayed cord clamping and no evidence of harm is associated with it. The Committee identified that an advantage of delayed cord clamping, not addressed by any of the studies in the evidence table, is that placental transfusion allows new-born infants to continue to receive oxygen via the placenta as long as the cord is pulsing. In babies born with fetal distress, it is believed that the passage of blood in the first minute can contribute to a better resuscitation. Leaving the cord intact does not necessarily preclude other actions being taken for the benefit of the baby simultaneously, for example giving oxygen. Given this, the Committee agreed that in most cases, clamping should not take place before 30 seconds from the birth of the baby, and that in situations where speed is of the essence, cord milking should be considered a reasonable alternative to delayed clamping.

15.1.8 Recommendations

59. If a preterm baby needs to be moved away from the mother for resuscitation, or there is significant maternal bleeding:

- consider milking the cord and
- clamp the cord as soon as possible.

60. Wait at least 30 seconds, but no longer than 3 minutes, before clamping the cord of preterm babies if the mother and baby are stable.

61. Position the baby at or below the level of the placenta before clamping the cord.

15.1.9 Research recommendations

Research question	7. Is there any advantage to preterm babies from delayed versus early cord clamping, or cord milking?
Why this is needed	
Importance to 'patients' or the population	Delay in cord clamping, or active cord milking, to ensure an adequate placental transfusion to the baby at the time of birth, has been shown to be beneficial (see NICE guideline on intrapartum care). Current evidence relates to term babies, but it is possible that benefit would be greater for the preterm baby, assisting transfer from the fetal circulation, and improving haemoglobin and iron stores.
Relevance to NICE guidance	Since the current guideline now recommends delayed cord clamping or cord milking (providing mother and baby are stable), the importance is low. The strength of the evidence base for the recommendation needs to be enhanced.
Relevance to the NHS	If a clear effect on improved neurodevelopmental outcome can be demonstrated, the intervention of delaying cord clamping, or cord milking, is likely to be highly cost-effective.
National priorities	NHS Outcomes Framework #1: Preventing people from dying prematurely
Current evidence base	The current evidence relates to term babies and does not look specifically at the preterm population.
Equality	This group is defined only by gestational age at delivery.

Research question	7. Is there any advantage to preterm babies from delayed versus early cord clamping, or cord milking?
Feasibility	There have been previous studies in the area so further ones would be feasible. Comparison of immediate with delayed cord clamping should allow ascertainment of any improvement in short-term stability or mortality, medium-term requirement for blood transfusion, and long-term neurodevelopment of the preterm infant. The potential difficulty in setting up such a trial is that researchers may feel the existing evidence on term babies is sufficient to extrapolate to preterm babies.
Other comments	Ascertainment of outcome could and should be masked from knowledge of treatment group, but masking of the allocated treatment from the healthcare professionals looking after the mother and baby at delivery is not feasible.

1

16 Health Economics

16.1 What is the clinical effectiveness of prophylactic progesterone (vaginal or oral) in preventing preterm labour in pregnant women considered to be at risk of preterm labour and birth?

16.1.1 Review of the literature

A search was undertaken for health economic evidence on prophylactic progesterone to prevent preterm labour in women considered to be at risk of preterm labour and birth. A total of 149 studies were identified by the search. After reviewing titles and abstracts, 5 papers were obtained. Three of these studies were excluded because they were not economic evaluations, were reporting a conference abstract or lacked the relevant comparator. Two studies were included in the literature review and are reported here (see Appendix H for Evidence Table).

A US study (Cahill 2010) used a decision analytic framework to evaluate the cost-utility of 4 strategies for the prevention of preterm labour (PTL) in women with threatened PTL.

i. Universal sonographic screening for cervical length and treatment with vaginal progesterone
ii. Cervical length screening for women with increased risk of preterm birth and treatment with vaginal progesterone
iii. Risk-based treatment with 17 α-hydroxprogesterone caproate (17-OHP-C) without screening
iv. No screening or treatment

The authors reported that the analysis was based on published evidence but the source of the evidence is unclear. The authors concluded that universal sonographic screening for cervical length and treatment with vaginal progesterone was the dominant strategy. They further reported that Monte Carlo simulation showed this to be the dominant strategy 96.9% of the time.

However, there are a number of quality concerns in the reporting of this analysis. The authors do not report values for model inputs nor sources for this data. The study has no reference list, does not give the perspective of the analysis list of references, perspective and study dates. Whilst the paper reports the results of sensitivity analysis and a “worst case” scenario the methods are barely reported. Furthermore, there is no attempt to quantify the uncertainty in any of the reported results.

A US study (Pizzi 2014) used a decision analytic approach to evaluate the cost-effectiveness of vaginal progesterone gel to a placebo for the prevention of preterm labour (PTL) in women with threatened PTL with a cervical length of 10-20 mm as measured by transvaginal ultrasound. The population was based on women in a multicentre RCT, who were pregnant with singleton pregnancies. The decision analysis included efficacy and safety data from the trial and used the cost per preterm birth averted as the measure of cost-effectiveness

The author reported a cost year of 2011 and included costs incurred until the intact infant is discharged from hospital. Based on a third party payer perspective, the author reported that vaginal progesterone was dominant with vaginal progesterone producing cost savings of USD 12,354 relative to placebo for an incremental benefit of 0.0426 births averted. A probabilistic sensitivity analysis suggested that vaginal progesterone gel dominated placebo in 79.2% of simulations.

The authors report a number of limitations with their study. In particular they note that PREGNANT had a multi-country study design and that it might not be appropriate to apply US costs to the services reported in the trial. The model had a relatively short-time horizon but as the authors note, including long term morbidity would have strengthened their reported conclusion.

16.2 What is the diagnostic accuracy of the following (alone or in combination) in women with intact membranes to identify preterm labour leading to preterm birth:

- **clinical assessment (such as symptoms expressed by women, strength and frequency of contractions, findings on vaginal examination)**
- **biochemical testing for markers for preterm labour namely cervicovaginal fetal fibronectin and IGF-BP1 insulin-like growth factor binding protein 1**
- **cervical ultrasound features (such as cervical length and funnelling)?**

16.2.1.1 Cost effectiveness of diagnostic tests to identify preterm labour

16.2.1.1.1 Introduction

Preterm labour is a common occurrence in pregnancy and is associated with adverse outcomes. However, there are interventions which can reduce the risks of adverse outcomes (maternal corticosteroids or tocolysis for example). Diagnosis of preterm labour has the potential to identify the women who would benefit from treatment whilst providing reassurance to the majority who are not in preterm labour.

The cost-effectiveness of diagnosis cannot usually be considered in the absence of treatment or management as it is the decisions that follow from a particular diagnosis that affect patient outcomes. So, for example, diagnosis would not usually be cost-effective if there was not an effective treatment for the condition being diagnosed.

Therefore, whilst this analysis focuses on the diagnostic decision it uses the output from Section16.4 on the effectiveness of tocolysis to quantify the benefit of diagnosing preterm labour.

16.2.1.1.2 Methods

A cost-utility decision analytic model was developed in Microsoft Excel® to assess different diagnostic strategies to identify preterm labour in women with suspected preterm labour and intact membranes between gestational ages of 24-34 weeks. A range of alternative diagnostic strategies to diagnose preterm labour within 48 hours were considered in the clinical review, an outcome considered important because it is related to the decision making regarding the timing of steroid and magnesium sulfate administration. In addition the strategies of no diagnostic test/no treatment and treating all women without a diagnostic test were also included as alternatives.

The evidence on the diagnostic accuracy of the various diagnostic strategies was of generally poor quality and often with serious limitations (see Chapter 9). Where there was more than 1 study reporting the diagnostic accuracy of the test it was not thought appropriate to synthesise these data and therefore data on the same diagnostic test were often conflicting. For this reason, the evaluation took the form of a “what-if” analysis. This involved calculating the cost-utility for all combinations of sensitivity and specificity between 0-100% (10,201 combinations in total) and determining what the cost-effective strategy would be for a diagnostic strategy with a certain cost at each of these different combinations – “what-if” a diagnostic strategy had this particular diagnostic accuracy, then what would the cost-effective strategy be?

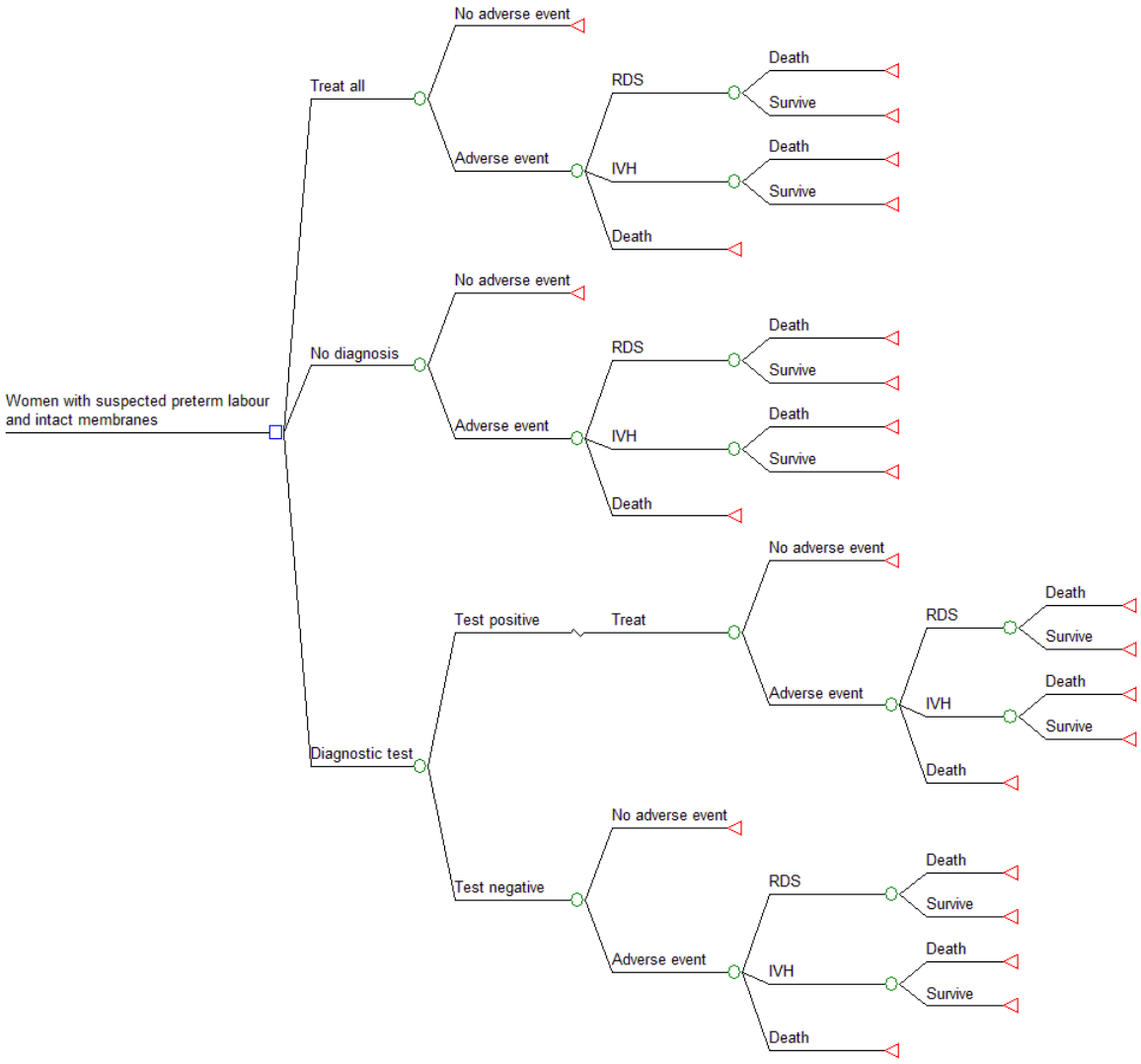
1. Treat based on the results of the diagnostic test
2. Do not perform a diagnostic test but treat all women
3. Do not perform a diagnostic test and do not treat

Where a strategy involves treatment (1 and 2) then It was assumed that women were given calcium channel blockers as a tocolytic, as that conforms with the Committee recommendation on tocolysis and what was assessed as the most cost-effective tocolytic in Section 16.4. However, the model assumed that only true positives derive the benefit of treatment. It was assumed that women not treated, either as a result of a negative test result or the strategy itself, would be sent home.

Whilst, the initial costs of diagnosis and treatment occur in the immediate term, the model takes a lifetime horizon of the baby both in terms of future costs and benefits as the outcomes assessed have lifelong consequences. This reflects the outcomes included in the treatment models, see Section 16.4.2.1.

The model did not consider the re-presentation of women with suspected preterm labour following a previous negative test result. A schematic of the model is shown in Figure 19. The model followed the approach of the model assessing the cost-effectiveness of tocolytics in women with suspected or diagnosed preterm labour by performing the analysis by gestational age.

Figure 19: Schematic of model to assess the cost-effectiveness of alternative diagnostic strategies for preterm labour



1*16.2.1.1.3 Diagnostic strategies, test accuracy and prevalence of actual preterm labour*

The various diagnostic strategies that were included in the clinical review are shown in Table 96. The diagnostic strategies evaluated based on clinical assessment (Bishop Score), transvaginal ultrasound (cervical length), fetal fibronectin and insulin-like growth factor binding protein 1 (pIGFBP-1) could be used alone or in combination.

Table 96: List of diagnostic test strategies and their reported test sensitivity and specificity

Study	Year	Diagnostic Strategy	Sensitivity	Specificity
		Treat all	100.0%	0.0%
		No diagnosis	0.0%	100.0%
Schmitz	2008	Bishop score ≥ 4	94.0%	43.0%
Schmitz	2008	Bishop score ≥ 8	35.0%	97.0%
Schreyer	1989	Bishop score 4 to 6	69.2%	73.7%

Study	Year	Diagnostic Strategy	Sensitivity	Specificity
Schmitz	2008	Bishop Score 4-7 and cervical length ≤ 20mm	60.0%	64.0%
Schmitz	2008	Bishop Score 4-7 and cervical length ≤ 25mm	80.0%	46.0%
Schmitz	2008	Bishop Score 4-7 and cervical length ≤ 30mm	90.0%	28.0%
Gomez	2005	Cervical length <30mm	88.2%	53.0%
Schmitz	2008	Cervical length <30mm	88.0%	40.0%
Tsoi	2005	Cervical length ≤10mm	81.0%	93.7%
Tsoi	2005	Cervical length ≤15mm	97.7%	84.8%
Gomez	2005	Cervical length ≤15mm	64.7%	90.4%
Bagga	2010	Cervical length ≤25mm	62.5%	89.5%
Tsoi	2005	Cervical length ≤5mm	42.9%	97.8%
Gomez	2005	Fetal fibronectin and cervical length < 15mm	41.2%	95.5%
Gomez	2005	Fetal fibronectin and cervical length < 30mm	58.8%	85.9%
LaShay	2000	Fetal fibronection test	75.0%	88.0%
Gomez	2005	Fetal fibronection test	58.8%	78.8%
Ting	2007	pIGFBP-1	100.0%	74.0%
Lembet	2002	pIGFBP-1	93.3%	81.0%
Brik	2010	pIGFBP-1	73.7%	64.9%
Kwek	2004	pIGFBP-1	66.7%	66.1%
Schmitz	2008	Selective Test (TVUS + Bishop score)	88.0%	58.0%

In order to determine the proportion of positives (true and false positives) who receive treatment and the proportion of negatives who do not receive treatment (true and false negatives) a prevalence of actual preterm labour of 10% was assumed based on the opinion of the Committee.

16.2.1.1.4 Costs

The approximate costs of the diagnostic tests are shown in Table 97. Whilst these values are used to inform the "what-if" analysis, they can also be varied themselves as part of a sensitivity analysis to determine to what extent the cost of the diagnostic strategy is an important driver of the optimal test/treat strategy.

Table 97: Diagnostic Test Costs

Test	Unit cost	Notes
Clinical assessment	£62.50	Honest (2009), based on 37.5 minutes of midwife time, see also Table 113
Ultrasound (TVUS)	£152	NHS Reference Costs 2013/14[a]
pIGFBP-1	£40	Based on 10 minutes of midwife time, see also Table 113 and £196.07 cost of HHH1206 Test kit partus (pack of 10)[b]
FFN	£70	Based on approximate cost of £45 to reflect costs associated with cassette and analyser, and 15 minutes of midwife time, see also Table 113

a. Currency code MA36Z, Obstetrics outpatient procedure
b. NHS supply chain catalogue February 2014

The costs of treatment following a positive diagnosis and the costs arising from adverse outcomes are shown in Table 112 and Table 127 respectively. The lifelong costs of adverse outcomes are discounted at an annual rate of 3.5%.

16.2.1.1.5 Baseline risk and treatment effectiveness

As noted earlier treatment effectiveness is based on the model used to assess the cost-effectiveness of tocolysis and the baseline data used in that model is described in Section 16.4.2.2 and 16.4.2.3.

However, in this model treatment effectiveness is handled in a deterministic fashion predominantly for ease of exposition in the "what-if" analysis. As noted earlier, this model assumes that women would be given calcium channel blockers as their tocolytic treatment and this model uses the mean relative treatment across the 100,000 iterations of the NMA to derive an absolute risk for each of the 3 model outcomes with treatment by gestational age. These absolute risks for these outcomes by gestational age are shown in Table 98 below.

Table 98: Absolute risk with calcium channel blockers by gestational age

Gestational age	Mortality	RDS (respiratory distress syndrome)	IVH (Intraventricular haemorrhage)
24 weeks	0.507	0.612	0.080
25 weeks	0.375	0.819	0.099
26 weeks	0.253	0.852	0.099
27 weeks	0.197	0.755	0.050
28 weeks	0.131	0.563	0.012
29 weeks	0.135	0.537	0.010
30 weeks	0.075	0.468	0.006
31 weeks	0.067	0.306	0.006
32 weeks	0.045	0.229	0.003
33 weeks	0.036	0.280	0.000
34 weeks	0.023	0.112	0.000

16.2.1.1.6 QALYs (Quality adjusted life years)

The benefits of diagnosis are based on providing the most cost-effective tocolytic treatment to true positives using the analysis described in Section 16.4. Benefits from treatment in terms of health related quality of life are derived from a potential of treatment to reduce the adverse outcomes of neo-natal/perinatal mortality, respiratory distress syndrome and intraventricular haemorrhage. The QALY loss associated with these adverse outcomes is given in Table 128 and Table 129.

A 3.5% annual discount rate is applied to QALY losses occurring in the future.

16.2.1.2 Results

In the following analyses the cost of a diagnostic test was set to £152 which was based on the cost of a transvaginal ultrasound (see Table 97). The cost-effective strategy for different diagnostic test sensitivity and diagnostic test specificity by gestational age is summarised in Figure 20 to Figure 32.

Figure 20: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 24 weeks

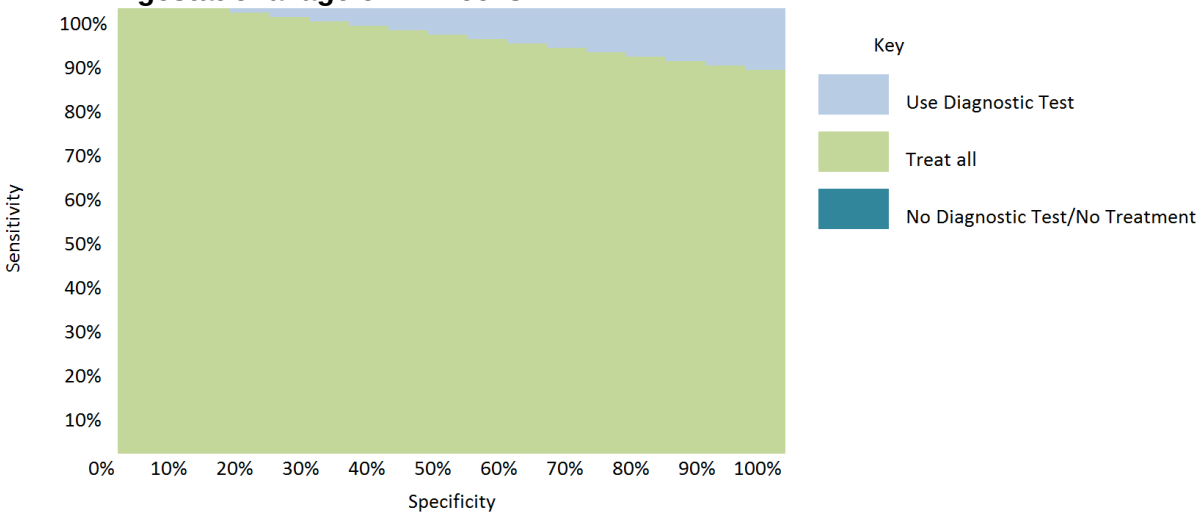
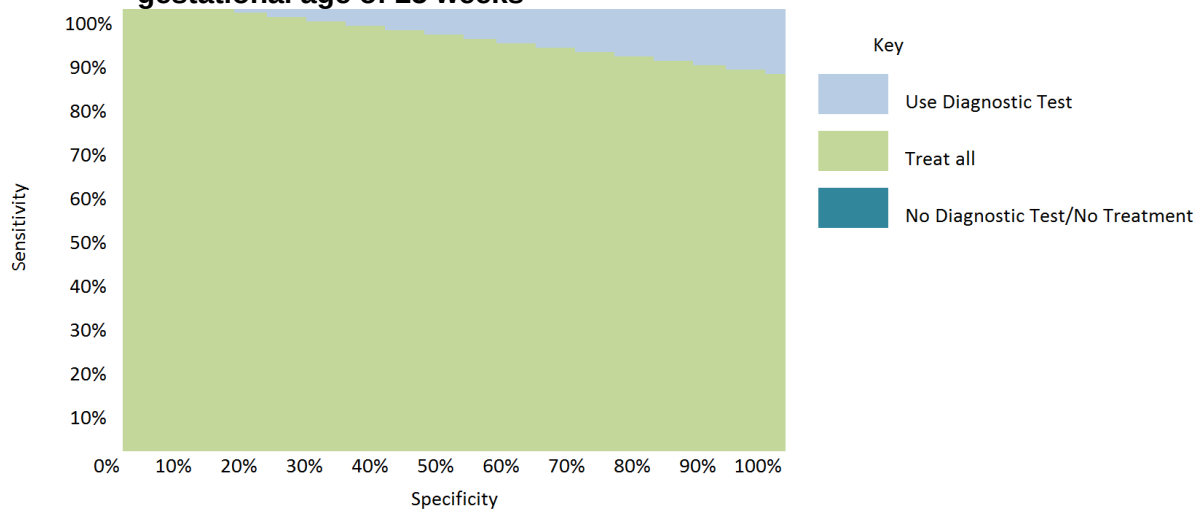


Figure 21: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 25 weeks



1

Figure 22: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 26 weeks

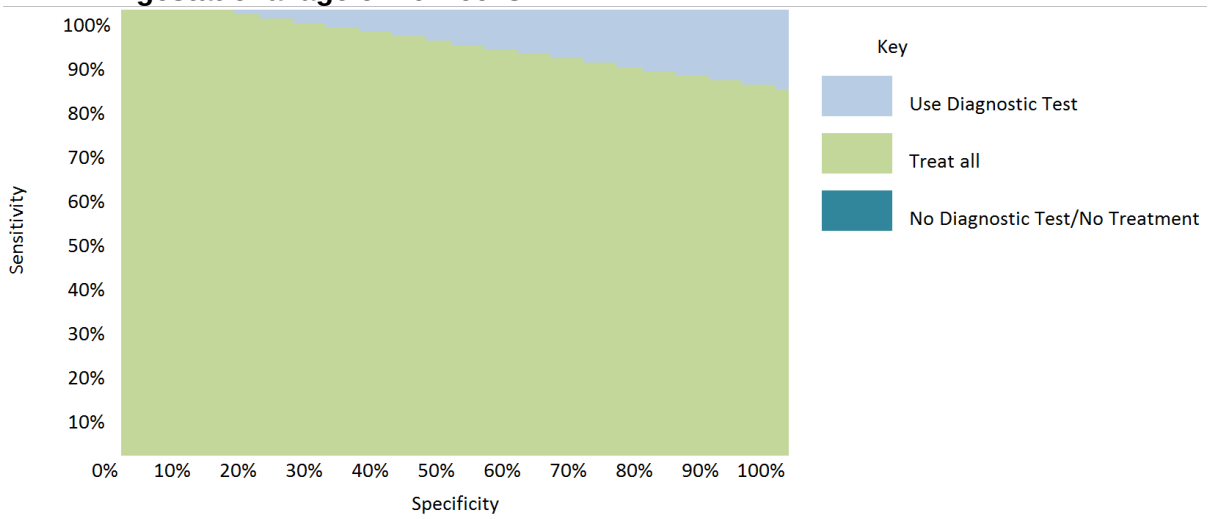
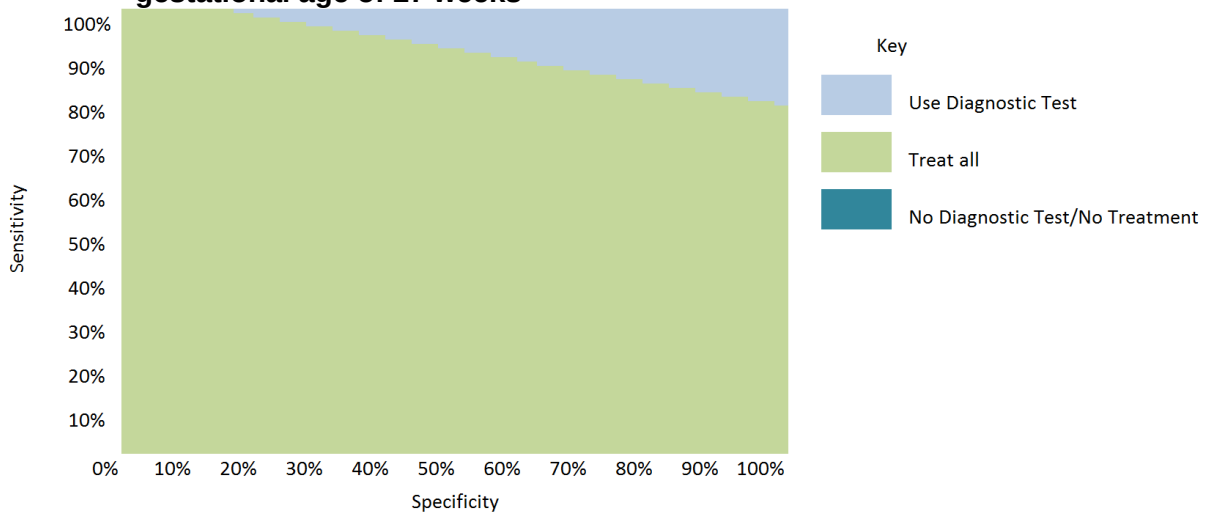


Figure 23: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 27 weeks



1

Figure 24: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 28 weeks

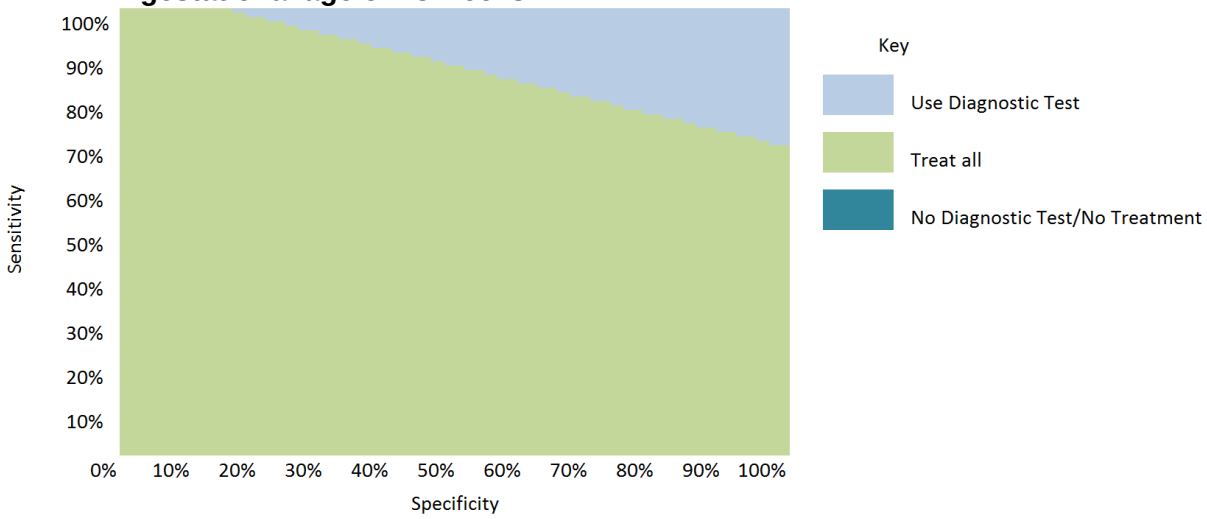


Figure 25: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 29 weeks

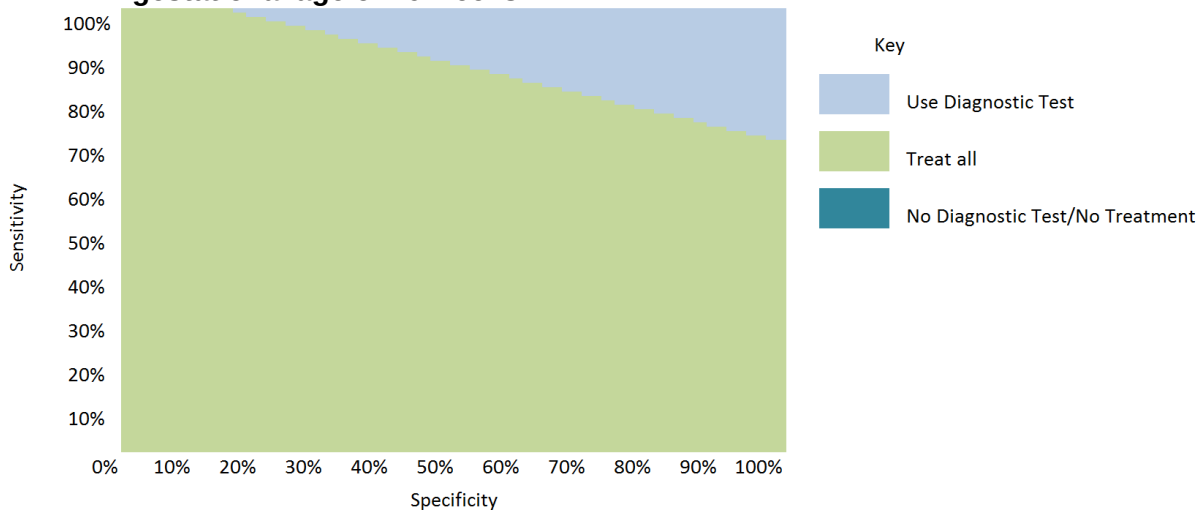


Figure 26: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 30 weeks

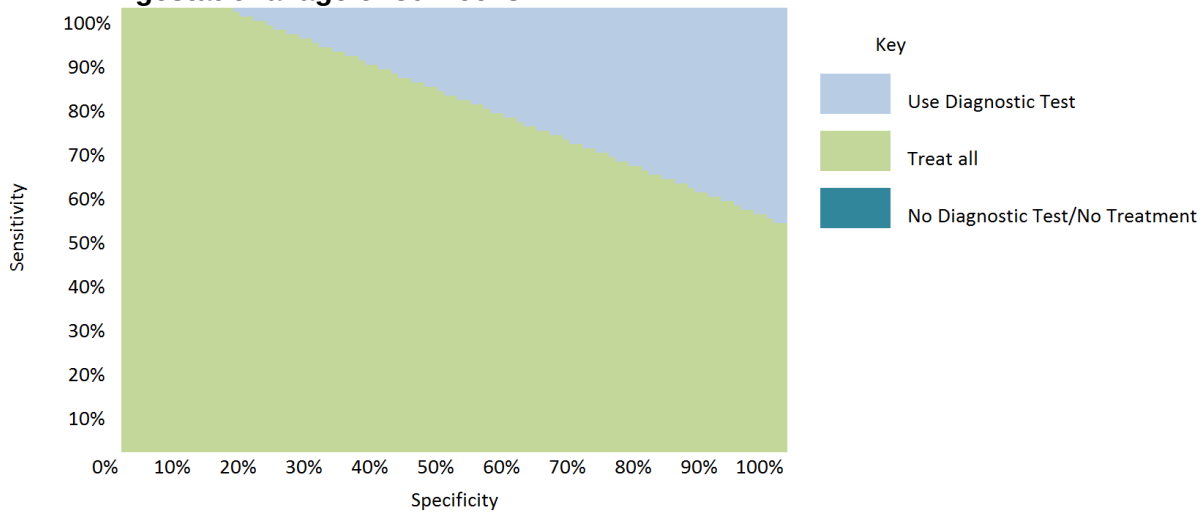


Figure 27: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 31 weeks

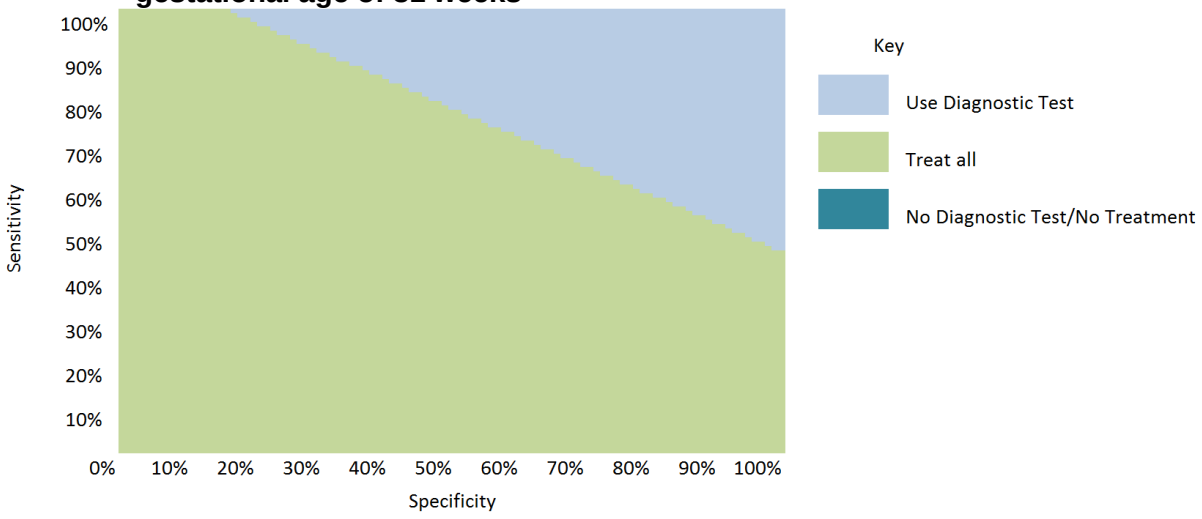


Figure 28: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 32 weeks

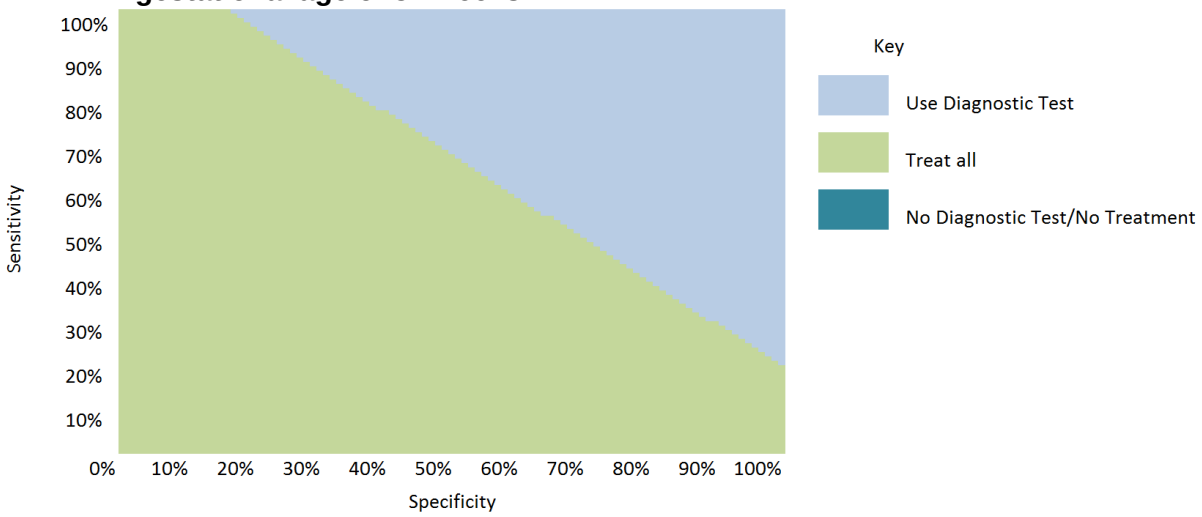


Figure 29: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 33 weeks

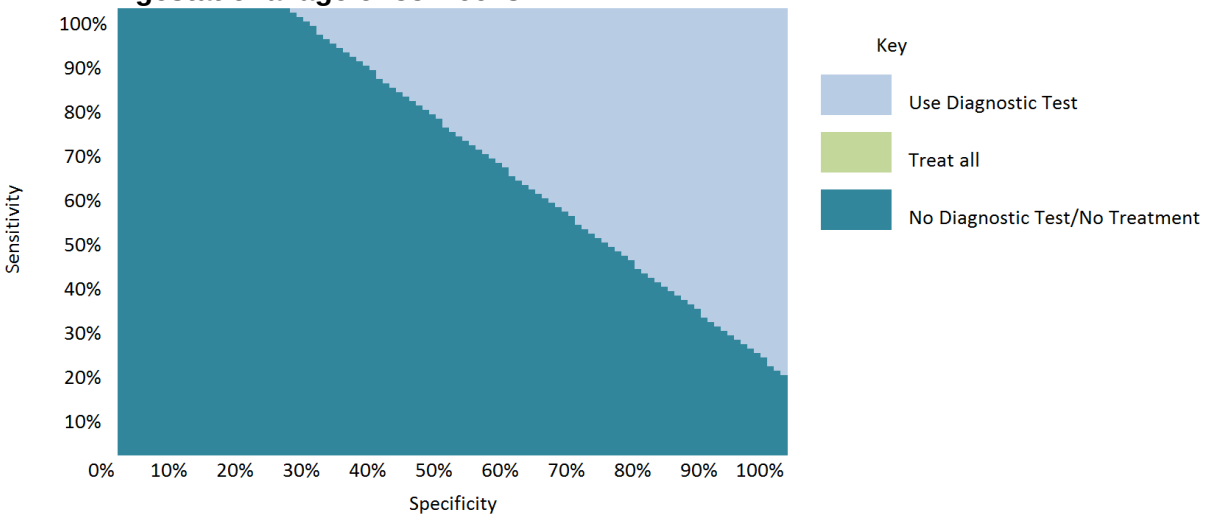


Figure 30: What-if analysis showing cost-effective strategies by test accuracy at a gestational age of 34 weeks

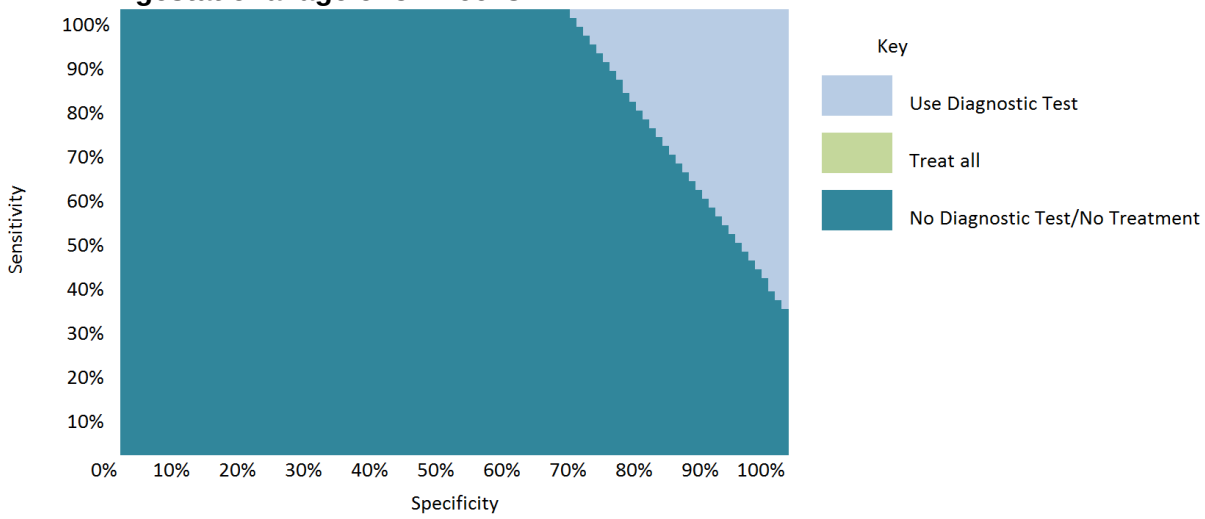


Figure 31: Summary of cost-effective strategies by diagnostic accuracy for gestational ages 24-32 weeks

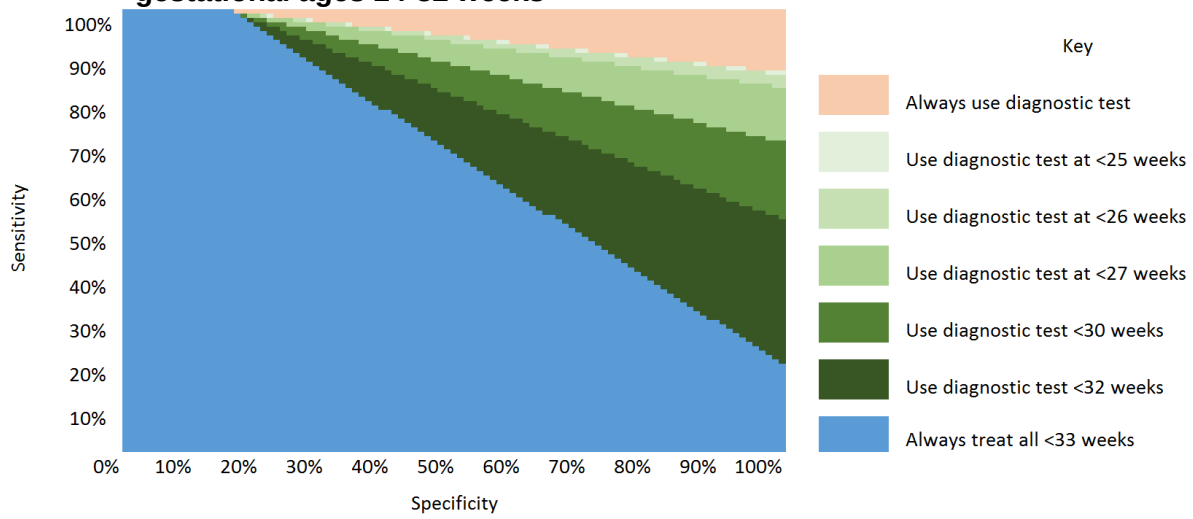
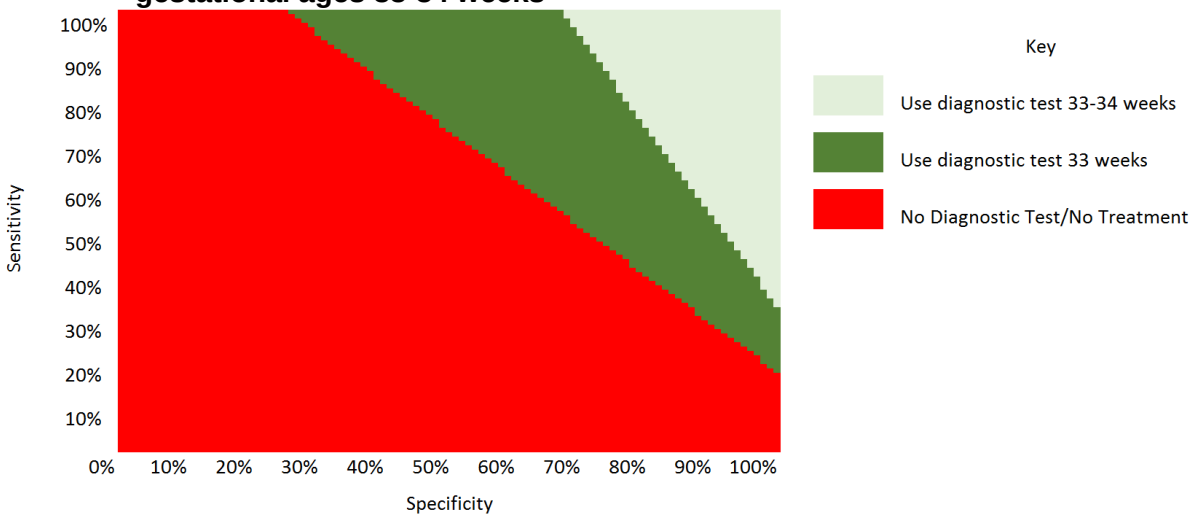


Figure 32: Summary of cost-effective strategies by diagnostic accuracy for gestational ages 33-34 weeks



Sensitivity analysis

i. Varying the cost of the diagnostic test

In this sensitivity analysis the "what-if" thresholds for cost-effectiveness were compared for a diagnostic cost of £40 versus £152 for gestational ages of 24, 30 and 34 weeks.

Figure 33: Cost-effective strategies by diagnostic accuracy varying diagnostic test cost at a gestational age of 24 weeks

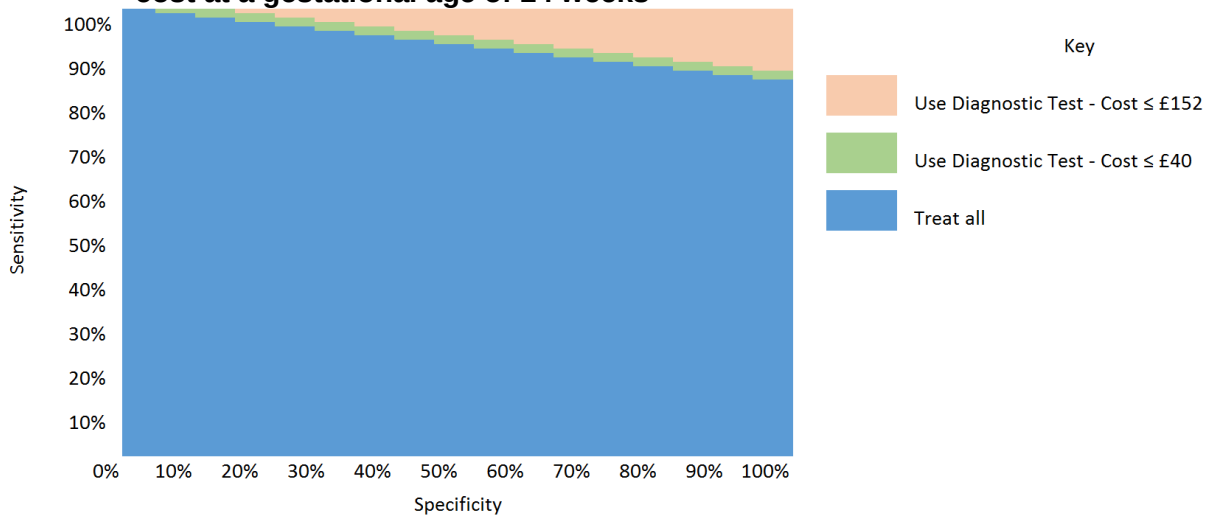


Figure 34: Cost-effective strategies by diagnostic accuracy varying diagnostic test cost at a gestational age of 30 weeks

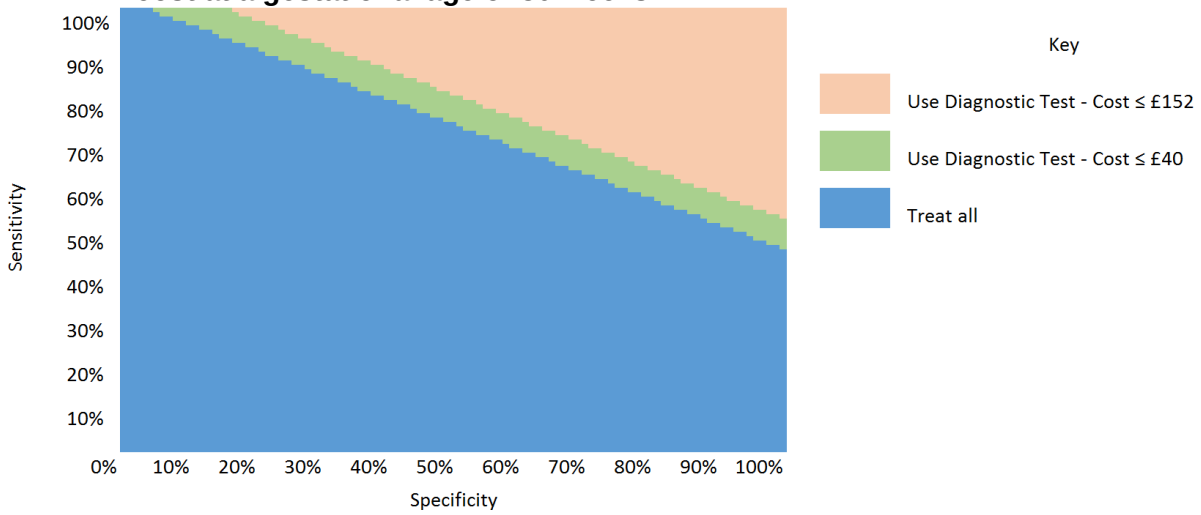
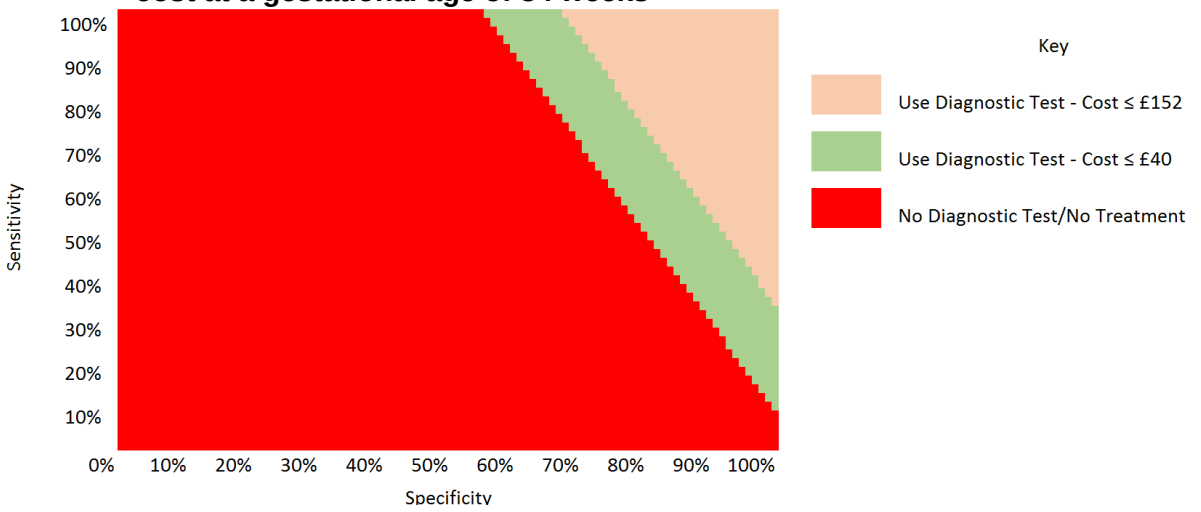


Figure 35: Cost-effective strategies by diagnostic accuracy varying diagnostic test cost at a gestational age of 34 weeks



ii. Varying the prevalence

In this sensitivity analysis the "what-if" thresholds for cost-effectiveness were compared for a prevalence of 5%, 10% and 20% for gestational ages of 24, 30 and 34 weeks and a diagnostic test cost of £152.

Figure 36: Cost-effective strategies by diagnostic accuracy varying prevalence at a gestational age of 24 weeks

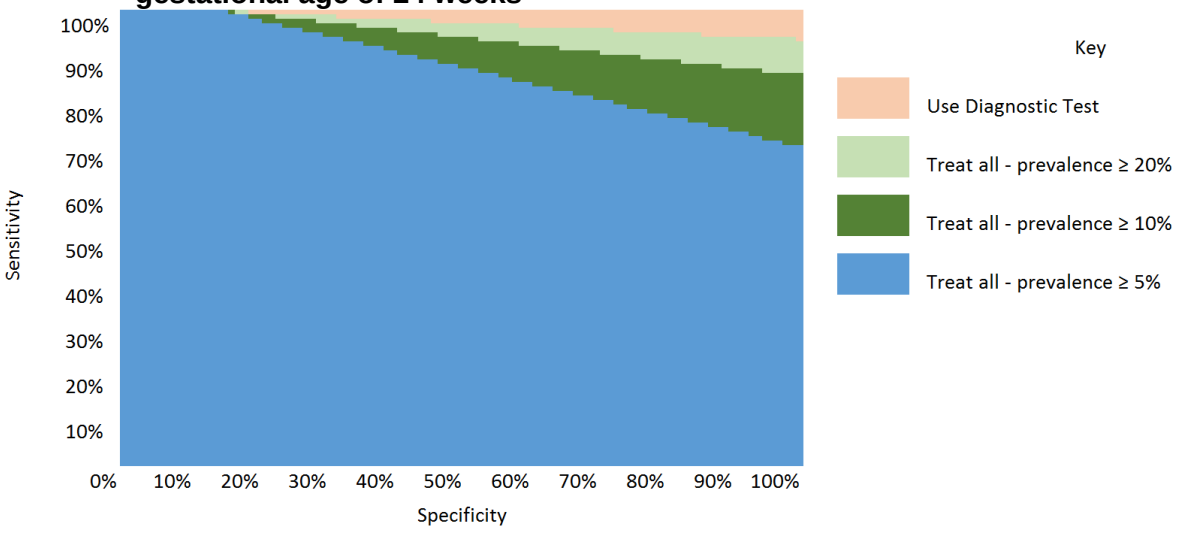


Figure 37: Cost-effective strategies by diagnostic accuracy varying prevalence at a gestational age of 30 weeks

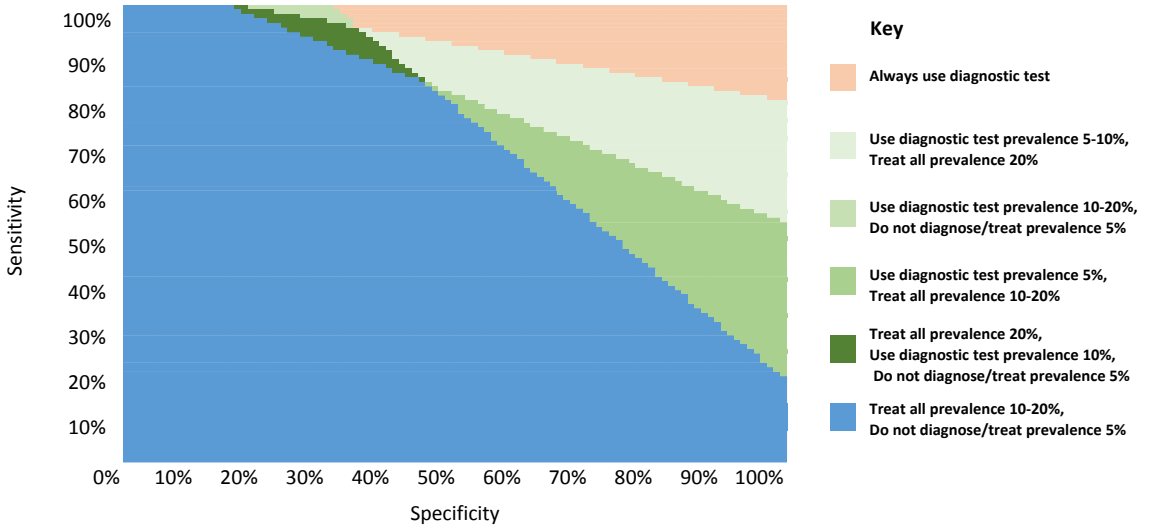
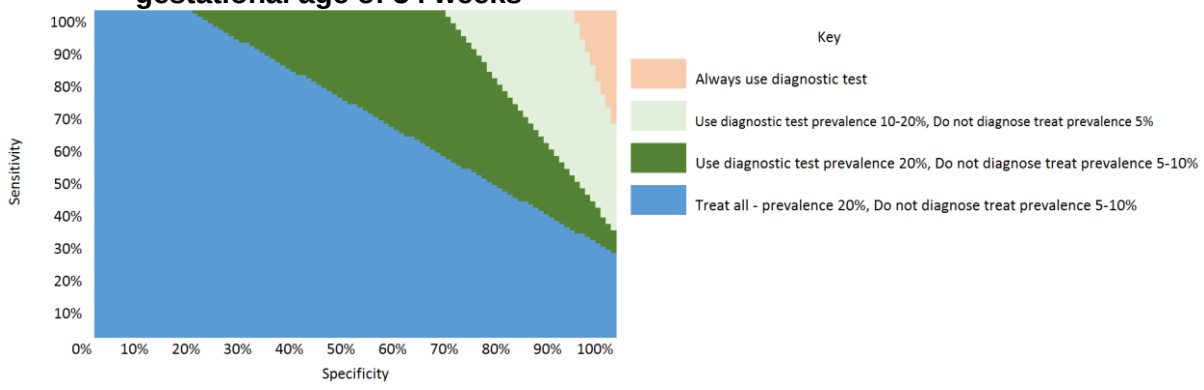


Figure 38: Cost-effective strategies by diagnostic accuracy varying prevalence at a gestational age of 34 weeks



iii. Varying the cost of false negatives

In the previous analyses it was assumed that there was no cost with a false negative, other than that associated with higher rates of adverse outcomes as a result of missing treatment.

In Figure 39 below the implication of changing from no cost of false negative to a cost of £20,000 per false negative is shown for a woman of 24 weeks gestational age. The figure of £20,000 per false negative was used as such a high value was necessary to demonstrate any effect of this input on cost-effectiveness thresholds at this gestational age.

Figure 39: Cost-effective strategies by diagnostic accuracy varying the cost of a false negative at a gestational age of 24 weeks

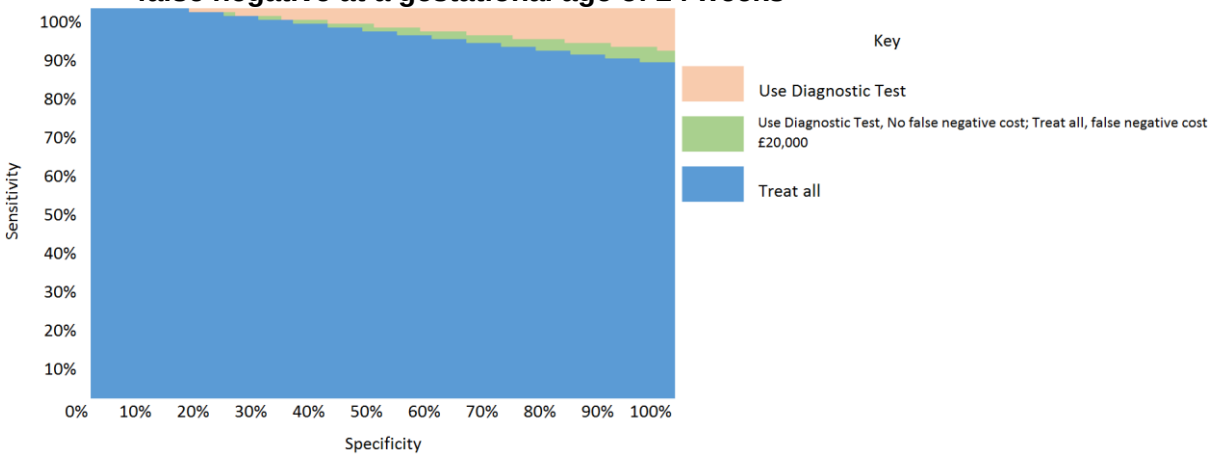


Figure 40 shows how cost-effective strategies by diagnostic accuracy change when assuming no cost of a false negative diagnosis, a £1,000 cost per false negative and a £20,000 cost of a false negative for women of a gestational age of 30 weeks

Figure 40: Cost-effective strategies by diagnostic accuracy varying the cost of a false negative at a gestational age of 30 weeks

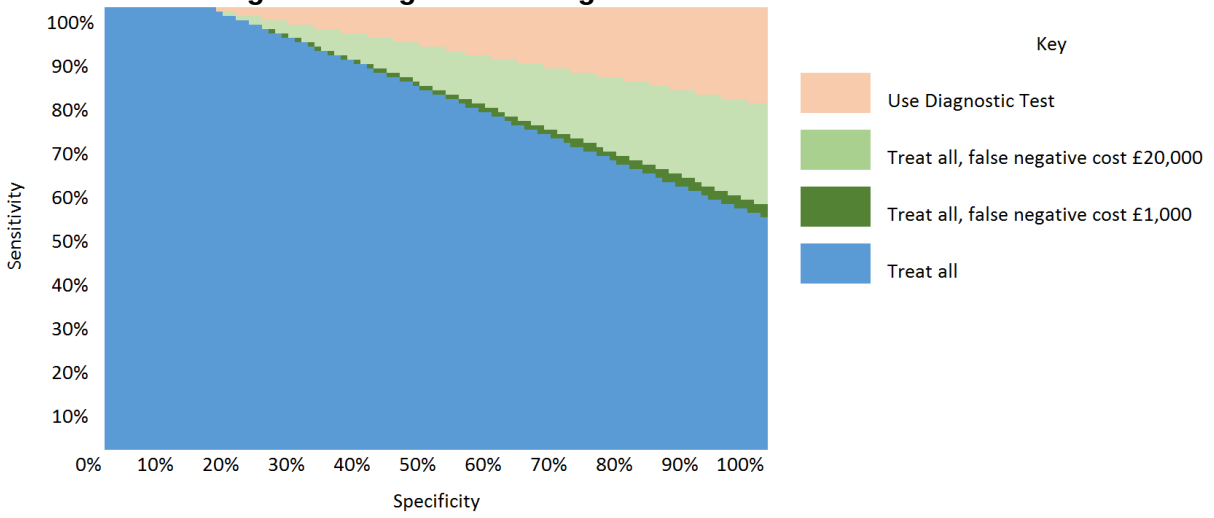
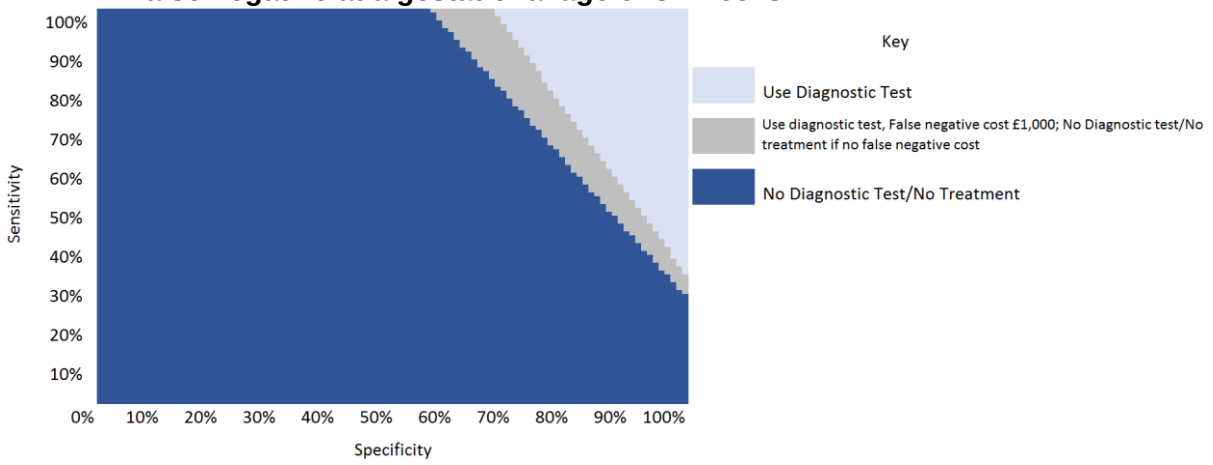


Figure 41 shows how the most cost-effective strategy by diagnostic accuracy when comparing no false negative costs with cost per false negative diagnosis of £1,000

Figure 41: Cost-effective strategies by diagnostic accuracy varying the cost of a false negative at a gestational age of 34 weeks



16.2.1.3 Discussion

The results presented in Section 16.2.1.2 do provide a cost-effective rationale for adopting a different diagnostic strategy by gestational age as summarised in Figure 31 and Figure 32. At the lower gestational age, tests require good diagnostic accuracy and particularly sensitivity in order to be preferred to a strategy of treat all, which for analytical purposes can be considered as having 100% sensitivity and 0% specificity. This is because the absolute benefits of treatment are much higher at the lower gestational ages and therefore there are greater implications of missing false negative in terms of benefits foregone. As the threshold in Figure 20 shows any percentage point reduction in sensitivity has to be compensated by a much larger percentage point increase in specificity to be considered of equivalent cost-effectiveness. In other words a relatively large reduction in false positives with their associated costs is necessary to compensate any increase in false negatives.

However, as gestational age increases the issue of false positives becomes more important as a determinant of cost-effectiveness. Less benefits are foregone by false negatives and therefore smaller reductions in false positives are needed to maintain cost-effectiveness. In Figure 28 for example, the threshold occurs where a percentage point reduction in sensitivity can be traded approximately for a percentage point increase in specificity. As a corollary of false positives becoming more important at higher gestational ages relative to false negatives, then less good diagnostic accuracy is required for treatment based on a diagnostic test to be cost-effective. This reflects the different trade-off between sensitivity and specificity at higher gestational ages. Compared with treat all, treating based on a diagnostic test can lead to a very large reduction in costs associated with false positives which can mean that the additional benefits of treating all, which are smaller in absolute terms, can no longer be achieved at acceptable cost.

Indeed at 33-34 weeks in the base case analysis, treat all was never a cost-effective strategy but rather with increasing gestational age better diagnostic accuracy was necessary in order to justify treatment. The rationale for this is that at 33-34 weeks treat all is less cost-effective than no diagnosis and treatment. However, treatment can still be cost-effective if the diagnostic test can identify sufficient true positives without too large cost in terms of false positives.

Between 32-33 weeks there is a “tipping”point between treat all and no treatment at lower levels of diagnostic accuracy (see Figure 28 and Figure 29). As diagnostic accuracy declines there is a smaller chance that a diagnostic test is preferred relative to either treat all/no treat. When the absolute benefits of treatment are relatively large then a treat all strategy is more

likely to be cost-effective. As absolute benefits of treatment declines then a no treat strategy is more likely to be cost-effective. In the model this change occurs at a gestational age of 33 weeks.

However, as Figure 38 suggests the finding that no diagnosis and no treatment is more cost-effective than treat all, is very sensitive to the prevalence of preterm birth. At 33 and 34 weeks gestational age, treat all was preferable to no diagnosis and no treatment at a prevalence of 11% and 19% respectively.

Sensitivity analysis suggested that the cost of the diagnostic test (within plausible ranges) was not an important driver of cost-effective thresholds for treat all, treat based on diagnostic test and no diagnosis and no treatment. This was especially the case at the lower gestational age where assuming an almost 4-fold increase in the cost of the diagnostic test had a negligible impact (see Figure 33). Again this is because the cost of diagnosis is relatively insignificant when compared with the losses in health related quality of life and "downstream" costs associated with false negatives. Nevertheless, the analysis did demonstrate, other things being equal, that less good diagnostic accuracy would be required for a cheaper test to be cost-effective compared with treat all (Figure 33 and Figure 34) and no treat and no diagnosis Figure 35.

As noted earlier, the sensitivity analysis did indicate that the cost-effective strategy by diagnostic accuracy was sensitive to changes in the prevalence of the preterm model, especially at the older gestational ages (Figure 36, Figure 37 and Figure 38). This is important because there is some uncertainty as to the precise prevalence in this population. The reason that the model is sensitive to assumptions about prevalence is because the importance of false negatives increased and the importance of false positives diminishes with increasing prevalence.

The base case analysis assumed that there were no additional costs with false negatives other than those arising from the "downstream" costs of adverse outcomes. However, the sensitivity analysis did not suggest that the results were sensitive to this assumption (Figure 39, Figure 40 and Figure 41). This is demonstrated by the fact that a false negative cost of £20,000 was used in order to demonstrate an impact, but this is a figure way in excess of what could be considered plausible. It is plausible that there are some costs associated with false negatives such as additional appointments with health care professionals for on-going symptoms for example and extra investigations as part of differential diagnosis, but these would typically be at least an order of magnitude less than £20,000 per patient.

As noted this analysis does provide a rationale for adopting a different approach according to gestational age. Figure 20 suggests that a test must have a sensitivity ≥ 87% (higher if specificity is ≤ 100%) at a gestational age of 24 weeks to be preferred to a strategy of treat all. There are a number of studies that report sensitivity of ≥ 87% (see Table 96) but they are either contradicted by another study of a similar test or don't have high enough specificity to make them a cost-effective option.

However, the model very clearly suggests that treat all does not remain a cost-effective option for all gestational ages. Given the limitations and quality of the diagnostic studies included it is not a straightforward matter to determine precisely the gestational age when the approach should change but there is some evidence, subject to uncertainty and limitations, from the clinical review that some tests might achieve the diagnostic accuracy to be considered cost-effective relative to treat all at 30 weeks. Figure 26 shows that the following combinations of sensitivity and specificity (or better) make treating based on a diagnostic test more cost-effective than treat all at a gestational age of 30 weeks (see Table 99). Both studies of transvaginal ultrasound using a cervical length of ≤15mm fall have diagnostic accuracy figures that are sufficient to make treatment based on a diagnostic test be considered cost-effective relative to treat all. Using transvaginal ultrasound and cervical length of ≤10 mm also has diagnostic accuracy figures that would support a recommendation when compared to treat all, but this is only based on a single study. In addition the model

suggested that treating based on a transvaginal ultrasound and cervical length of ≤ 15 mm could be considered cost effective relative to a strategy of no test and no treat at the higher gestational ages.

Table 99: Test accuracy threshold for cost-effectiveness of using a diagnostic test at a gestational age of 30 weeks

Sensitivity	Specificity
100%	17%
99%	19%
98%	21%
97%	22%
96%	24%
95%	26%
94%	28%
93%	29%
92%	31%
91%	33%
90%	35%
89%	36%
88%	38%
87%	40%
86%	42%
85%	43%
84%	45%
83%	47%
82%	49%
81%	51%
80%	52%
79%	54%
78%	56%
77%	58%
76%	59%
75%	61%
74%	63%
73%	65%
72%	66%
71%	68%
70%	70%
69%	72%
68%	74%
67%	75%
66%	77%
65%	79%
64%	81%
63%	82%
62%	84%
61%	86%

Sensitivity	Specificity
60%	88%
59%	89%
58%	91%
57%	93%
56%	95%
55%	96%
54%	98%
53%	100%

There are a number of considerations that need to be taken into account when the results of the above health economic analysis are interpreted. First, whilst the analysis took a standard incremental approach with respect to broad categories of treat all, treat based on a diagnostic test, or do no treat do not diagnose it did not do so with respect to different combinations of sensitivity and specificity that would be considered cost effective relative to treat all and do not diagnose or treat. Nor within the context of the what-if analysis did it do this for specific diagnostic accuracy data provided by the included studies. This was because it was thought that this diagnostic accuracy data had such severe limitations that the broader what-if approach would be more useful. However, the uncertainty inherent in the reported diagnostic accuracy evidence adds another level of uncertainty with respect to what may considered the most cost-effective diagnostic test or combinations of test.

Secondly the analysis departed from the NICE reference case by not including a probabilistic sensitivity analysis. However, it is important to note that this model took into account the interdependence of treatments and diagnosis in determining cost-effectiveness and the treatment effect size was derived from the mean relative treatment effect of a treatment found to be cost-effective in a probabilistic sensitivity analysis. It would have been possible to sample the relative treatment effect from the iterations produced for the NMA but if treatment was considered cost-effective based on the mean relative treatment effect it is almost certain that a probabilistic sensitivity analysis would confirm this finding, albeit with some quantification of the uncertainty. Furthermore, this what-if analysis is by construction deterministic with respect to sensitivity and specificity and yet the uncertainty surrounding the diagnostic accuracy of the various strategies is more difficult to quantify than the uncertainty surrounding treatment. Extensive 1-way sensitivity analysis was undertaken to investigate the important drivers of cost-effectiveness (including sensitivity, specificity, gestational age, prevalence, diagnostic and cost of false negatives).

Finally, in assessing the benefits of diagnosis the model assumes that false negatives miss the benefits of treatment and experience the baseline risk of various outcomes. However, this reflects a worst case scenario and at least a proportion of women sent home as negatives are likely to re-present in sufficient time to still benefit from treatment.

16.2.1.4 Conclusion

This what-if analysis provides strong evidence that treatment is cost-effective even when considering the costs of identifying the women suitable for treatment. It also provides evidence that the most cost-effective diagnostic strategy varies with gestational age. At earlier gestational ages when the absolute risks are high then treating all women with suspected preterm labour and intact membranes can be cost-effective even when allowing for the fact that 90% of those treated might not derive any treatment benefit.

The model also suggest that treatment can remain cost-effective at higher gestational ages when absolute risks are lower providing a diagnostic test can be applied with sufficiently good diagnostic accuracy.

Whilst a change in diagnostic strategy according to gestational age is indicated by this analysis the gestational age at which this change should take place is difficult to precisely identify given the uncertainty with respect to the precise diagnostic accuracy of the various tests. Nevertheless, Figure 26 and Figure 37 suggest that at 30 weeks and beyond may be reasonable gestational age at which to require treatment to be guided by a positive diagnostic test, and thereby reduce inconvenience to women and costs to the health service when absolute risks are relatively low. At 30 weeks there is some suggestion from the diagnostic studies reviewed that transvaginal ultrasound using a cervical length of ≤15mm could have sufficient diagnostic accuracy to be considered cost-effective relative to treat all, do not diagnose do not treat or other diagnostic tests or combinations of tests which do not have a cost-effective sensitivity/specificity combination.

16.3 What is the clinical and cost effectiveness of magnesium sulfate given to women at high risk of giving birth preterm (defined as those suspected to be in preterm labour or diagnosed as being in preterm labour and those having planned preterm birth) for preventing cerebral palsy and other neurological disorders in babies born at different preterm gestations?

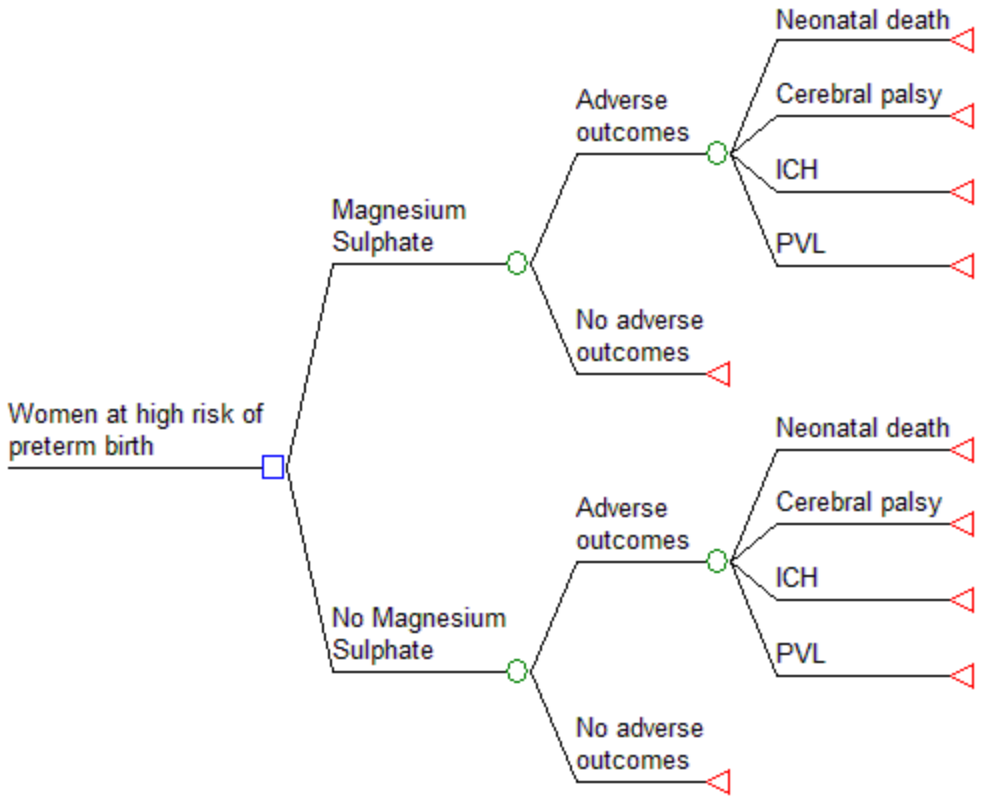
16.3.1 Introduction

Preterm labour carries a high risk of neonatal mortality and morbidity which can result in large on-going costs for the health care system, as a result of cerebral palsy for example. There has been increasing research interest in recent years in magnesium sulfate as a treatment that may offer some degree of neuroprotection for preterm birth. Clearly, there is potential for any treatment that reduces the adverse consequences of preterm birth to be cost-effective given the substantial losses in health related quality of life and healthcare costs resulting from these adverse consequences. A small published literature (see 16.3.2 **Error! Reference source not found.**) suggested that magnesium sulfate is a cost-effective treatment for neuroprotection but these analyses were performed outside a UK setting and therefore it was thought that it would be useful to develop a de novo model utilising the clinical review that was undertaken for this guideline.

16.3.2 Methods

A decision analytic model was developed in Microsoft Excel® to assess the cost-effectiveness of magnesium sulfate given to women for neuroprotection between 24+0 and 32+0 weeks of pregnancy and at high risk of preterm birth. A schematic of the model is shown below in Figure 42.

Figure 42: Schematic of the model for the use of magnesium sulfate for neuroprotection



16.3.2.1 Model probabilities and treatment effect size

A clinical review undertaken for the guideline assessed the following outcomes:

- stillbirth
- neonatal mortality before discharge
- neonatal/paediatric mortality between discharge and follow-up
- total perinatal, neonatal and paediatric mortality*
- grade III or IV intracranial haemorrhage (ICH)*
- periventricular leukomalacia (PVL)*
- any cerebral palsy
- moderate or severe cerebral palsy at 2 years*
- gross motor dysfunction at 2 years
- any developmental delay at 2 years
- cognitive dysfunction at 2 years
- blindness at 2 years
- deafness at 2 years
- maternal death
- any maternal adverse effects
- maternal adverse effects leading to stopping of infusion

- maternal cardiac or respiratory arrest
- drop in maternal blood pressure of more than 15mmHg
- maternal hypotension

Only the outcomes marked with an asterisk were included in the model. Priority was given to outcomes that have the most important impact on health related quality of life and/or where a substantial saving to the NHS would result from an averted case. With respect to the outcomes included in the model, a lifetime approach was considered with respect to costs and health related quality of life.

Clearly there would be double counting if more than 1 neonatal/paediatric mortality outcome was included and therefore the most comprehensive measure of mortality was chosen.

Moderate or severe palsy was chosen over any cerebral palsy as mild cerebral palsy, which will explain most of the difference, has a much smaller impact on future health service costs and health related quality of life.

The outcomes relating to motor function, cognitive development and any developmental delay were not included as, although important, it was thought their impact and costs would occur predominantly outside of the health domain (e.g. the education sector). Furthermore, their exclusion from the model was unlikely to have an important bearing on the overall cost-effectiveness as treatment effects were small and not statistically significant.

Outcomes relating to deafness, blindness, maternal mortality, maternal cardiac or respiratory arrest and maternal hypotension were excluded as the number of events was negligible and any differences fell a long way short of statistical significance.

The effect of treatment on maternal blood pressure of more than 15mmHg was not included in the analysis as it was felt to be a more intermediate marker of outcomes of interest and unlikely to have an important independent effect on health related quality of life and costs.

Broad measures of maternal adverse effects were not included as it was thought they most likely would contribute to treatment failure and/or would have only a very short term effect on health related quality of life.

Meta analyses undertaken for the clinical review were used to estimate the baseline risk and the treatment effect size as shown in Table 100 and Table 101 respectively. In addition to the point estimate used in deterministic analysis the tables also show the parameters that are used for probabilistic sensitivity analysis.

Table 100: Baseline risks for model of magnesium sulfate for neuroprotection

Outcome	Risks	Distribution	Alpha[a]	Beta[a]	Source
Neonatal/paediatric mortality	10.8%	Beta	242	2,001	Guideline meta-analysis
Cerebral palsy	3.4%	Beta	59	1,656	Guideline meta-analysis
ICH	5.0%	Beta	90	1,709	Guideline meta-analysis
PVL	2.7%	Beta	48	1,751	Guideline meta-analysis

a. The Alpha parameter is given by the number of events in the controls in the meta- analysis. The Beta parameter is the number without events in the controls

Table 101: Treatment effect size for model of magnesium sulfate for neuroprotection

Outcome	Relative risk	Distribution	Mu[a]	Sigma[a]	Source
Neonatal/paediatric mortality	0.95	Log-normal	-0.044	0.088	Guideline meta-analysis

Outcome	Relative risk	Distribution	Mu[a]	Sigma[a]	Source
Cerebral palsy	0.61	Log-normal	-0.490	0.211	Guideline meta-analysis
ICH	0.81	Log-normal	-0.189	0.154	Guideline meta-analysis
PVL	0.94	Log-normal	-0.053	0.206	Guideline meta-analysis

a. Mu is calculated as the natural log of the relative risk and sigma is calculated as the standard error of the log of the relative risk

16.3.2.2 Costing and resource use

In accordance with the NICE Guidelines Manual (NICE, 2012) costing was undertaken from the perspective of the NHS and personal social services. Costs are based on 2015 prices unless otherwise stated.

Discounting of costs was not exactly as per the NICE Reference Case. No discounting is necessary for treatment as this all occurs at the start of the intervention. However, the outcomes evaluated are often lifetime in their impact and their costs should be discounted. However, the paper used to estimate these costs used a discount rate of 5% rather than the 3.5% discount rate suggested by NICE (Kruse 2009). It was not possible to re-calculate the costs using a 3.5% discount rate as a temporal breakdown of the overall lifetime cost was not provided.

16.3.2.2.1 Treatment cost

The cost of magnesium sulfate are based on an initial 4g IV bolus followed by 1g per hour IV infusion thereafter for a period of 24 hours. The pharmaceutical component of treatment costs are shown in Table 102.

Table 102: Magnesium sulfate costs

Dose	Per unit cost	Quantity	Total	Source
20mL (4-g) amp	£16.98	1	£16.98	BNF (March 2015)
2mL (1-g) amp	£1.15	24	£27.60	BNF (March 2015)
Total	-	-	£44.58	

It addition it was assumed that the woman would require ante-natal monitoring for the whole 24 hour period for which she was on treatment and this was based on NHS Reference costs, see Table 103. So the total treatment cost was £1,081

Table 103: Ante-natal observation

Currency Code	Currency Description	Cost	Source
NZ16Z	Ante-natal routine observation	£1,036	NHS Reference Costs 2013/14

16.3.2.2.2 Downstream costs

In addition to the costs of the intervention it is also important to compare the alternatives in terms of their impact of costs that arise subsequent to the intervention decision which are attributable either to the intervention itself, such as adverse events, or to preterm birth. Table 104 details the costs associated with model outcomes.

Table 104: Outcome related costs

Outcome	Cost	Distribution	SE	Source
Perinatal, neonatal or paediatric mortality	£1,480	Normal	£159	NHS Reference Costs 2013/14[a]

Outcome	Cost	Distribution	SE	Source
Cerebral palsy	£74,608[b]	Deterministic	-	Kruse 2009
Intracranial haemorrhage	£22,382[c]	Deterministic	-	Kruse 2009
Periventricular leukomalacia	£74,608[d]	Deterministic	-	Kruse 2009

a. XB03Z Paediatric critical care, advanced critical care 3
b. Kruse 2009 estimated in year 2000 prices that the lifetime health care costs for cerebral palsy using an annual discount rate of 5% was €66,155 for men and €65,288 for women. The mid-point of this estimate was used and converted into GBP using an exchange rate of £0.737 = €1 (http://www.exchangerates.org.uk/ - accessed 03/04/2015). It was then converted into 2013/14 prices using the HCHS (The Hospital & Community Health Services) Index
c. It was assumed that Grade III and Grade IV ICH would be similar in cost to cerebral palsy. Alvarez 1994 suggest that 30% of ICH is of severity Grade III and Grade IV and therefore the cost of ICH was estimated as 0.3 x £74,608
d. It was assumed that the costs of PVL were the same as the costs for cerebral palsy

16.3.2.3 QALYs

A lifetime QALY loss was assigned to each of the 4 outcomes assessed in the model as shown in Table 105. In the case of a neonatal or paediatric death it was assumed that this would result in a loss of 80 years of life based on current life expectancy in the United Kingdom. It was assumed that all 80 years would have be lived with a health state utility of 0.82 based on UK population norms (Kind, 1983). An annual discount rate of 3.5% was applied for each year lived in full health in accordance with NICE methods.

The health state utility for moderate to severe cerebral palsy was taken from the literature (Cahill 2011) with a value of 0.55. A life expectancy of 60 years was assumed and the total discounted QALY for that life expectancy calculated assuming there were no additional co-morbidities. This was then subtracted from the discounted QALY of an individual who lived 80 years of life with a health state utility of 0.82 in order to estimate the overall lifetime QALY loss associated with moderate to severe cerebral palsy.

It was assumed that the QALY loss from PVL would be the same as for moderate to severe cerebral palsy and that the QALY loss from ICH would be one-third of that from cerebral palsy.

Table 105: QALY losses associated with adverse model outcomes

Outcome	QALY loss
Perinatal, neonatal or paediatric mortality	22.70
Cerebral palsy	8.50
Intracranial haemorrhage	2.80
Periventricular leukomalacia	8.50

16.3.2.4 Sensitivity analysis

Probabilistic sensitivity analysis, one-way and multi-way sensitivity analyses were undertaken to assess the robustness of the model results given uncertainty surrounding various model inputs.

16.3.3 Results

Deterministic base case results are presented in Table 106, Figure 43, Figure 44 and Figure 45. This analysis suggests that magnesium sulfate for neuroprotection is dominant, being

cheaper and more effective than no magnesium sulfate for neuroprotection. It is cheaper because the savings from a reduction in adverse outcomes more than offsets the cost of treatment.

Table 106: Incremental costs and QALYs of magnesium sulfate for neuroprotection

Outcome	Incremental costs	Incremental QALYs
Treatment	£1,081	N/A
Neonatal/paediatric mortality	-£7	0.11
Cerebral palsy	-£995	0.11
Intercranial haemorrhage	-£193	0.02
PVL	-£102	0.01
Total	-£215	0.26

(a) <Insert Note here>

Figure 43: Deterministic base case analysis of magnesium sulfate for neuroprotection shown on a cost-effectiveness plane

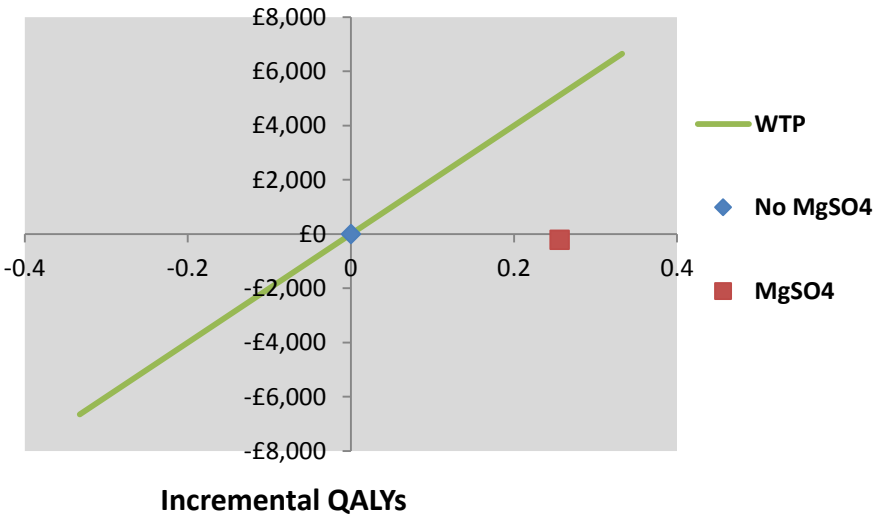


Figure 44: Breakdown of base case analysis of costs for magnesium sulfate for neuroprotection versus no magnesium sulfate for neuroprotection

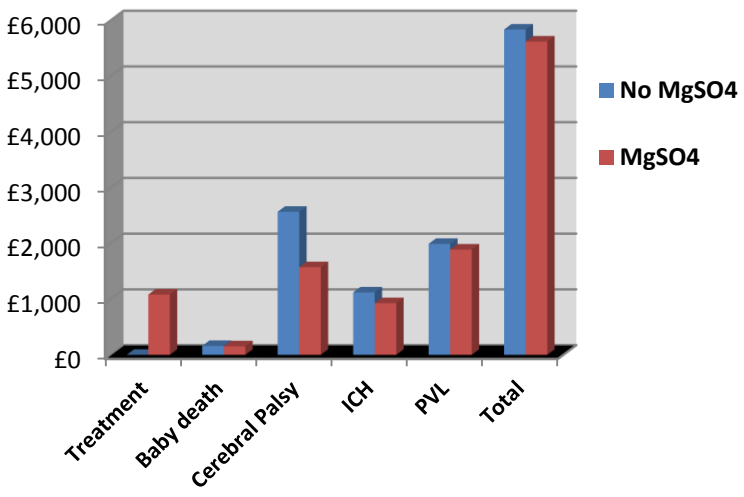


Figure 45: Breakdown of base case analysis of QALY losses for magnesium sulfate for neuroprotection versus no magnesium sulfate for neuroprotection

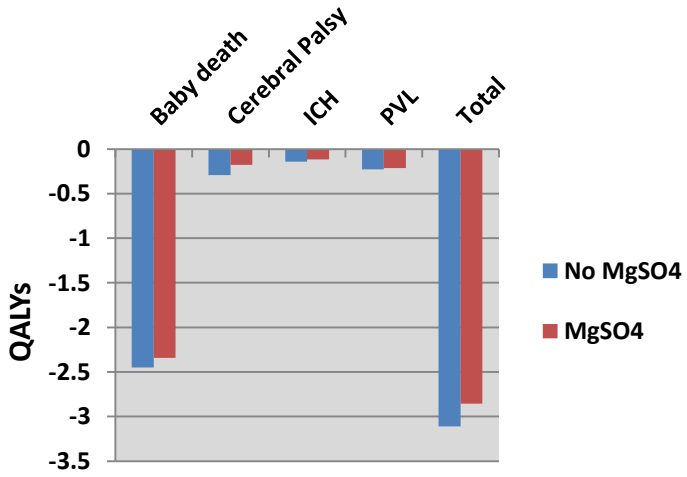
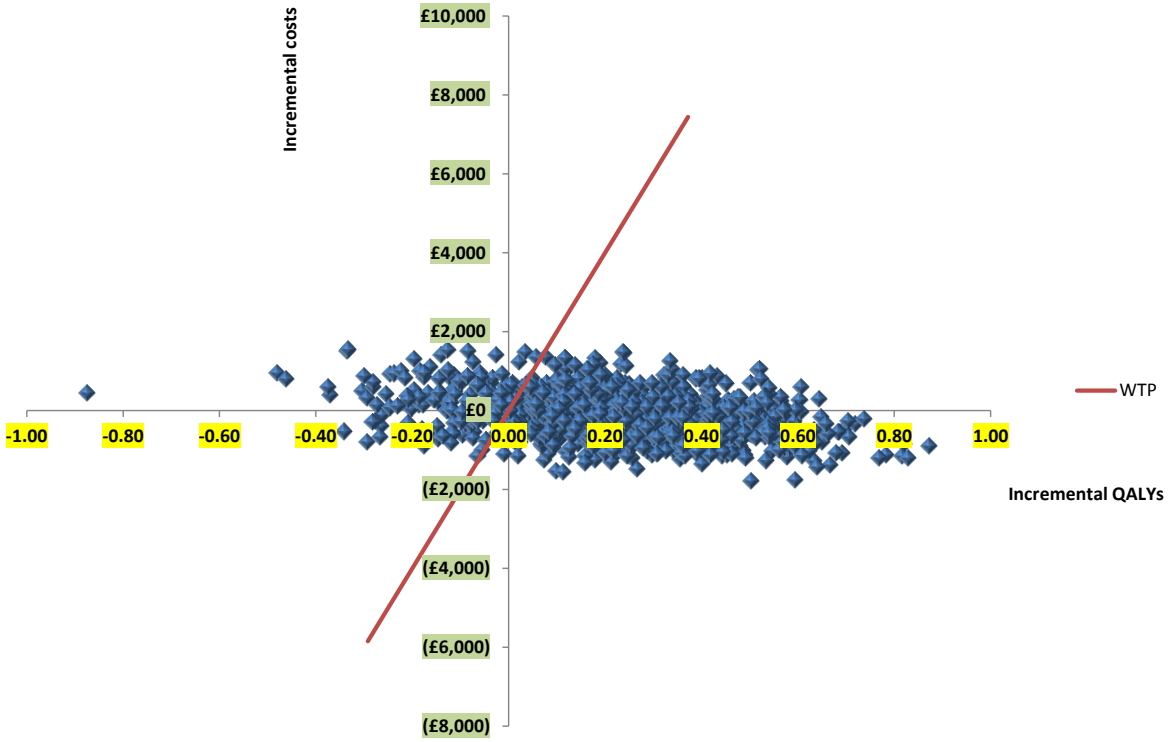


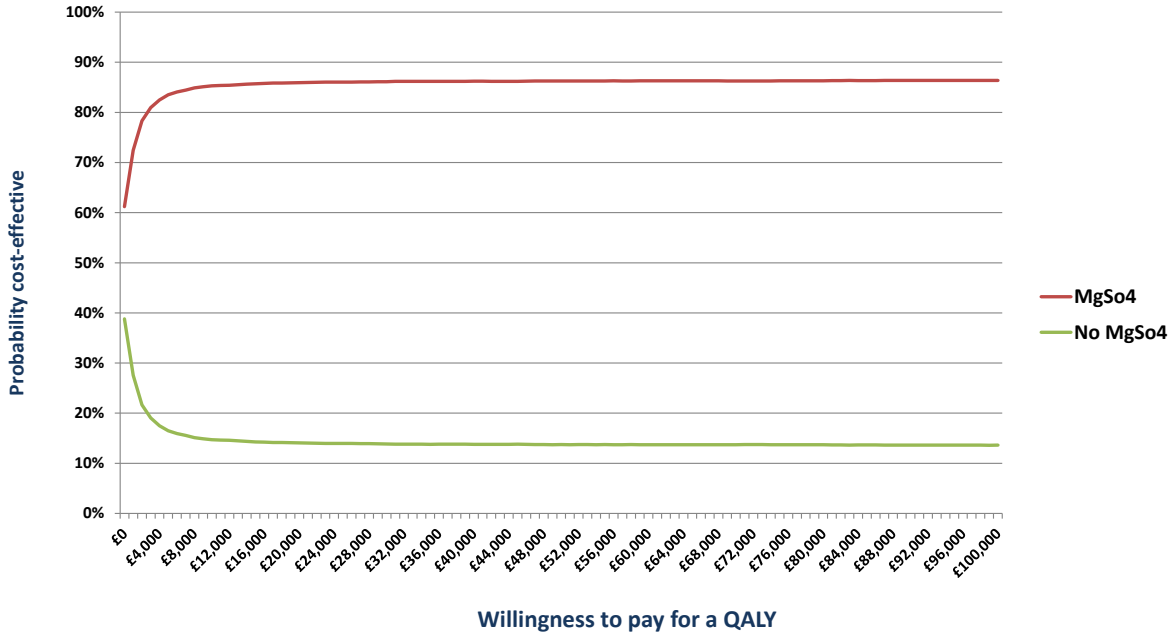
Figure 46 shows results from the probabilistic sensitivity analysis of 10,000 Monte Carlo simulations plotted on a cost-effectiveness plane. This suggests that the probability magnesium sulfate is cost-effective for neuroprotection compared to not giving magnesium sulfate for neuroprotection is 85.9% using a £20,000 willingness to pay for a QALY decision threshold. Across the 10,000 simulations magnesium sulfate for neuroprotection had a net mean benefit of £4,900 when compared to a strategy of no magnesium sulfate for neuroprotection.

Figure 46: Base case probabilistic sensitivity analysis of magnesium sulfate for neuroprotection versus no magnesium sulfate for neuroprotection



The cost-effectiveness acceptability curve for the base case probabilistic sensitivity analysis is shown in Figure 47. Across all willingness to pay thresholds magnesium sulfate has the highest probability of being cost-effective. This is so even when the decision maker is not willing to pay anything for a QALY and this is because magnesium sulfate for neuroprotection was the cheapest strategy for approximately 60% of the simulations as a result of savings from reduced adverse outcomes more than offsetting treatment outcomes.

Figure 47: Cost-effectiveness acceptability curve for base case probabilistic sensitivity analysis



16.3.3.1 Sensitivity analysis

The base case analysis strongly suggests that magnesium sulfate for neuroprotection is cost-effective. Below are presented a number of sensitivity analyses to test how robust that conclusion is with respect to changes in model parameters. Clearly lowering the treatment cost and/or increasing the QALY associated with adverse outcomes would only re-inforce the base case conclusion. Therefore, these sensitivity analyses and intended to explore the thresholds at which magnesium sulfate for neuroprotection would cease to be cost-effective. If those input values required to achieve those thresholds are not thought plausible then confidence in the base case result can be strengthened.

16.3.3.1.1 Increasing the treatment cost

The treatment cost would have to be increased to £6,413 for magnesium sulfate for neuroprotection to be no longer considered cost-effective at a £20,000 willingness to pay for a QALY according to the deterministic analysis. In a probabilistic sensitivity analysis of 10,000 Monte Carlo simulations at this treatment cost and using a £20,000 willingness to pay threshold the results indicated that magnesium sulfate would have a 47.0% probability of being the most cost-effective treatment option and had a net mean benefit of -£513 across the 10,000 simulations. This result is displayed in Figure 48 and the associated cost-effectiveness acceptability curve in Figure 49.

Figure 48: Monte Carlo simulation of magnesium sulfate for neuroprotection assuming a treatment cost of £6,413 and a willingness to pay of £20,000 per QALY

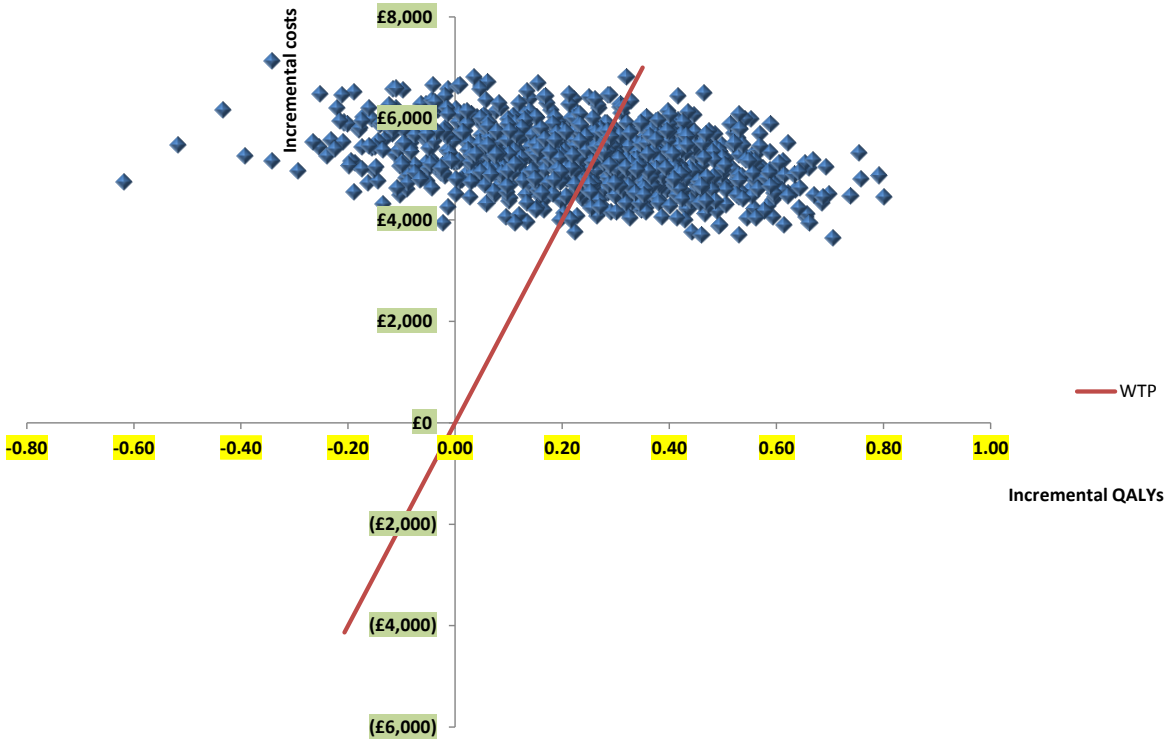
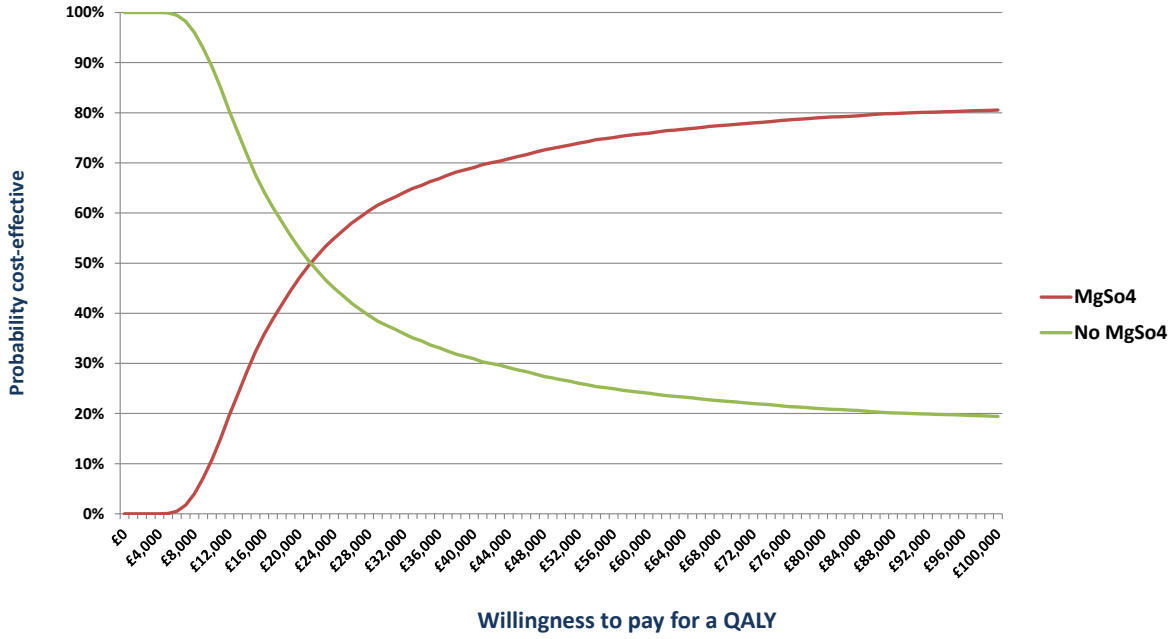


Figure 49: Cost-effectiveness acceptability curve for magnesium sulfate for neuroprotection assuming a treatment cost of £6,413



If a £30,000 per QALY threshold was used to assess cost-effectiveness then the treatment cost would have to exceed £8,971. A probabilistic sensitivity analysis of 10,000 Monte Carlo

simulations at this treatment cost and £30,000 willingness to pay threshold suggested that magnesium sulfate would have a 47.1% probability of being the most cost-effective treatment option and had a net mean benefit of -£698 across the 10,000 simulations. This result is displayed in Figure 50 and the associated cost-effectiveness acceptability curve in Figure 51.

Figure 50: Monte Carlo simulation of magnesium sulfate for neuroprotection assuming a treatment cost of £8,971 and a willingness to pay of £30,000 per QALY

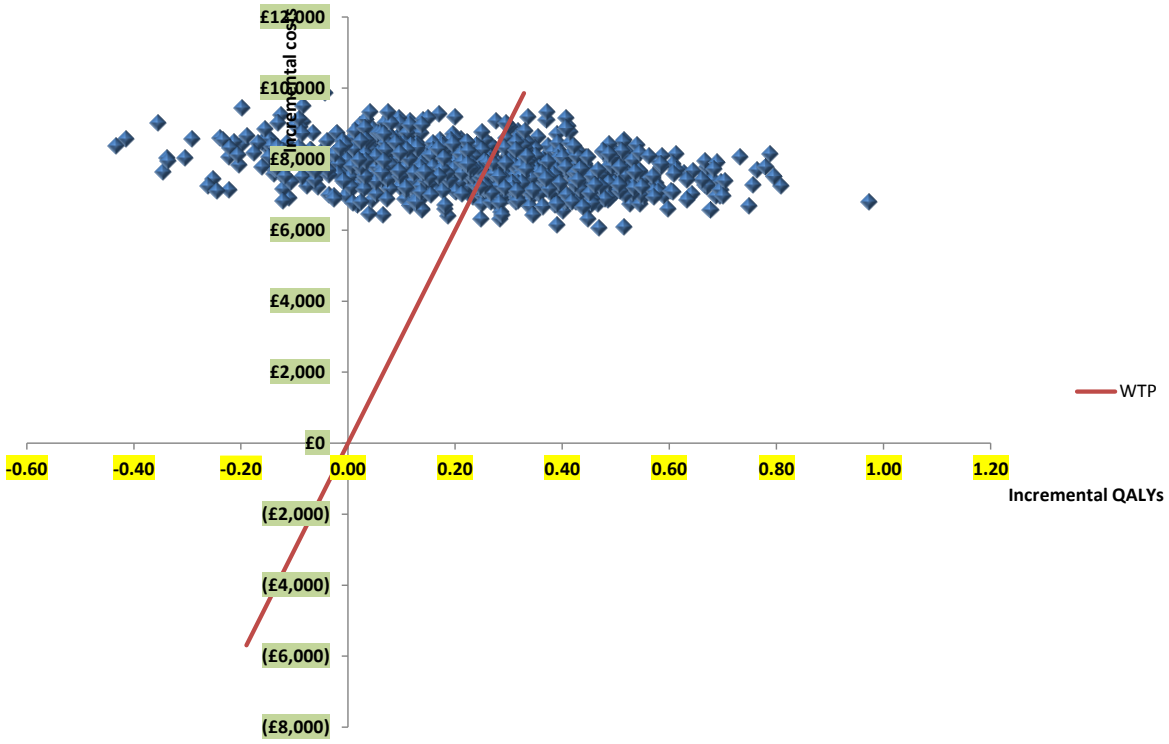
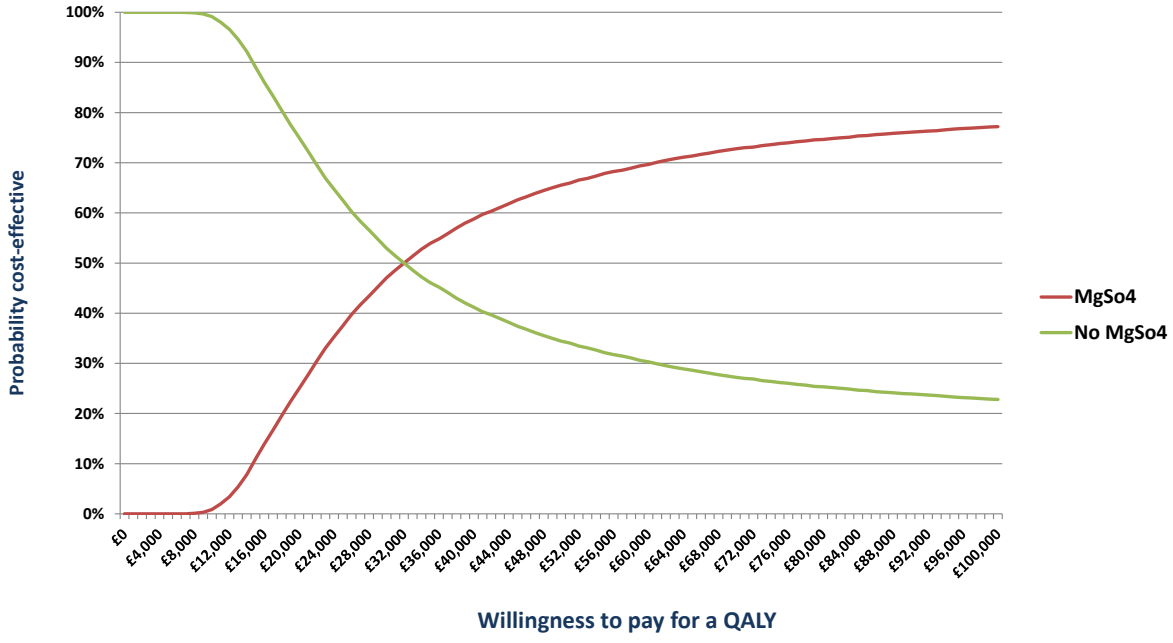


Figure 51: Cost-effectiveness acceptability curve for magnesium sulfate for neuroprotection assuming a treatment cost of £8,971



1*16.3.3.1.2 Reducing the QALY loss from adverse outcomes*

There is considerable uncertainty around the health state utility loss associated with adverse outcomes and they are treated deterministically even in the probabilistic sensitivity analysis. In this analysis we restrict the QALY loss to that arising from mortality and assume in this analysis that this is 11 QALYs, about half what was used in the base case analysis.

A probabilistic sensitivity analysis of 10,000 Monte Carlo simulation found that there was a 70.9% probability of magnesium sulfate for neuroprotection being cost-effective with these assumption about reduced QALY loss from adverse outcomes while keeping all other model inputs at their base case values. Across the 10,000 simulations the net mean benefit of magnesium sulfate was £1,075. The plot of these 10,000 simulations is shown in Figure 52 and the associated cost-effectiveness analysis acceptability curve is shown in Figure 53.

Figure 52: Monte Carlo simulation of magnesium sulfate for neuroprotection assuming that a QALY loss only arises from mortality and that this loss is only half the value assumed in the base case analysis.

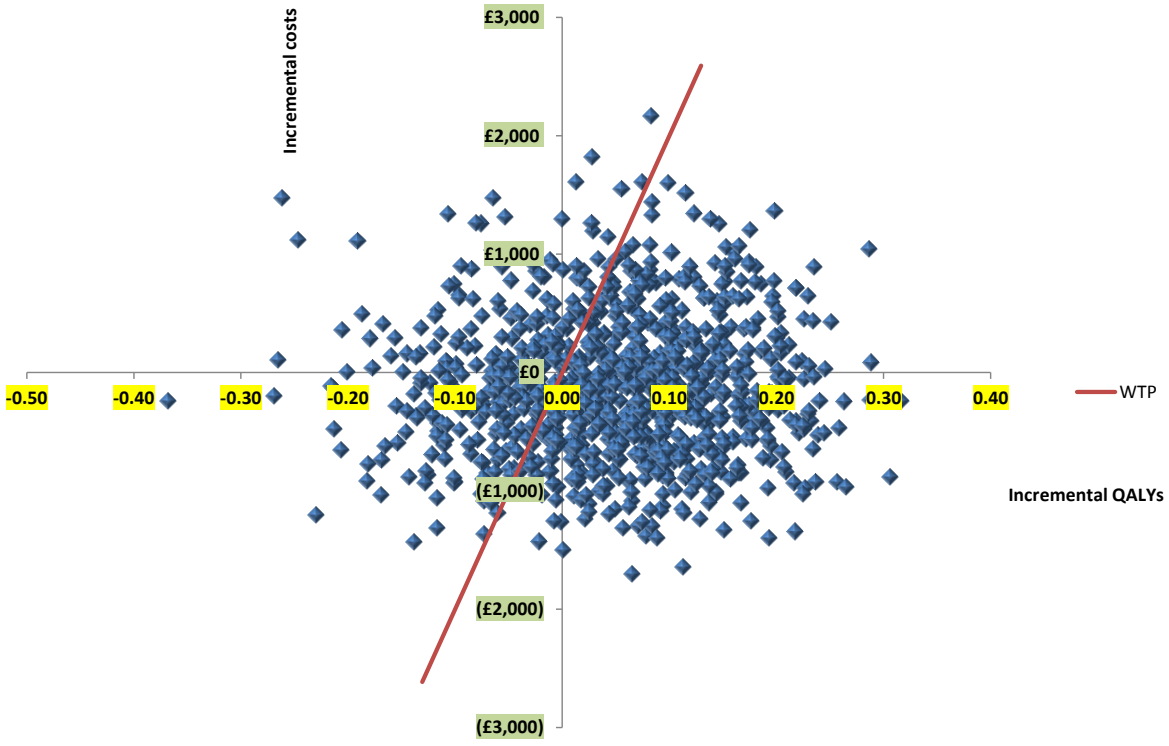
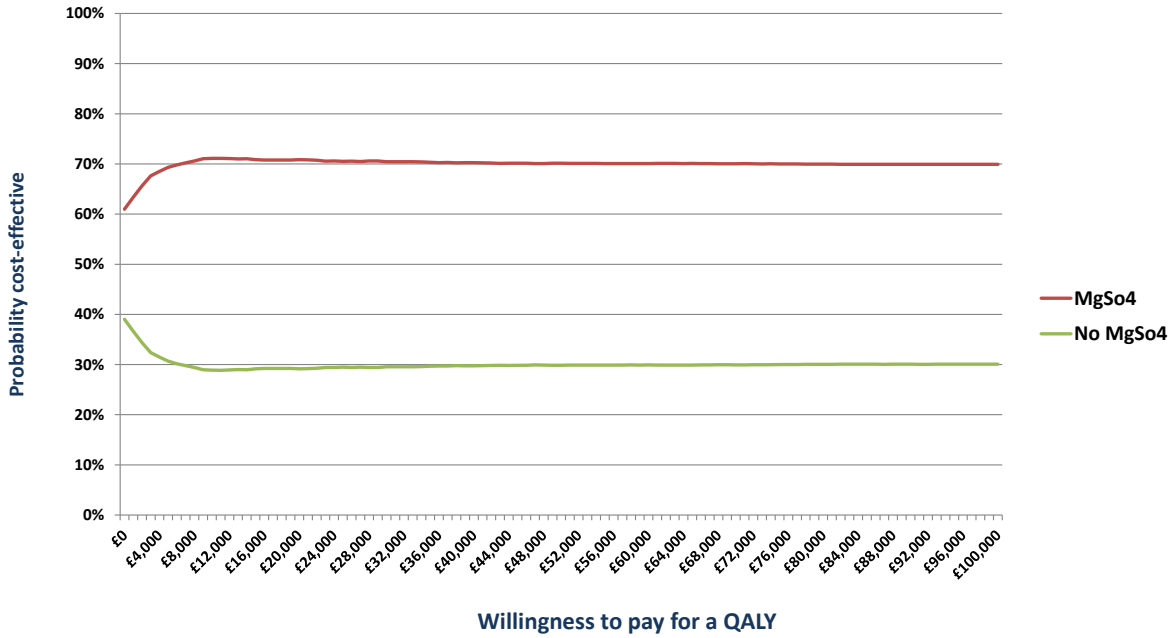


Figure 53: Cost-effectiveness acceptability curve magnesium sulfate for neuroprotection assuming that a QALY loss only arises from mortality and that this loss is only half the value assumed in the base case analysis.



1*16.3.3.1.3 Reducing the cost of adverse outcomes*

In this sensitivity analysis the cost of all adverse outcomes are removed and the treatment cost is increased until magnesium sulfate is no longer cost-effective in the deterministic analysis at a willingness to pay of £20,000 and £30,000 per QALY. This occurs at treatment costs of £5,117 and £7,675 respectively. Probabilistic sensitivity analysis for these 2 scenarios is shown in Figure 54 and Figure 56 with their respective cost-effectiveness acceptability curves depicted in Figure 55 and Figure 57.

The probability of magnesium sulfate being cost-effective in the Monte Carlo simulation with the lower treatment cost and willingness to pay for a QALY was 48% with a net mean benefit of -£346. In the Monte Carlo simulation with a higher treatment cost and willingness to pay for a QALY the probability of magnesium sulfate given for neuroprotection being cost-effective was also 48% with a net mean benefit of -£597 across the 10,000 simulations.

Figure 54: Monte Carlo simulation of magnesium sulfate for neuroprotection assuming no costs from adverse outcomes and a treatment cost of £5,117 and a willingness to pay of £20,000 per QALY

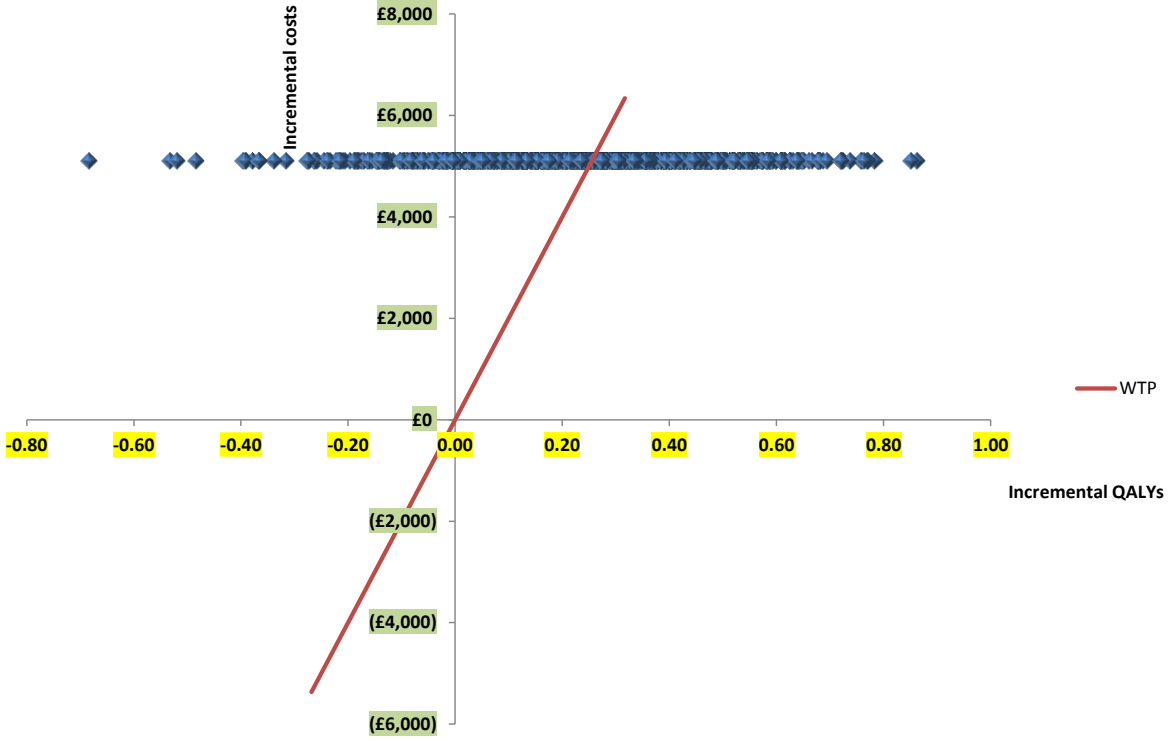


Figure 55: Cost-effectiveness acceptability curve magnesium sulfate for neuroprotection no costs from adverse outcomes and a treatment cost of £5,117

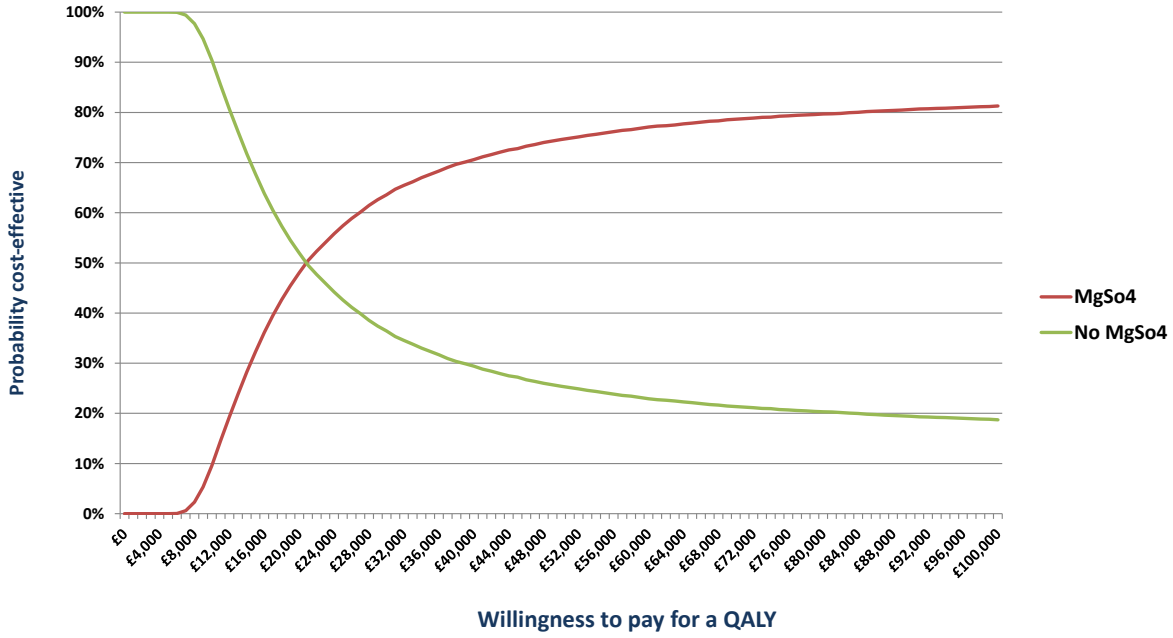


Figure 56: Monte Carlo simulation of magnesium sulfate for neuroprotection assuming no costs from adverse outcomes and a treatment cost of £7,675 and a willingness to pay of £30,000 per QALY

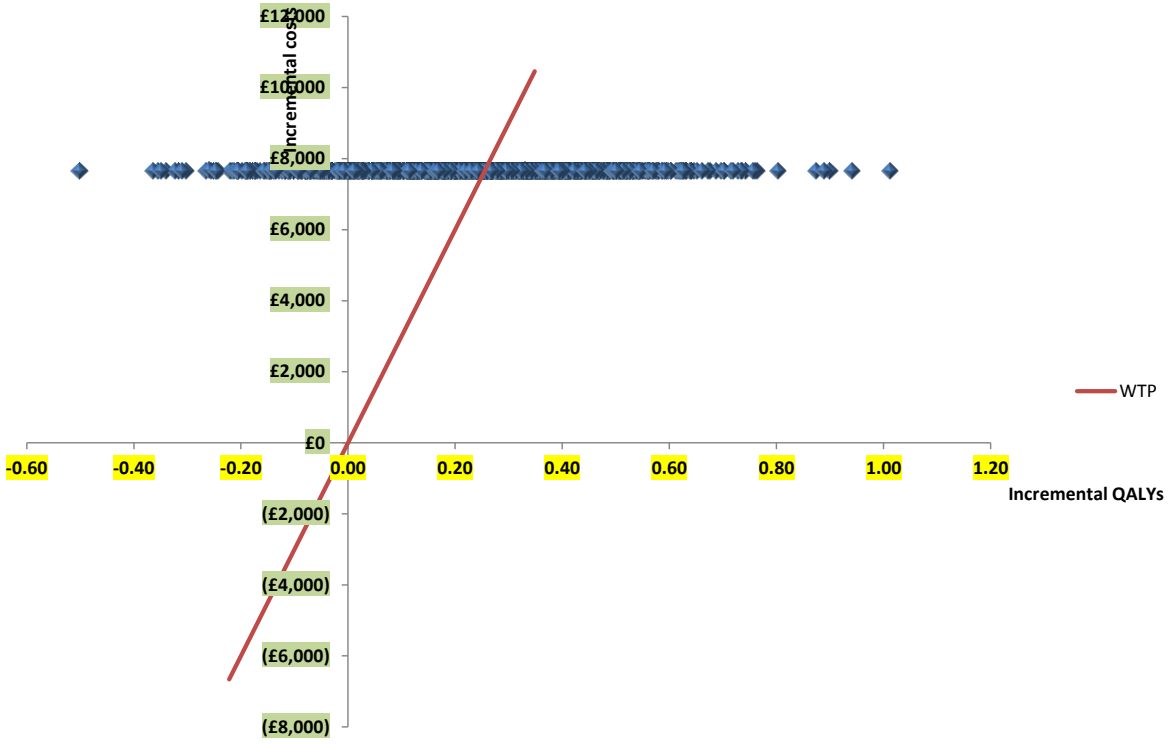
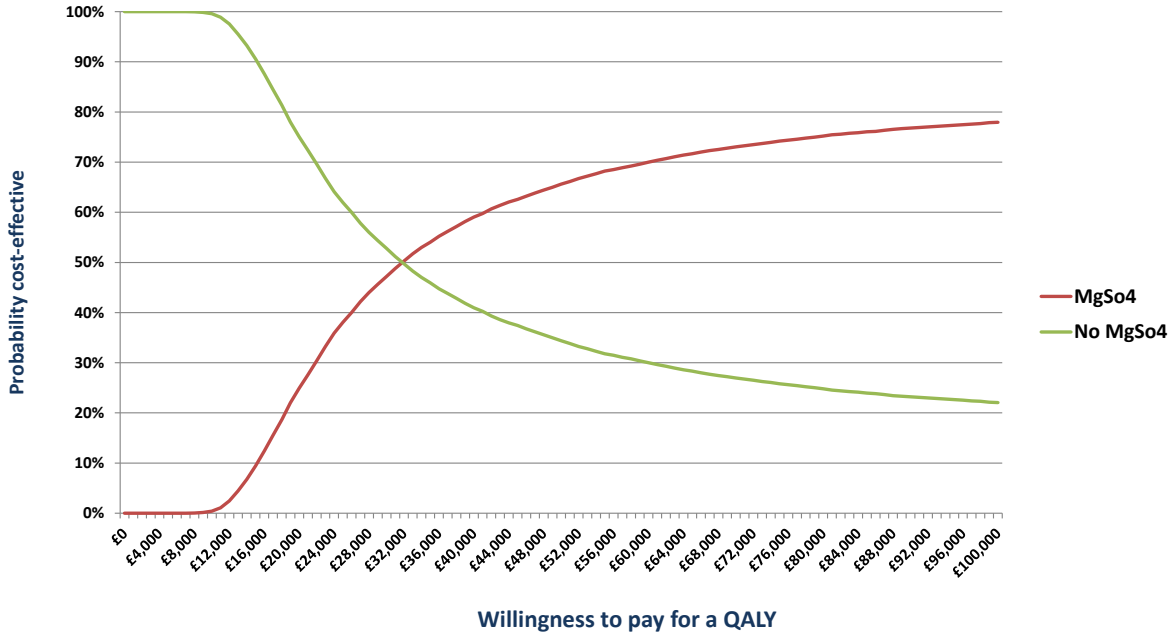


Figure 57: Cost-effectiveness acceptability curve magnesium sulfate for neuroprotection no costs from adverse outcomes and a treatment cost of £7,675

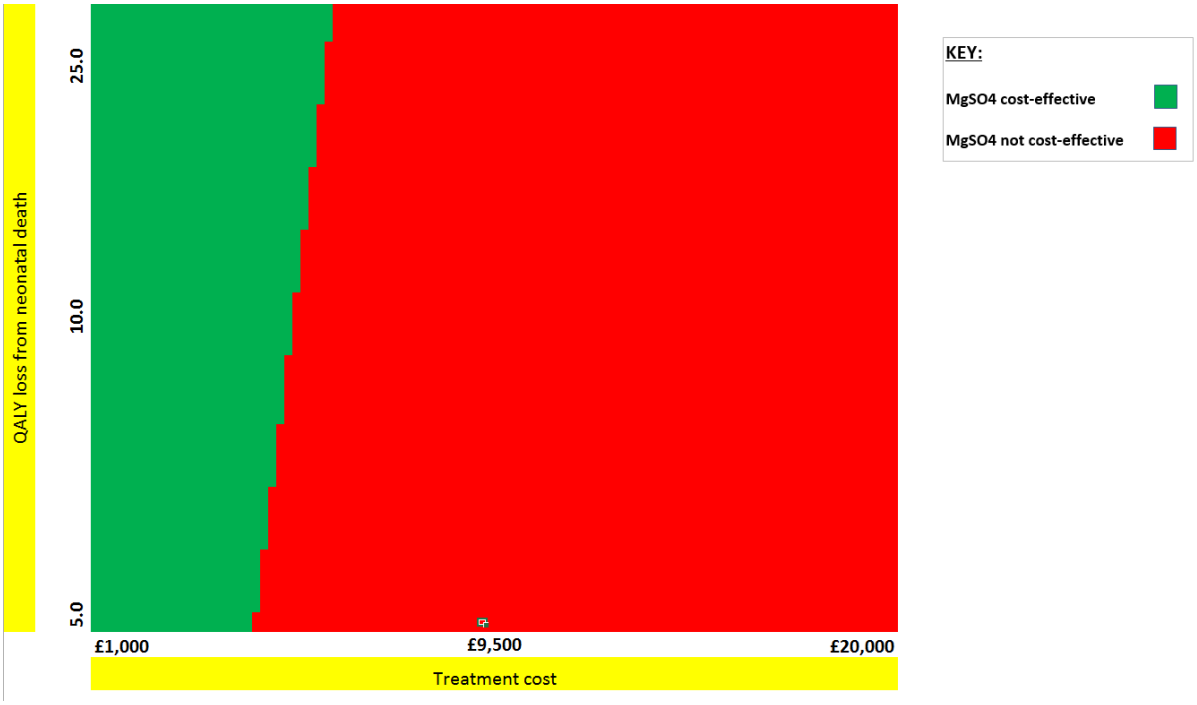


16.3.3.1.4 Two-way sensitivity analysis varying the treatment cost and the QALY loss from mortality

In this sensitivity analysis 2 model inputs are varies across a wide range of values to estimate a combined cost-effectiveness threshold. It also indicates the trade-off necessary in order across these 2 variables necessary for cost-effectiveness.

Treatment costs are varied between £1,000 and £20,000 and the QALY loss from mortality is varied between 5 and 25 QALYs and the results of this sensitivity analysis are shown in Figure 66 for a £20,000 willingness to pay for a QALY. The base case analysis falls a long way from the threshold with a treatment cost of just over £1,000 and a 22.70 QALY loss from mortality.

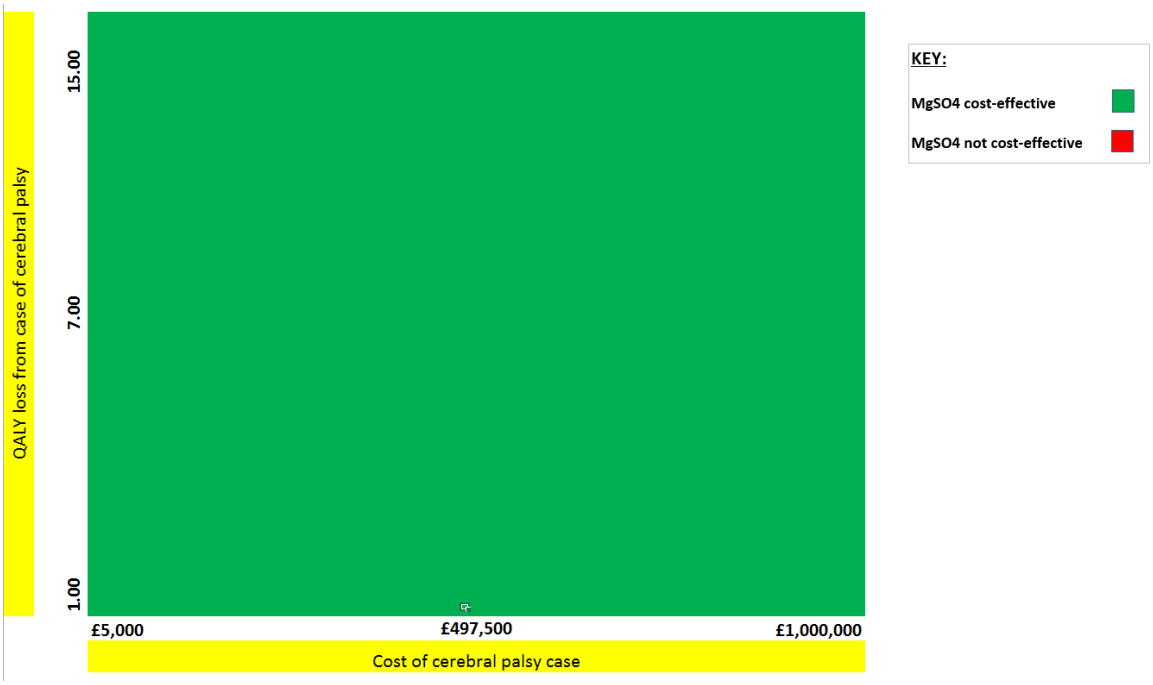
Figure 58: Graph to show the two-way relationship between QALY loss from mortality and treatment cost holding all other values constant at their base case value



16.3.4 Two-way sensitivity analysis varying the cost of cerebral palsy and the QALY loss from cerebral palsy

In this sensitivity analysis the cost and QALY loss from cerebral palsy are varied between wide levels, including much lower values assumed in the base case analysis. However, the conclusion that magnesium sulfate given for neuroprotection is cost-effective remains the case for all the cost QALY combinations assessed in this analysis as displayed in Figure 59

Figure 59: Graph to show the two-way relationship between QALY loss from cerebral palsy and cerebral palsy cost holding all other values constant at their base case value



16.3.5 Discussion

This model supports other published economic evaluation in finding magnesium sulfate to be a cost-effective intervention for neuroprotection in preterm birth (Cahill 2011; Bickford 2013). Like those studies magnesium sulfate was found to be dominant when compared with an alternative of no magnesium sulfate.

Probabilistic sensitivity analysis took into account treatment uncertainty where magnesium sulfate was only found to offer a statistically significantbenefit for only 1 of the 4 outcomes, but nevertheless found that magnesium sulfate had an 86% probability of being the most cost-effective treatment using base case inputs.

The finding that magnesium sulfate was cost-effective was generally robust to changes in model inputs that were made to favour the alternative of no magnesium sulfate for neuroprotection. A sensitivity analysis demonstrated a treatment cost threshold at which magnesium sulfate for neuroprotection would no longer be cost-effective but this treatment cost had to be more than 5 times the treatment cost used in the model, which included a cost of hospitalisation for all women.

The assumptions made with respect to cerebral palsy were potentially important to the results of the analysis as this is where the clinical review for our guideline found the greatest evidence of treatment benefit. However, sensitivity analyses assuming a much lower cost of cerebral palsy and a much lower QALY gain from an averted case did not alter the finding that magnesium sulfate was cost-effective.

The guideline did not consider the cost-effectiveness by gestational age within women who could be considered for treatment. If relative treatment effects do not vary with gestational age then treatment is likely to be relatively more cost-effective at the younger gestational ages as the risk of adverse outcomes declines with increasing gestational age.

16.3.6 Conclusion

The model produced for this guideline provides strong support for the cost-effectiveness of magnesium sulfate given for neuroprotection and this is reflected in the recommendations made by the Guideline Committee.

16.4 What is the clinical and cost effectiveness of tocolytics given to women with suspected or diagnosed preterm labour to improve outcomes:

- **progesterone/progestogens**
- **beta-sympathomimetics**
- **oxytocin receptor antagonists**
- **calcium channel blockers**
- **cyclo-oxygenase enzyme inhibitors**
- **non-steroidal anti-inflammatory drugs**
- **nitric oxide donors**
- **magnesium sulfate?**

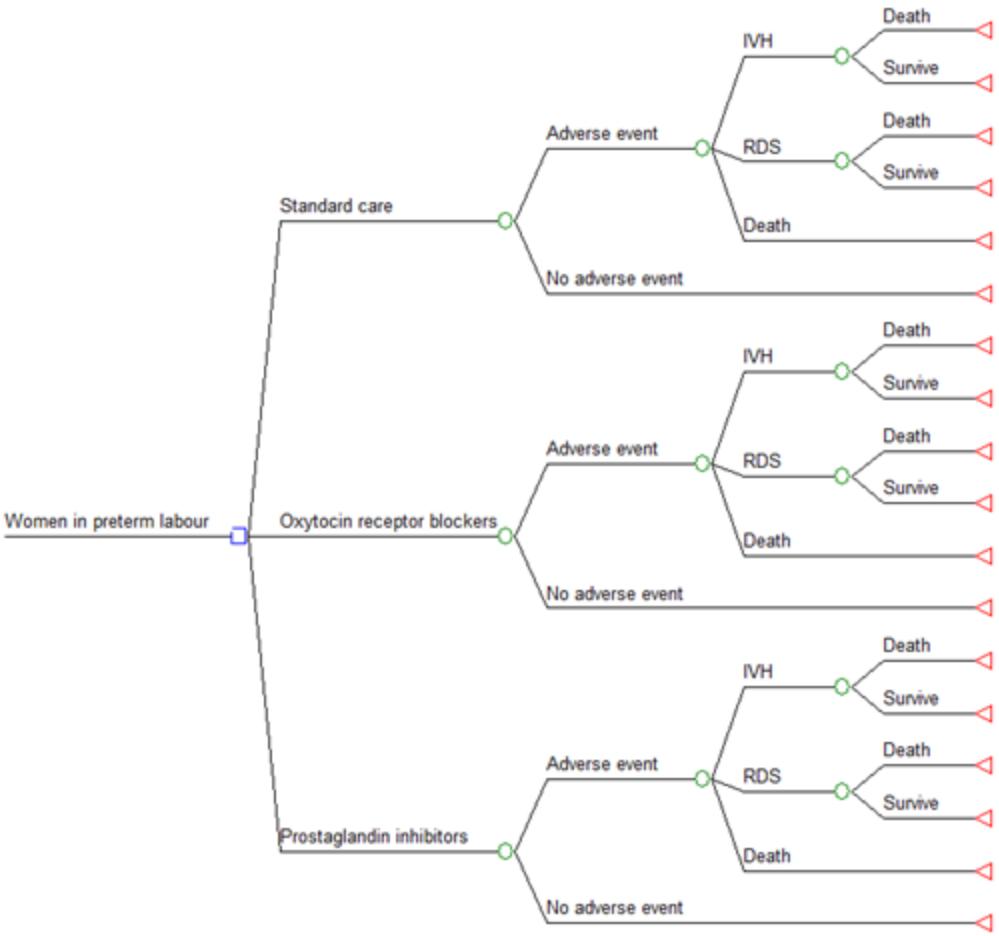
16.4.1 Introduction

Preterm birth can be costly for the health services and accounts for a disproportionate amount of infant death and morbidity. Indeed, very preterm birth which accounts for just 1% of UK births is implicated in more than half of infant deaths (https://www.rcog.org.uk/globalassets/documents/guidelines/gtg1b26072011.pdf). Therefore, prevention of preterm birth is important in order to improve child outcomes. There are a range of medications which have been proposed as having a tocolytic function and there is considerable variation in their cost. Therefore, cost-effectiveness analysis can potentially be helpful in making decisions between the use of these particular alternatives or none.

16.4.2 Methods

A decision analytic model was developed in Microsoft Excel® to assess the cost effectiveness of drugs given to women with suspected or diagnosed preterm labour in order to delay birth and by so doing improve neonatal outcomes. A simplified schematic of the model is shown in Figure 60.

Figure 60: Schematic of decision tree model to assess the cost-effectiveness of drugs given to women with suspected or diagnosed preterm labour



The model has been developed so that cost effectiveness can be assessed by gestational age in weekly increments from 24 weeks to 34 weeks. In total 6 different drug classes can be compared in addition to standard care although the model can be set to run with 1 or more treatment alternatives excluded. The various treatment alternatives compared within the model are as follows:

- standard care
- betamimetics
- calcium channel blockers
- magnesium sulfate
- oxytocin receptor blockers
- prostaglandin inhibitors
- nitrates

Treatment is expected to work by delaying birth and to reflect this in the schematic (see Figure 60) it might have seemed more intuitive if the first chance nodes following treatment had been delayed birth and not delayed. However, it was not possible from the clinical data to determine different probabilities for adverse events according to whether labour was delayed or not. Furthermore, in calculating overall health gain from a particular treatment it is the risk of the outcome by treatment which is required which, if the data existed, would simply be a weighted average of the risks in the delayed/not delayed group.

16.4.2.1 Clinical outcomes

As part of the protocol for the clinical review the Committee prioritised the outcomes listed in Table 107 below.

Table 107: Committee prioritised outcomes for clinical review of tocolysis

Outcome	Included in the NMA
Maternal outcomes	
Maternal mortality	
Adverse events – discontinuation of treatment	Included
Maternal infection	
Neonatal outcomes	
Perinatal mortality	Included
Neonatal mortality	Included[a]
Delay of birth by more than 48 hours	Included
Mean gestational age at birth	Included
Respiratory distress syndrome (RDS)	Included[a]
Chronic lung disease/bronchopulmonary dysplasia	
Intraventricular haemorrhage (IVH)	Included[a]
White matter injury/periventricular leucomalacia	
Neonatal infection/sepsis	Included
Neurodevelopmental disability	

a. Used as an outcome in the health economic model

The health economic model was restricted to outcomes for which a NMA was undertaken in order to ensure all the drug treatment classes could be compared in a consistent manner. However, if it was thought evidence from pairwise comparison of outcomes might have a bearing on cost effectiveness, the Committee would be able to use that information additionally in making their recommendations.

Of the 8 outcomes included in the NMA, 3 outcomes were chosen for inclusion in the model – neonatal mortality, respiratory distress syndrome (RDS) and intraventricular haemorrhage (IVH). Whilst delay of birth is the objective of treatment, that was considered to be an intermediate marker for the real end-points of interest, namely improved neonatal and maternal outcomes. Similarly, discontinuation of treatment and mean gestational age were considered to be proxies for other outcomes influencing health related quality of life. Neonatal infection/sepsis was not included because prognosis is often good unless it leads to a death or neurodevelopmental problems in which case it would usually be captured within the included outcomes. Neonatal and perinatal mortality are both important outcomes but is thought that there would be issues of double counting if both outcomes were included and therefore it was felt that neonatal mortality was the most useful outcome of the two.

16.4.2.2 Baseline data

The NMA generates a measure of treatment effect for each drug class relative to placebo. It was assumed that placebo could be used to represent a standard care or no drug treatment to delay birth option. However, many of the studies in the NMA were quite dated and therefore the Committee considered that the outcomes in the placebo arms in those trials were unlikely to represent the current risk of standard care for the outcomes included in the health economic analysis.

Therefore, rather than use the placebo risk in trials the baseline risks for mortality, RDS and IVH were estimated from different sources. However, this is not without its own limitations as it is likely that the data that is used to inform the baseline risk includes women with suspected or diagnosed preterm labour who will often have been given some drug treatment in order to delay birth. If a single drug class dominated current practice then it would have been possible to use this as the baseline treatment and measure the other treatments effectiveness relative to this drug class. However, the Committee considered there was too much variation in current practice to do this. Therefore, if drug treatment to delay birth is effective the baseline risks used in this analysis may under-estimate the risk associated with standard care or no drug treatment.

Whilst the health economic model used data from the NMA on neonatal mortality a more useful outcome would have been all deaths including stillbirths but, as noted in Section 16.4.2.1 above, the trial data did not record the data in a manner that allowed it to be analysed in this way. The baseline data presented in Table 108 gives the risk of all perinatal (including stillbirth) and neonatal death and it is assumed that the relative treatment effect derived from the NMA on neonatal death will be the same when applied across all deaths. This is a strong assumption but at least the NMA did not find statistically significant differences in perinatal mortality for the drug treatment classes in this analysis.

Table 108: Baseline death rate with no drug treatment to delay birth by gestational age[a]

Gestational age (weeks)	Births	Deaths[b]	Mortality rate
24	759	473	0.623
25	801	386	0.482
26	887	299	0.337
27	1,009	268	0.266
28	1,178	211	0.179
29	1,339	210	0.184
30	1,623	169	0.104
31	2,140	197	0.092
32	3,094	191	0.062
33	4,141	209	0.050
34	6,975	226	0.032

a. Source: Live births, stillbirths and infant deaths by gestational age at birth, 2011 birth cohort (ONS, 2011)
b. Includes stillbirths, early perinatal deaths and late perinatal deaths

Published data was used in order to estimate the baseline risk of RDS and IVH (Ross, 2014) and is shown in Table 109 and Table 110 respectively. In order to sample the baseline risk for probabilistic sensitivity analysis it was useful to estimate an actual number of events and this was done by multiplying the births in Table 108 by the mortality rate reported in the paper.

Table 109: Baseline respiratory distress syndrome rate with no drug treatment to delay birth by gestational age

Gestational age (weeks)	Births	RDS cases[a]	RDS rate
24	759	531	0.700
25	801	720	0.899
26	887	824	0.929
27	1,009	847	0.839
28	1,178	765	0.649
29	1,339	709	0.622

Gestational age (weeks)	Births	RDS cases[a]	RDS rate
30	1,623	892	0.555
31	2,140	791	0.370
32	3,094	866	0.280
33	4,141	1,407	0.340
34	6,975	976	0.140

a. Estimated

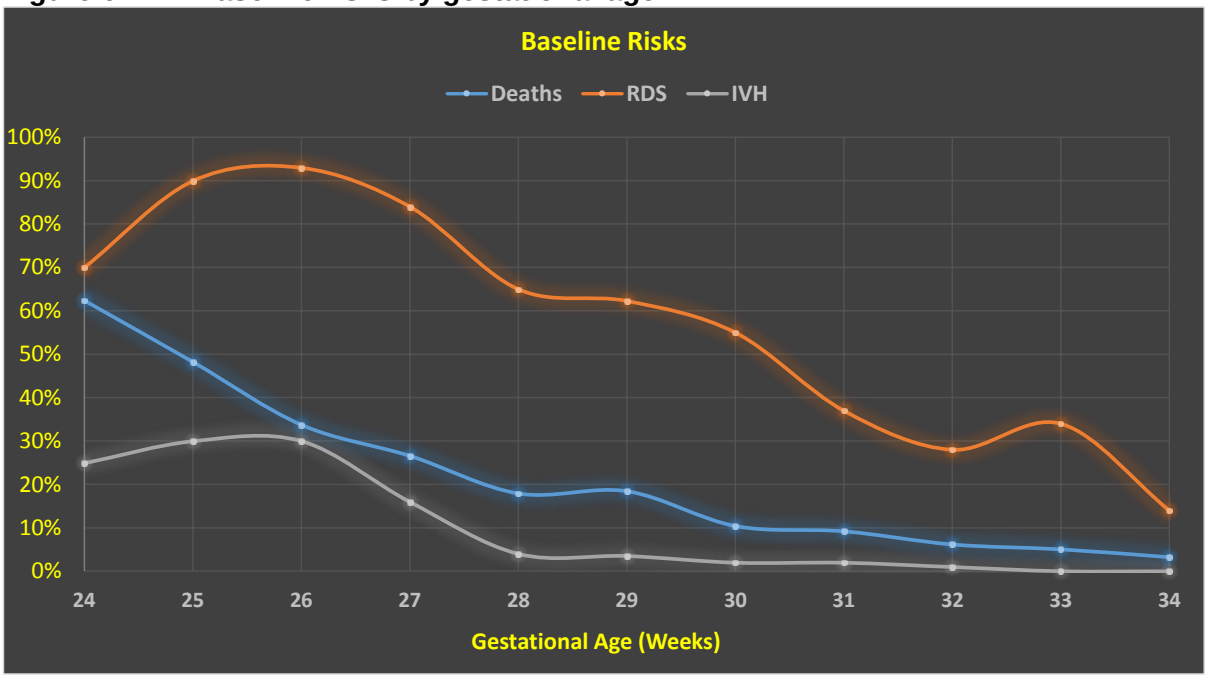
Table 110: Baseline intraventricular haemorrhage rate with no drug treatment to delay birth by gestational age

Gestational age (weeks)	Births	IVH cases[a]	IVH rate
24	759	189	0.249
25	801	240	0.300
26	887	266	0.300
27	1,009	161	0.160
28	1,178	47	0.040
29	1,339	40	0.035
30	1,623	32	0.020
31	2,140	42	0.020
32	3,094	30	0.010
33	4,141	0	0.000
34	6,975	0	0.000

a. Estimated

The baseline risks for all 3 model outcomes by gestational age are depicted graphically in Figure 61.

Figure 61: Baseline risks by gestational age



16.4.2.3 Mortality due to RDS and IVH

Trials included in the NMA will have counted RDS and IVH cases even where those cases resulted in death. This double counting is accounted for in the model by estimating the mortality associated with RDS and IVH. This enables the model to estimate the proportion of babies with RDS and IVH that survive, which is important when calculating the QALYs (see Section 16.4.2.6).

It is assumed that the mortality rate from RDS and IVH does not vary by gestational age. RDS mortality was estimated using published US data (American Lung Association Lung Disease Data, 2008). This data suggested that RDS affected 16,268 babies born in the United States in 2005 with 875 of those cases resulting in death. The mortality rate of IVH has been estimated at between 27-50% from severe (high-grade) IVH and at 5% for from low-grade haemorrhage. Table 111 below summarises the RDS and IVH mortality risk used in the model's base case analysis.

Table 111: RDS and IVH mortality rates by gestational age

Gestational age (weeks)	RDS mortality rate	IVH mortality rate
24	0.054	0.300
25	0.054	0.300
26	0.054	0.300
27	0.054	0.300
28	0.054	0.300
29	0.054	0.300
30	0.054	0.300
31	0.054	0.300
32	0.054	0.300
33	0.054	0.300
34	0.054	0.300

16.4.2.4 Treatment effectiveness

Section 16.4.2.2 outlines how the baseline risk of 3 model outcomes has been estimated. This baseline is assumed to represent the risk when no drugs are given to delay preterm birth. The NMA estimates a treatment effect size for each of the drug classes in this analysis relative to this baseline risk. A baseline risk and a relative treatment effect allow the absolute risk or probability of the outcome to be calculated for each treatment class.

Absolute risk for Betamimetics = Baseline risk x Relative risk

This absolute risk can then be used to generate the weighted QALYs and costs associated with different drug classes.

The model assumes that the relative treatment effect, derived from the NMA, will be the same across gestational age. Women with pregnancies less than 26 weeks were not included in the studies that made up the NMA but the Guideline Committee considered it reasonable to assume that treatment would be equally effective in these women. However, the absolute treatment effect varies with gestational age, reflecting the different baseline risks at different gestational ages.

16.4.2.5 Costs

Costs were based on an NHS and Personal Social Services perspective as outlined in the NICE reference case in Developing NICE guidelines: the manual (https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf). Costs were based on a 2015 price year or as close to 2015 prices as could be estimated when sourced from an earlier price year.

16.4.2.5.1 Treatment costs

For the NMA that provides the estimates of treatment effectiveness, the Committee agreed that it was reasonable to construct the network by treatment class as opposed to individual drugs. The rationale for this being that they expected there to be little variation in treatment effectiveness by class. Where treatments are assumed to be equally effective it follows that the most cost effective treatment among them will be the cheapest. Therefore, the Committee agreed that the costing for each treatment class would be based on the cheapest drug in class that they would be willing to recommend if this drug was shown to be cost-effective. The costs used for treatment costs are shown in Table 112. How these costs were derived is described below.

Table 112: Treatment costs

Drug class	Cost	Source
Betamimetics	£81	BNF 2015
Calcium channel blockers	£14	BNF 2015
Magnesium sulfate	£169	BNF 2015
Oxytocin receptor blockers	£517	BNF 2015
Prostaglandin inhibitors	£14	BNF 2015
Nitrates	£16	BNF 2015

In addition to the pharmaceutical costs of the treatment it is assumed that some staff time is required in order to administrate. For each treatment the staffing time is estimated and a total cost of staff time is calculated based on the unit costs of staff time shown in Table 113.

Table 113: Staff unit costs

Staff	Unit cost per hour	Source
Nurse	£120[a]	Curtis (2014)
Doctor	£41[b]	Curtis (2014)

a. Based on per hour of patient contact for the full-time equivalent basic salary for Agenda for Change band 6. This cost per hour is based on observations about the ratio of direct to indirect time on face to face contact. It is assumed that 41% of a nurses time is spent on direct patient with the remaining 59% spent on non-patient activities such as administration and paperwork

b. Based on a Foundation House officer 2 on a 48 hour week

Betamimetics

Two betamimetics were compared to obtain the lowest cost in class:

i. Terbutaline sulfate

A summary of the doses and mode of administration used in the included studies is shown below:

- mavaldi: 0.25 mg of loading dose subcutaneously. Same dose repeated every 45 minutes

- laohapojanart: 10 µg/min with an increase of 5µg/min every 10 min if required until 25 µg/min reached
- motazadian: 250 µg subcutaneous followed by the same dose every 45 minutes
- how: oral 5 to 10 mg every 4 to 6 hours

For the purposes of this costing dosing was based on the following (http://www.fpnotebook.com/ob/Pharm/Trbtln.htm):

Intravenous

1. start: 10 mcg/minute
2. increase rate by 5 mcg per minute every 10 minutes
3. maximum: 25 mcg per minute
4. once controlled, decrease dose 5 mcg every 30 minutes
5. titrate dose down to lowest effective dose

It was assumed that a total of 50.55 mg is administered over a period of 48 hours.

a. 1-10 minutes = 10 x 10 mcg = 0.1 mg
b. 11-20 minutes = 10 x 15 mcg = 0.15 mg
c. 21-30 minutes = 10 x 20 mcg = 0.2 mg
d. 31minutes-24hours = 1410 x 25 mcg = 35.25 mg
e. 24h-24h30min = 30 x 20 mcg = 0.6 mg
f. 24h-30min-25h = 30 x 15 mcg = 0.45 mg
g. 25h-48h = 1380 x 10 mcg = 13.8 mg

From the BNF (accessed 30 March, 2015), the price of 1-mL ampule of terbutaline sulfate 500 micrograms/mL for injection was £0.43. The treatment cost was calculated for terbutaline sulfate as shown in Table 114.

Table 114: Calculation of treatment cost of terbutaline sulfate

	Unit Cost	**Unit cost per mg**
Injection (0.5 mg)	£0.43	£0.86
Total infused	50.55 mg	£43.37
Staff	**Minutes**	**Cost**
Nurse	15	£30.00
Doctor	15	£10.25
Total Cost		**£83.72**

ii. Salbutamol

A summary of the doses and mode of administration used in the included studies is shown below:

- jannet: dilution of 2.5 mg in a 500 ml 5% weight per volume glucose solution, with an initial flow rate of 30 ml/h, 0.15 mg/h
- motazadian: IV bolus 0.1 mg followed by the same boluses every 5 minutes
- jannet: IV infusion; initial dose 12 µg/min. The dose was increased by 6 µg/min at 10 minute intervals up to maximum of 50 µg/min until the desired effect was achieved

It was assumed that a total of 79 mg is administered over a period of 48 hours:

a. 1-10 minutes = 10 x 10 mcg = 0.1 mg
b. 11-20 minutes = 10 x 15 mcg = 0.15 mg
c. 21-30 minutes = 10 x 20 mcg = 0.2 mg
d. 31-40 minutes = 10 x 25 mcg = 0.25 mg
e. 41-50 minutes = 10 x 30 mcg = 0.3 mg
f. 51-60 minutes = 10 x 35 mcg = 0.35 mg
g. 61-70 minutes = 10 x 40 mcg = 0.4 mg
h. 71 minutes-24h = 1370 x 45 mcg = 61.65 mg
i. 24h-30h = 360 x 22.5 mcg = 8.1 mg
j. 30h-36h = 360 x 11.3 mcg = 4.05 mg
k. 36h-42h = 360 x 5.65 mcg = 2.03 mg
l. 42h-48h = 360 x 2.83 mcg = 1.01 mg

From the BNF (accessed 30 March, 2015), the price of a 5-mL ampule of 1mg/1mL salbutamol (as sulfate) solution for intravenous infusion was £2.48. The treatment cost for salbutamol was calculated as shown in Table 115.

Table 115: Calculation of treatment cost of salbutamol

	Unit Cost	Unit cost per mg
Solution for IV infusion	£2.48	£0.50
Drug	Maximum	Total drug cost
IV infusion	79 mg	£39.18
Staff	**Minutes**	**Cost**
Nurse	15	£30.00
Doctor	15	£10.25
Total cost		**£79.43**

Calcium channel blockers

Two calcium channel blockers were compared to obtain the lowest cost in class:

i. Nifedipine

A summary of the doses and mode of administration used in the included studies is shown below:

- haghighi; 10 mg capsule given sublingually repeated every 20 min (up to maximum of 40 mg during the first hour of treatment)
- mavaldi: 30 mg of oral loading dose followed by 20 mg orally after 90 minutes
- taherian: Start with 10 mg orally repeated every 20 minutes (Max dose 40 mg in the first hour)
- laohapojanart: 10 mg crashed and swallowed, 10 mg every 20 min with maximum 40 mg in the first hours. After that 20 mg every 4 hours up to 72 hours.
- al-quattan: 30 mg of oral loading dose followed by 20 mg orally after 120 minutes
- papatsonis: 10 to 40 mg of oral loading dose at the first hour followed maintenance dose of 60 -160 mg of slow releasing daily until 34 weeks

- koks: 30 mg of oral loading dose followed by 20 mg 2 to 4 times daily
- kashanian: 10 to 40 mg of oral up to maximum of 4 doses if contraction subsided 20 mg every 6 hours for the first 24 hours; 20 mg every 8 hours for the second 24 hours; 10 mg every 8 hours for the next 24 hours
- al-omari: 10 mg orally, by chewing every 15 min. maximum dose 40 mg in the first hour then 10 mg every 4 – 6 hours
- klauser: 30 mg of oral loading dose followed by 20 – 30 mg orally every 4 to 6 hours

For the purpose of this costing it was assumed that 40 mg is given as a loading dose followed by 20 mg every 6 hours for a period of 72 hours (240 mg). From the BNF (accessed 30 March, 2015), the price of a 90 capsule pack of 10 mg nifedipine was £7.30. The treatment cost for nifedipine was calculated as shown in Table 116.

Table 116: Calculation of treatment cost of nifedipine

	Unit cost	**Unit cost per mg**
90 capsules	£7.30	£0.01
Drug	**Dose**	**Cost**
Loading dose	40	£0.32
Subsequent dose	240	£1.95
Staff	**Minutes**	**Cost**
Nurse	5	£10.00
Doctor	5	£3.42
Total cost		**£15.69**

ii. Nicardipine

A summary of the doses and mode of administration used in the included studies is shown below:

- jannet: dilution of 50 mg in a 500 ml 5% weight per volume glucose solution with an initial flow rate of 30 ml/h, 3 mg/h
- larmon: 40 mg oral dose, after 2 hours if needed further 20 mg oral dose

For the purposes of this costing it was assumed that an initial dose of 40 mg is followed by a subsequent dose of 20 mg. From the BNF (accessed 30 March, 2015), the price of a 56 capsule pack of 20 mg nicardipine was £6.00. The treatment cost for nicardipine was calculated as shown in Table 117.

Table 117: Calculation of treatment cost of nicardipine

	Unit cost	Unit cost per mg
56 capsules	£6.00	£0.005
Drug	**Dose**	**Cost**
Capsules	60mg	£0.32
Staff	**Minutes**	**Cost**
Nurse	5	£10.00
Doctor	5	£3.42
Total cost		**£13.74**

Magnesium sulfate

A summary of the doses and mode of administration that was used in the included studies is shown below:

- Haghighi; loading dose 6 g IV followed by an infusion of 2 g/h increase to max 4g/h
- Taherian: loading dose 4 g IV over 15 min then followed by an IV infusion of 2-3 g/h
- Wilkins: loading dose 4 g IV over 15 min then followed by an IV infusion of 2 g/h
- Cotton: loading dose 4 g IV over 15 min then followed by a continuous IV infusion of 2 g/h
- Mcwhorter: loading dose 4 to 6 g in 20% solution followed by an continuous infusion at the rate of 2 to 4 g/h
- Klauser: loading dose 6 g IV over 20 min followed by an infusion of4-6 g/h
- Borna: loading dose 4 to 6 g in 20% solution followed by an continuous infusion of 2 to 4 g/h
- El-Sayed: 4 g iv bolus, then at the rate of 2 – 4 g/h
- Larmon: loading dose 6 g IV (2g/h) increase to max 4g/h

For the purpose of this costing it was assumed that 4 g is given as a loading dose followed by 2g/hour by intravenous infusion for a period of 48 hours (96 g). From the BNF (accessed 30 March, 2015), the price of a 20-mL (4-g) ampule of magnesium sulfate for injection was £16.98 and the cost of a 2-mL (1-g) ampule of magnesium sulfate was £1.18. The treatment cost for magnesium sulfate was calculated as shown in Table 118.

Table 118: Calculation of treatment cost of magnesium sulfate

	Unit cost	Unit cost per mg
20ml (4-g) ampule	£16.98	£0.004
2mL (1-g) ampule	£1.15	£0.001
Drug	**Dose**	**Cost**
Loading dose	4000 mg	£16.98
Continuous infusion	96000 mg	£110.40
Staff	**Minutes**	**Cost**
Nurse	15	£30.00
Doctor	15	£10.25
Total cost		**£167.63**

Oxytocin receptor blockers

A summary of the doses and mode of administration used in the included studies is shown below:

- al-omari: bolus 6.7mg IV over 1 min followed by an IV infusion of 18 mg/h for 3 hours followed by 6 mg/h for 24 -48 hours

It was assumed that the dosing was as indicated by the BNF (accessed 30 Match 2015), intravenous injection, initially 6.75 mg over 1 minute, then by intravenous infusion 18 mg/hour for 3 hours, then 6 mg/hour for 45 hours. From the BNF (accessed 30 March, 2015), the price of a 0.9-mL (6.75 mg) vial of atosiban (as acetate) for injection was £18.41 and the cost of a 5-mL vial (7.5 mg/mL) for intravenous infusion was £52.82. The treatment cost for atosiban was calculated as shown in Table 119.

Table 119: Calculation of treatment cost of oxytocin receptor blockers

	Unit cost	Unit cost per mg

Injection vial (6.75 mg)	£18.41	£2.73
IV vial (37.5 mg)	£52.82	£1.41
Drug	**Dose**	**Cost**
Intravenous injection	6.75	£18.41
IV first infusion first 3 hours	54	£76.06
IV infusion next 45 hours	270	£380.30
Staff	Minutes	Cost
Nurse	15	£30.00
Doctor	15	£10.25
Total cost		**£515.02**

Nitrates

A summary of the doses and mode of administration used in the included studies is shown below:

- el-sayed: 100 µg IV bolus, then at a rate of 1 to 10 µg/kg/minute
- smith: transdermal patch

For the purpose of this costing it was assumed that women would be given 4 '10' patches releasing 10 mg/24 hours of glyceryl trinitrate over the course of 48 hours. From the BNF (accessed 30 March 2015) the price of a 28 pack '10' patch of glyceryl trinitrate is £14.06. The treatment cost for nitrates was calculated as shown in Table 120.

Table 120: Calculation of treatment cost of nitrates

	Unit cost	**Unit cost per mg**
'10' patch	£14.06	£0.05
Drug	**Dose**	**Cost**
4 patches	40 mg	£2.01
Staff	**Minutes**	**Cost**
Nurse	5	£10.00
Doctor	5	£3.42
Total cost		**£15.43**

Prostaglandin inhibitors

Five prostaglandin inhibitors were compared to obtain the lowest cost in class:

i. Indomethacin

A summary of the doses and mode of administration used in the included studies is shown below:

- besinger: 50mg orally initially then 25 to 50 orally every 4 hours until the contraction ceased then 25 mg maintenance therapy orally every 4-6 hours
- kashanian: 100 mg suppository, a repeat administrated 1 hour later with the same dose. The maximum dose 200 mg daily
- klauser: 100 mg suppository, a repeat administrated 2 hours later with the same dose, followed by 50 mg orally every 6 h for 12 hours
- spearing: 100 mg suppository, a repeat administrated 12 hours later with the same dose, followed by 25 mg orally every 6 h for 48 hours

For the purposes of this costing it was assumed that a 100 mg suppository is administered followed by a repeat administration with the same dose. This is then followed by 25 mg orally every 6 hours for 48 hours. From the BNF (accessed 30 March 2015) the price of a 10 pack of 100 mg indomethacin suppositories is £17.61 and the price of a 28 capsule pack of 25 mg indomethacin is £1.32. The treatment cost for indomethacin was calculated as shown in Table 121.

Table 121: Calculation of treatment cost of indomethacin

	Unit cost	Unit cost per mg
28 capsules	£1.32	£0.002
Suppositories	£17.61	£0.18
Drug	Dose	Cost
Suppositories	200 mg	£35.22
Capsules	200 mg	£0.38
Staff	**Minutes**	**Cost**
Nurse	5	£10.00
Doctor	5	£3.42
Total cost		**£49.01**

ii. Celecoxib

A summary of the doses and mode of administration used in the included studies is shown below:

- borna: 100 mg orally twice daily

For the purposes of this costing it was assumed that the women receive 100 mg orally twice daily over 48 hours (400 mg in total). From the BNF (accessed 30 March 2015) the price of a 60 capsule pack of 100 mg celecoxib is £21.55. The treatment cost for celecoxib was calculated as shown in Table 122.

Table 122: Calculation of treatment cost of celecoxib

	Unit cost	Unit cost per mg
60 capsules	£21.55	£0.004
Drug	**Dose**	**Cost**
Capsules	400 mg	£1.44
Staff	**Minutes**	**Cost**
Nurse	5	£10.00
Doctor	5	£3.42
Total cost		**£14.85**

iii. Sulindac

For the purposes of this costing it was assumed that the dose is 200mg every 12 hours for 48 hours (800 mg in total). From the BNF (accessed 30 March 2015) the price of a 56 tablet pack of 200 mg sulindac is £38.29. The treatment cost for sulindac was calculated as shown in Table 123.

Table 123: Calculation of treatment cost of sulindac

	Unit cost	Unit cost per mg
56 tablets	£38.29	£0.003
Drug	**Dose**	**Cost**

Tablets	800 mg	£2.74
Staff	**Minutes**	**Cost**
Nurse	5	£10.00
Doctor	5	£3.42
Total cost		**£16.15**

iv. Ketorolac

For the purposes of this costing it was assumed that ketorolac is administered intramuscularly as a 60-mg loading dose followed by 30 mg every 6 hours for 48 hours. From the BNF (accessed 30 March 2015) the price of a 1-mL ampule of 30 mg/mL ketorolac trometamol for injection was £1.07. The treatment cost for ketoreloc was calculated as shown in Table 124.

Table 124: Calculation of treatment cost of ketorolac

	Unit cost	**Unit cost per mg**
Injection (1-mL ampule)	£1.07	£0.036
Drug	**Dose**	**Cost**
Loading dose	60 mg	£2.14
Repeat doses	240 mg	£8.56
Staff	**Minutes**	**Cost**
Nurse	15	£30.00
Doctor	15	£10.25
Total cost		**£50.95**

v. Mefenamic acid

A study (Mital, 1992) reported using Mefenamic acid 500mg 3 times daily, although the duration was not specified in the abstract. For the purposes of this costing this dosing was used and it was assumed this was administered over 48 hours. From the BNF (accessed 30 March 2015) the price of a 28 tablet pack of 500 mg mefenamic acid is £2.62. The treatment cost for mefenamic acid was calculated as shown in Table 125.

Table 125: Calculation of treatment cost of mefenemic acid

	Unit cost	**Unit cost per mg**
28 tablets	£2.62	£0.0002
Drug	**Dose**	**Cost**
Tablets	3000 mg	£0.56
Staff	**Minutes**	**Cost**
Nurse	5	£10.00
Doctor	5	£3.42
Total cost		**£13.98**

A summary of all the costs of the different tocolytic by class is given in Table 126 and indicating the lowest cost option for use in the model.

Table 126: Summary of tocolytic treatment costs

	Terbutaline	Salbutamol	Nifedipine	Nicardipine	Magnesium sulfate	Nitroglycerine	Atosiban	Indomethacin	Celecoxib	Sulindac	Ketorolac	Mefenemic acid	Lowest cost
Betamimetics	£85	£79											**£79**
Calcium channel blockers			£16	£14									**£14**
Magnesium sulfate					£168								**£168**
Nitrates						£15							**£15**
Oxytocin receptor blockers							£515						**£515**
Prostaglandin inhibitors								£49	£15	£16	£51	£14	**£14**

Costs rounded to nearest pound

16.4.2.5.2 Costs relating to adverse outcomes

In addition to the costs of treatment it is important that "downstream" costs are also factored into the analysis as more effective treatments are likely to lead to lower costs arising from adverse outcomes. The costs relating to adverse outcomes used in the base case analysis are shown in Table 127.

Table 127: Costs of adverse outcomes

Outcome	Cost	Standard error	Distribution
Death[a]	£0	£106	Deterministic
RDS[b]	£7,000	£36	Deterministic
IVH[c]	£23,700	-	Deterministic

a. *NHS Reference Costs 2012/13, XB03Z Paediatric critical care, intensive care, advanced*
b. *NHS Reference Costs 2011/12, XB01Z Paediatric Critical Care, Intensive Care, ECMO/ECLS*
c. *It was assumed that IVH would have the same cost as ICH. It was additionally assumed that Grade III and Grade IV ICH would be similar in cost to cerebral palsy. A European paper (Kruse, 2009) estimated in year 2000 prices that the lifetime health care costs for cerebral palsy using an annual discount rate of 5% was €66,155 for men and €65,288. The mid-point of this estimate was used and converted into GBP using an exchange rate of £0.83 = €1 (http://www.exchangerates.org.uk/ - accessed 26/03/2014). It was then converted into 2011/12 prices using the HCHS (The Hospital & Community Health Services) Index. One study (Alvarez, 1994) suggested that 30% of ICH is of severity Grade III and Grade IV and therefore the cost of ICH was estimated as 0.3 x £79,000*

16.4.2.6 Quality Adjusted Life Years (QALYs)

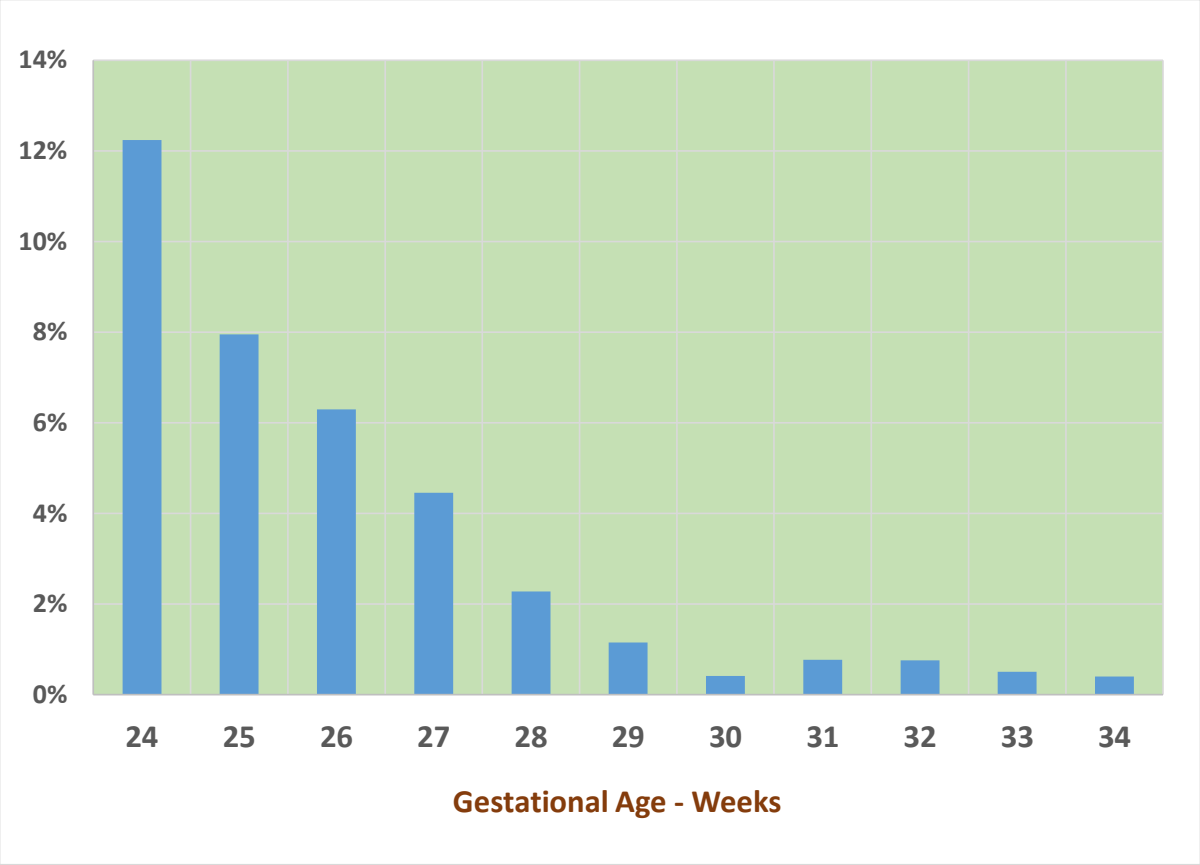
A lifetime QALY loss was assigned to each of the 3 outcomes included in the model (mortality, RDS and IVH). To calculate the QALY loss from mortality it was assumed that the maximum QALY loss associated with neonatal/perinatal death would be 22.7 QALYs. This is based on the current life expectancy in England and Wales of 80 years and the simplifying assumption that each year would be lived with a health state utility of 0.82, a value based on a UK population norm (Kind, 1982), and with an annual discount rate of 3.5% applied to that health state utility. However, in practice a lower QALY gain is likely to arise from averting mortality at earlier gestational ages as births at these ages are associated with shorter life expectancy and greater morbidity, other than that already captured by RDS and IVH. To proxy this effect, data on 1 year survival was used to estimate the proportion of averted neonatal/perinatal deaths that would not survive the first year of life. It was assumed that deaths within the first year would not have any QALY associated with them but that all those surviving the first year would receive the maximum 22.7 QALYs. Therefore, the overall QALY gain from averting a neonatal/perinatal loss is the weighted average of the QALY gain

experienced by those surviving the first year (22.7 QALYs) and the zero QALY from deaths occurring within that first year. The proportions surviving the first year varies with gestational age as is shown in Figure 62 and the QALY loss assigned to neonatal/perinatal mortality by gestational age is shown in Table 128.

Table 128: QALY loss from neonatal/perinatal mortality by gestational age

Gestational age	Survive 1st Year	Death 1st year	Max QALY	Weighted QALY
24	0.878	0.122	22.7	19.92
25	0.920	0.080	22.7	20.89
26	0.937	0.063	22.7	21.27
27	0.955	0.045	22.7	21.69
28	0.977	0.022	22.7	22.18
29	0.988	0.012	22.7	22.44
30	0.996	0.004	22.7	22.61
31	0.992	0.008	22.7	22.52
32	0.992	0.008	22.7	22.53
33	0.995	0.005	22.7	22.58
34	0.996	0.004	22.7	22.61

Figure 62: Mortality rate in first year by gestational age at birth



Source: ONS 2011

The QALY loss from RDS and IVH, not leading to neonatal/perinatal mortality are as shown in Table 129.

Table 129: QALY loss from RDS and IVH

Outcome	QALY loss
RDS	3.85
IVH	4.5

The QALY loss from RDS values was essentially a “dummy” value reflecting the highly variable prognosis. Therefore it is reasonable to posit a significant QALY loss to capture the proportion having poor long term outcomes but a proportion would also have good long term outcomes and therefore the QALY loss can expected to be considerably less than that arising from mortality. The QALY loss from IVH was based on the value we used for ICH in the model looking at the cost-effectiveness of magnesium sulfate for neuroprotection

In the base case analysis it is assumed that the NHS has a willingness to pay of £20,000 per QALY.

16.4.2.7 Probabilistic sensitivity analysis

It is usual and good practice when reporting a relative treatment effect to provide confidence intervals in addition to the point estimate. This is because there is uncertainty around the point estimate due to sampling error. As the sample size is increased this uncertainty is reduced which, everything else being equal, is reflected in narrower confidence intervals.

Similarly, in health economic analysis it is important to take into account of the uncertainty around model inputs. This can sometimes be achieved through 1 way sensitivity analysis, where 1 input value at a time is varied in order to assess what change that input has on the model’s results. However, whilst that can often provide useful insights into what inputs are driving the models results it is inadequate to address the uncertainty which exists simultaneously across all model inputs.

Probabilistic sensitivity analysis, using Monte Carlo simulation techniques, allows for uncertainty across all model inputs to be addressed. Simulation involves running the model many times. In each simulation, rather than taking an input’s point estimate it is sampled from its probability distribution. For inputs that are based on a large sample the probability distribution will be relatively narrow and the sampled inputs will reflect that. Conversely, input values derived from a small sample will have a relatively wide probability distribution.

For example, take the baseline risk of death in this model at a gestational age of 30 weeks. From Table 108 it can be seen that the point estimate is 0.104. In the probabilistic sensitivity analysis this is sampled from a beta distribution which is constrained to lie between 0-1 as a mortality rate must do. Two parameters, alpha and beta, are needed to describe a particular beta distribution. Alpha is simply the number of events, in this case 169 deaths. Beta is the number of non-events or surviving babies and is 1,454 at 30 weeks gestation. Using a random number generator, computer software (Microsoft Excel® for example) can be used to sample from the beta probability distribution for this combination of alpha and beta. Below, 10 sample from this distribution are shown.

Table 130: Example of Monte Carlo simulation of baseline mortality risk

Sample number	Sampled value
1	0.101
2	0.114
3	0.102
4	0.096
5	0.103
6	0.107
7	0.099

Sample number	Sampled value
8	0.112
9	0.095
10	0.095

If these values were being sampled as part of a Monte Carlo simulation, then these are the values that would be used for baseline mortality in the first 10 simulations of the model. The average of these 10 sampled values is 0.103 which is close to the point estimate. This is to be expected and consequently the larger the number of simulations the less role for sampling error to effect the result.

The NMA was used to simulate 100,000 log odds ratios for each treatment class relative to baseline from the posterior distribution of the probability of event (IVH, RDS or neonatal mortality). For each outcome, these values are a random sample from the joint distribution of the probabilities and therefore maintain any correlation between them. Using a mathematical transformation these log odds ratios are then converted into absolute probabilities. For the RDS outcome, nitrates were not included in the NMA and therefore it was conservatively assumed that they would have the same risk of RDS as baseline.

When the user runs the probabilistic sensitivity analysis, they can stipulate the number of simulations to run up to a maximum of 100,000. If a number less than a 100,000 is chosen the model will run a sequence of the 100,000 previously generated simulations but starting at a random point.

16.4.3 Results

Table 131 shows the results of the probabilistic sensitivity analysis based on 10,000 simulations for a gestational age of 24 weeks for the base case analysis.

Table 131: Base case PSA result based on 10,000 simulations for a gestational age of 24 weeks

Treatment	Mean cost	Mean QALY	Mean Net Benefit	Probability cost-effective	ICER
Nitrates	-£962	1.136	£23,685	0.36	n/a
Calcium channel blockers	-£916	2.467	£50,246	0.34	£35 per QALY
Prostaglandin inhibitors	-£277	-0.39	-£7,525	0.04	Dominated
Magnesium sulfate	-£221	-1.516	-£30,101	0.01	Dominated
Betamimetics	-£206	0.011	£425	0.01	Dominated
Standard care	£0	0	£0	0.04	Dominated
Oxytocin receptor blockers	£270	1.613	£31,985	0.20	Dominated

a. Mean costs and QALYs are calculated relative to standard care

Using the mean net benefit would suggest that calcium channel blockers are the most cost-effective treatment however nitrates have a marginally higher probability of being cost-effective.

Figure 63 shows a plot of 1,000 simulations of the base case analysis on the cost-effectiveness plane with incremental costs and QALYs shown relative to standard care (origin) in women with a gestational age of 24 weeks. Figure 64 is a similar plot of the same 1,000 simulations but restricted to the 3 most cost-effective treatments as assessed by net mean benefit.

The cost-effectiveness acceptability curve is shown in Figure 65. This shows the probability of different treatments being cost-effective as the willingness to pay for a QALY is varied and therefore at a willingness to pay of £20,000 per QALY, the probabilities graphed in Figure 65 correspond to the probabilities in Table 131.

Figure 63: Cost-effectiveness plane for the base case analysis showing all treatments in the analysis for women with a gestational age of 24 weeks

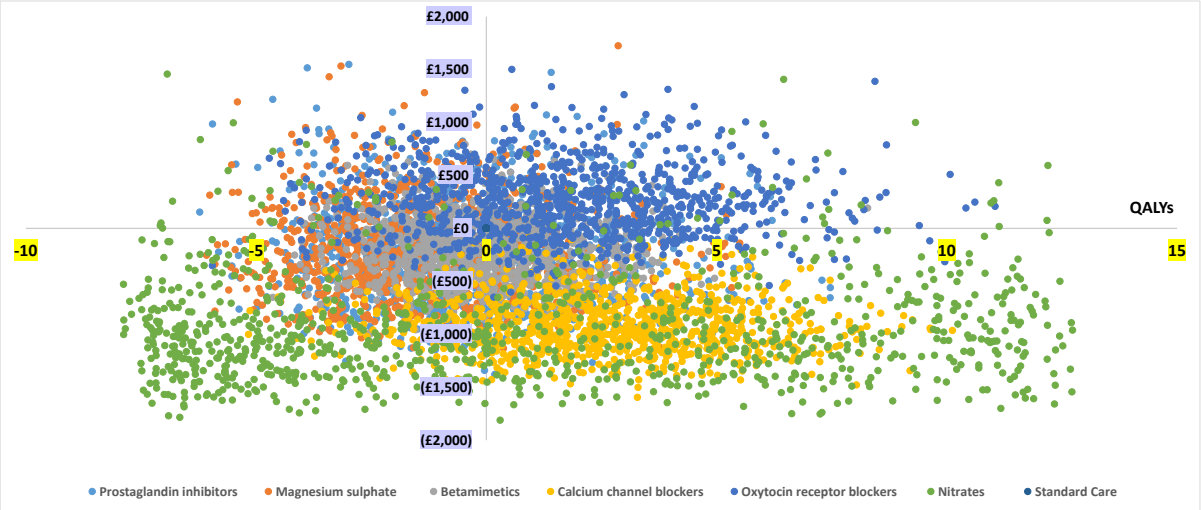


Figure 64: Cost-effectiveness plane for the base case analysis, restricted to the 3 most cost-effective treatments for women with a gestational age of 24 weeks

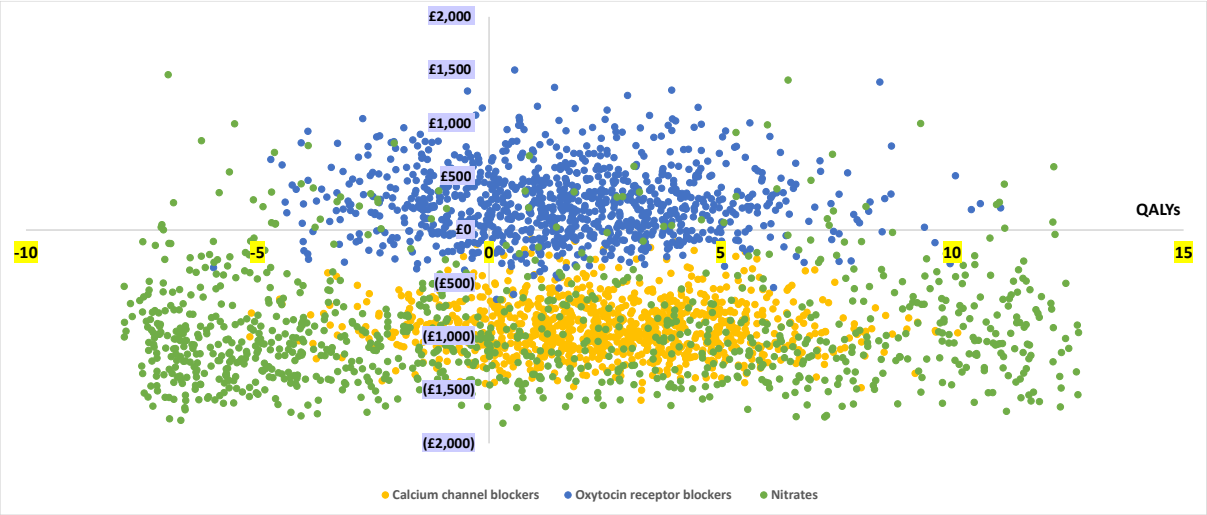
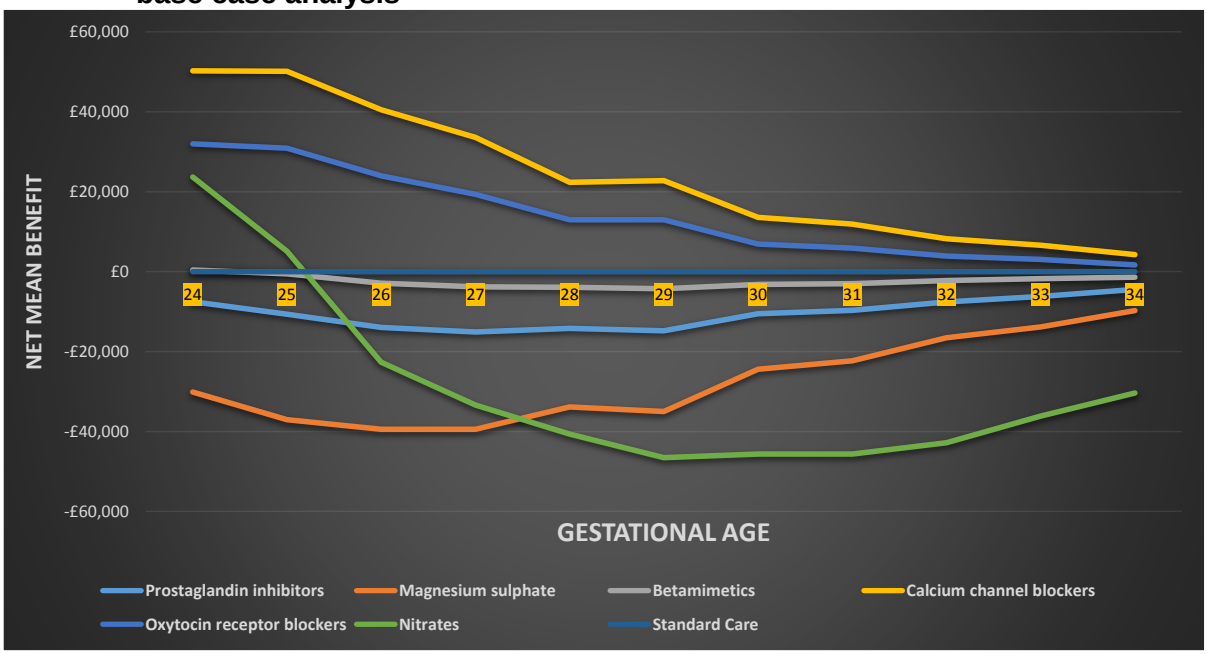


Figure 66 and Figure 67 respectively show how the net mean benefit and the probability that each treatment is cost-effective varies with gestational age. This shows that although treatment becomes relatively less cost-effective with increasing gestational age, as shown by declining mean net benefit, that calcium channel blockers continues to be cost-effective treatment for all diagnosed and suspected preterm births between 24-34 weeks at a willingness to pay threshold of £20,000 per QALY.

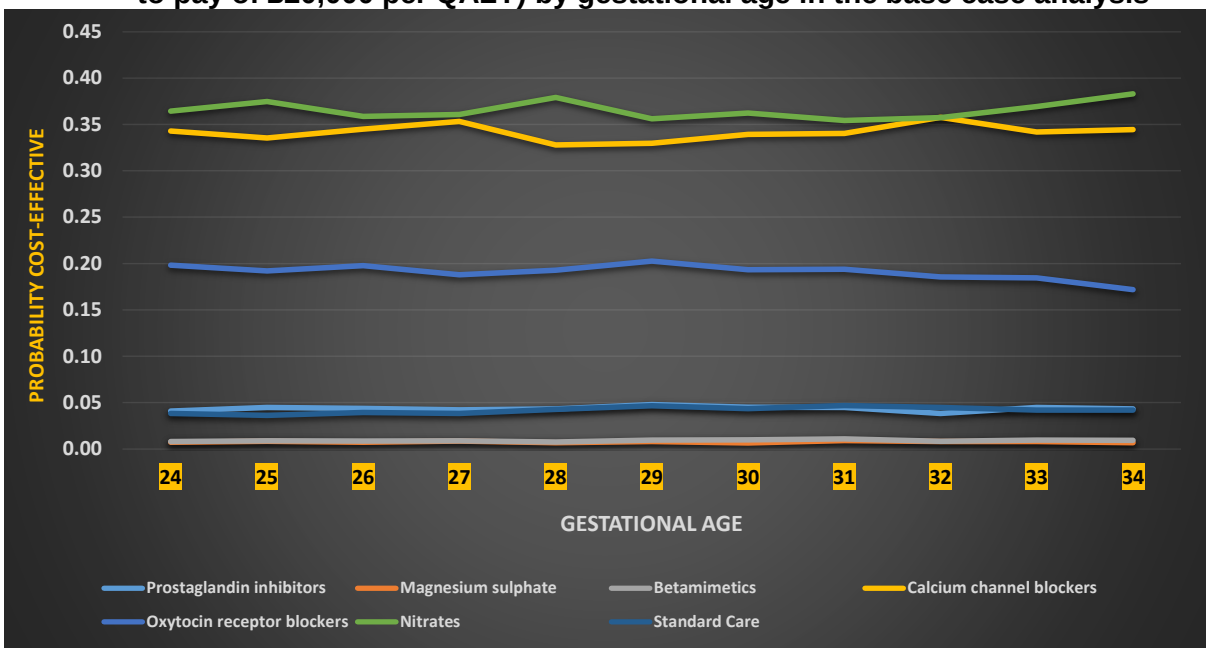
Figure 65: Cost-effectiveness acceptability curve for base case analysis and at a gestational age of 24 weeks

Figure 66: Chart to show net mean benefit by treatment and gestational age in the base case analysis



1

Figure 67: Chart to show probability a treatment is cost-effective (at a willingness to pay of £20,000 per QALY) by gestational age in the base case analysis



16.4.3.1 Sensitivity Analysis

A number of sensitivity analyses are presented below to illustrate how sensitive the model's results are to some key assumptions and input parameters.

16.4.3.1.1 Estimating the effect of treatment on mortality using the NMA undertaken on perinatal mortality

In the base case analysis the effect of treatment on mortality was estimated using the NMA on neonatal mortality. However, a NMA was also undertaken on the outcome of perinatal mortality, although for reasons outlined earlier it was not possible to combine neonatal and perinatal mortality outcomes. In this sensitivity analysis the effect of treatment on mortality is estimated from the relative treatment effects derived from the NMA on perinatal mortality and is based on 10,000 simulations.

Table 132: PSA result based on 10,000 simulations for a gestational age of 24 weeks with treatment effect on mortality estimated using data on relative treatment effect size from the NMA on perinatal mortality

Treatment	Mean cost	Mean QALY	Mean Net Benefit	Probability cost-effective	ICER
Nitrates	-£945	8.672	£174,381	0.89	Dominates
Calcium channel blockers	-£905	1.579	£32,491	0.03	Dominated
Prostaglandin inhibitors	-£250	1.742	£35,086	0.04	Dominated
Magnesium sulfate	-£205	-0.386	-£7,521	0.01	Dominated
Betamimetics	-£198	0.117	£2,354	0.00	Dominated
Standard care	£0	0	£0	0.00	Dominated
Oxytocin receptor blockers	£283	0.965	£19,018	0.02	Dominated

Table 132 reports the results of this sensitivity analysis in detail for women of a gestational age of 24 weeks with the cost-effectiveness acceptability curve for this analysis shown in Figure 68. This sensitivity analysis suggests that nitrates are the most cost-effective treatment dominating all alternatives, with the lowest mean costs and highest mean QALYs across the simulation. Allowing for uncertainty in model inputs, and treatment effect size in particular, they also have a very high probability of being the most cost-effective treatment alternative.

Figure 69 shows a plot of 1,000 simulations of the base case analysis on the cost-effectiveness plane with incremental costs and QALYs shown relative to standard care (origin) in women with a gestational age of 24 weeks. Figure 70 is a similar plot of the same 1,000 simulations but restricted to the 4 most cost-effective treatments as assessed by net mean benefit.

Figure 71 and Figure 72 respectively show how the net mean benefit and the probability that each treatment is cost-effective varies with gestational age. These suggest that whilst the relative cost-effectiveness of treatment diminishes with increasing gestational age, that nitrates continue to have a very high probability of being the most cost-effective treatment across all gestational ages considered in the model.

Figure 68: Cost-effectiveness acceptability curve using the NMA on neonatal mortality to estimate the relative treatment effect on mortality at a gestational age of 24 weeks

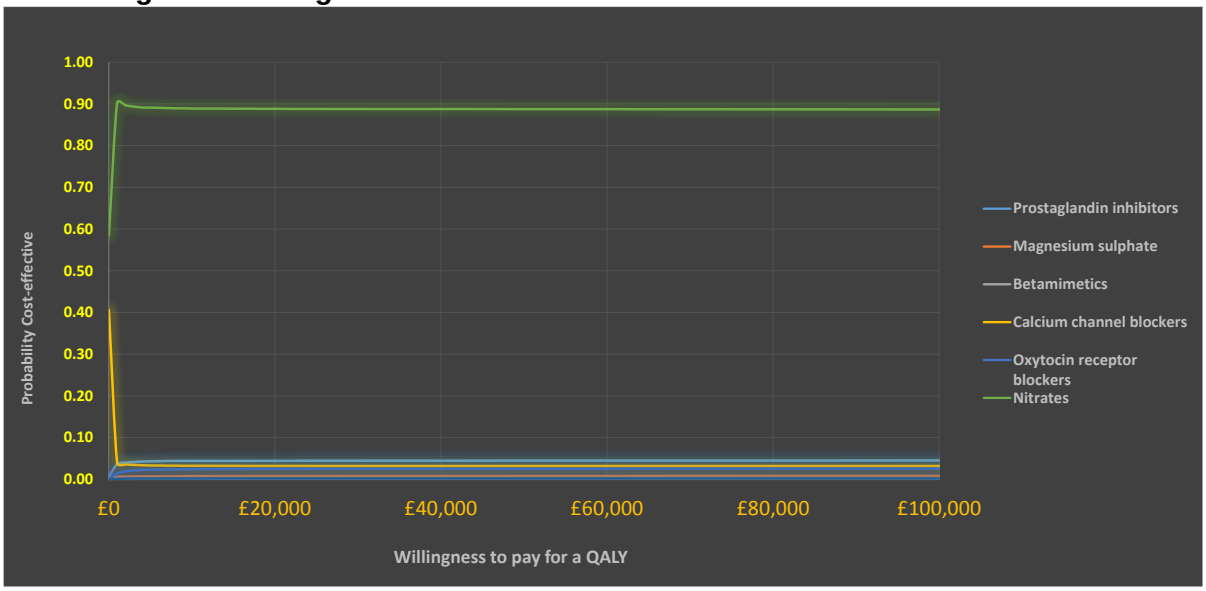
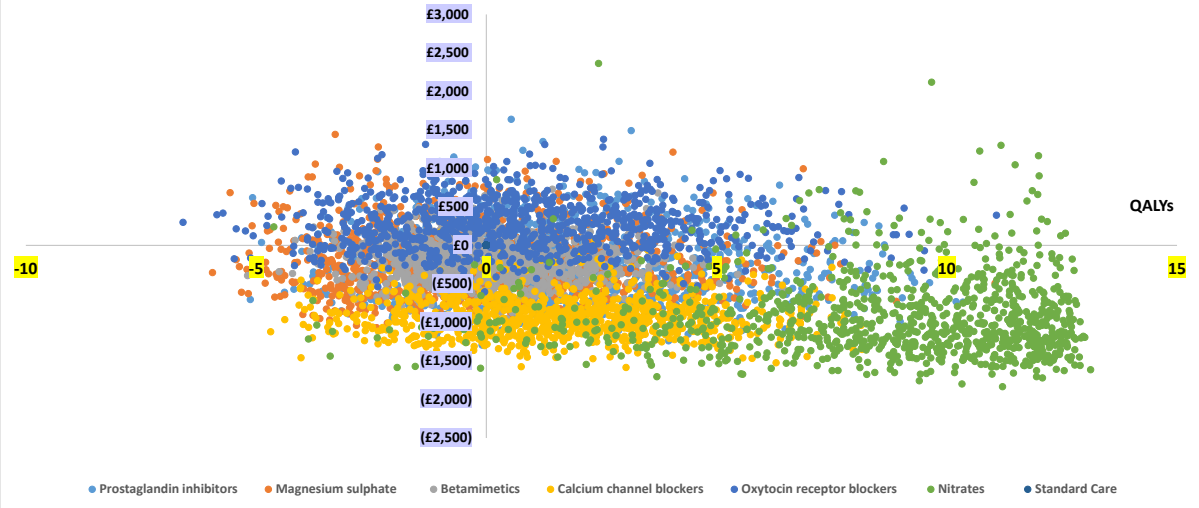
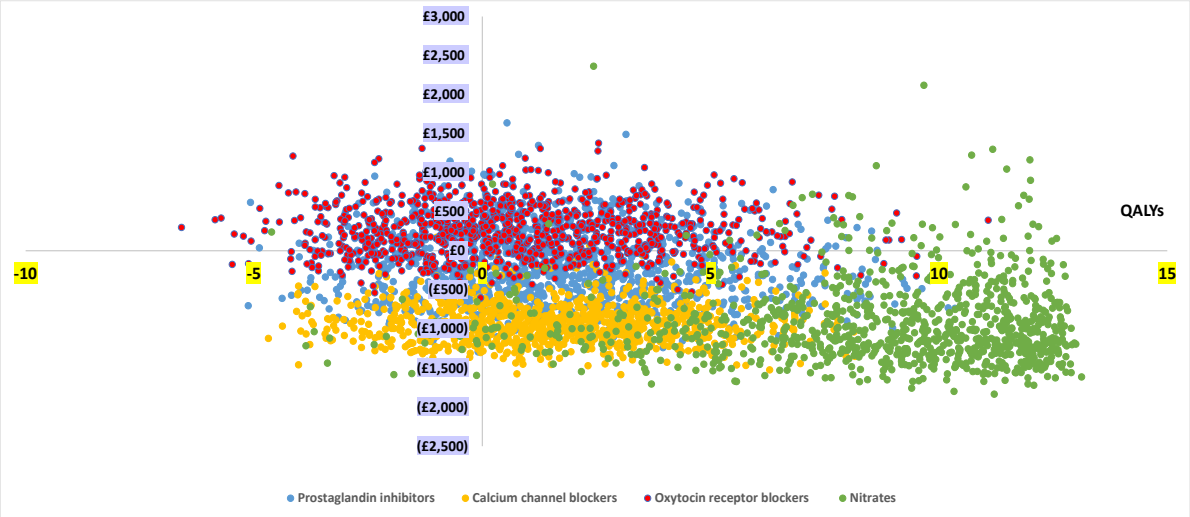


Figure 69: Cost-effectiveness plane with the relative treatment effect on mortality based on the NMA for perinatal mortality and showing all treatments in the analysis for women with a gestational age of 24 weeks



1

Figure 70: Cost-effectiveness plane with the relative treatment effect on mortality based on the NMA for perinatal mortality, restricted to the 4 most cost-effective treatments for women with a gestational age of 24 weeks



2

Figure 71: Chart to show net mean benefit by treatment and gestational age using the NMA on neonatal mortality to estimate the relative treatment effect on mortality

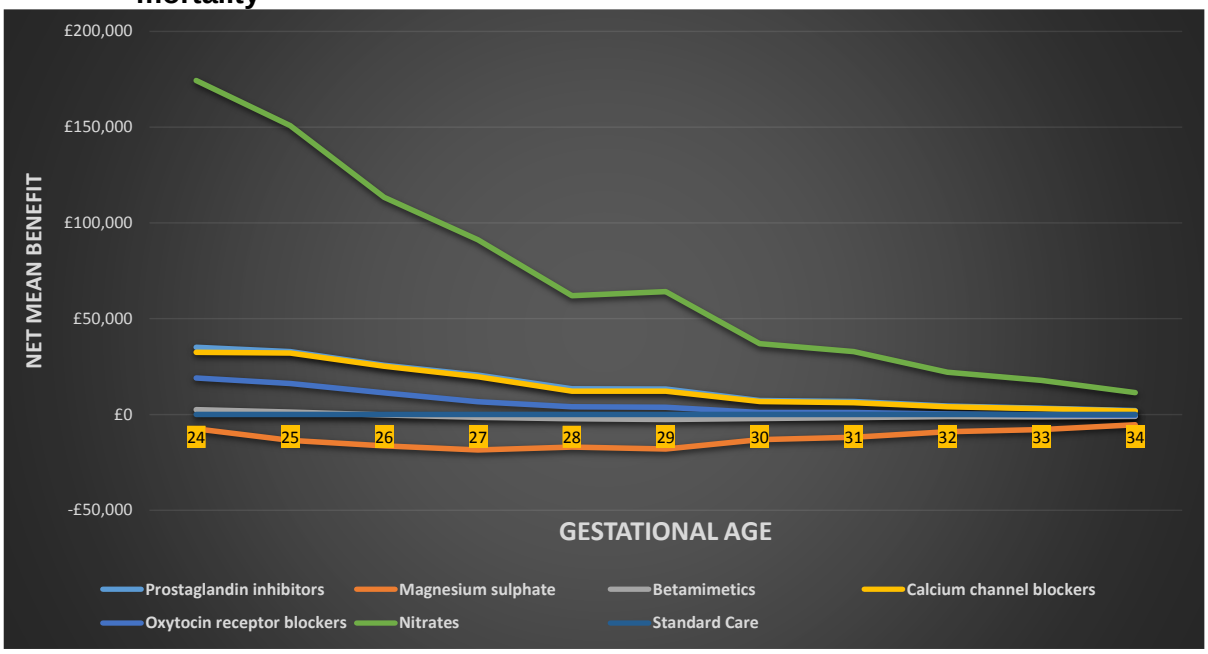
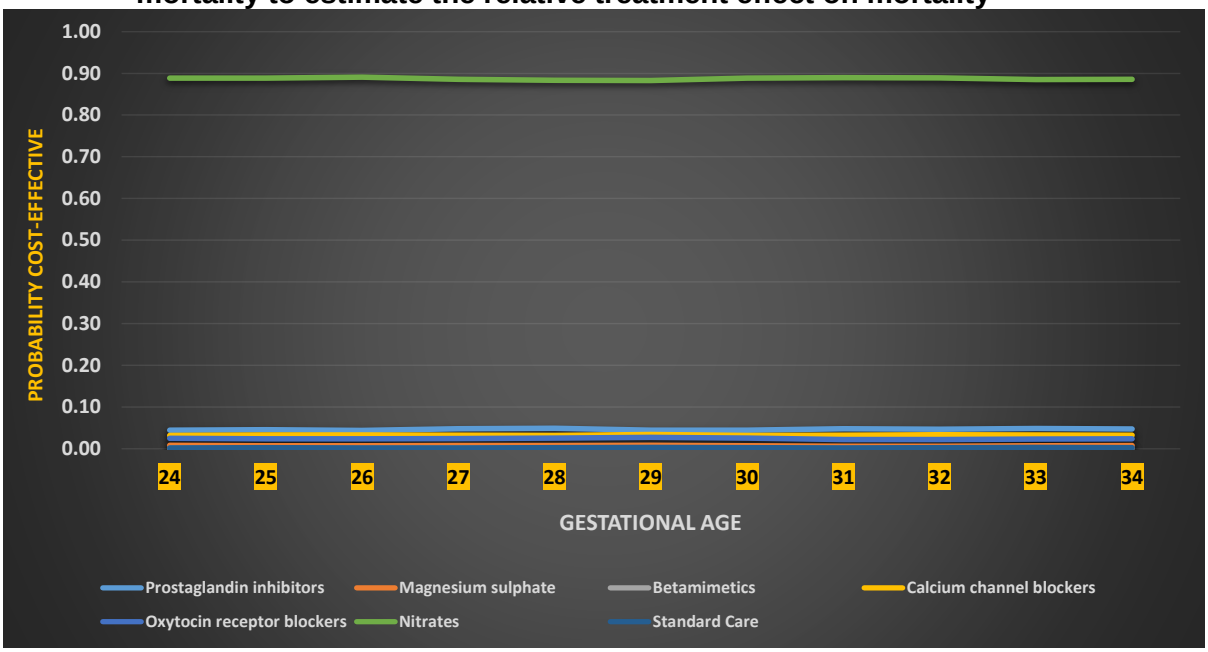


Figure 72: Chart to show probability a treatment is cost-effective (at a willingness to pay of £20,000 per QALY) by gestational age using the NMA on neonatal mortality to estimate the relative treatment effect on mortality



1*16.4.3.1.2 Varying the QALY loss from IVH and RDS*

In this sensitivity analysis the QALY loss from IVH and RDS were both reduced to 0.5 QALYs. Calcium channel blockers remained the most cost-effective treatment at 24 weeks and 34 weeks respectively with the highest net mean benefit. At 24 weeks the ICER for calcium channel blockers was £20 per QALY relative to nitrates which dominated other

treatment alternatives at that gestational age. At 34 weeks calcium channel blockers had an ICER of £25 per QALY relative to standard care whilst dominating all other treatments.

In addition another sensitivity analysis was performed in which the QALY loss from RDS and IVH was increased to 10 QALYs. At 24 weeks the ICER for calcium channel blockers was £24 per QALY relative to nitrates which dominated other treatment alternatives at that gestational age. At 34 weeks calcium channel blockers had an ICER of £24 per QALY relative to standard care whilst dominating all other treatments.

16.4.4 Discussion

One of the advantages of an economic analysis of this type is that is allows benefits across different outcome measures to be synthesised into a single measure of effect. In the base case analysis calcium channel blockers were found to be the most cost-effective intervention with the highest net mean benefit across 10,000 Monte Carlo simulations and a very low ICER relative to non-dominated treatment alternatives. Whilst the model showed that the cost-effectiveness declined with increasing gestational age calcium channel blockers remained the most cost-effective treatment across all gestational ages (see Figure 63). Interestingly, whilst calcium channel blockers had a similar probability of being the most cost effective to nitrates, it was the latter which had the highest probability of being the most cost-effective treatment. This apparent discrepancy reflected the wide confidence intervals reported for the nitrates treatment log odds ratios in the NMA as can be visualised in Figure 63 and Figure 64.

The base case model did not address the possible costs of diagnosis which can cause the costs of achieving particular outcomes to be under-estimated. Nor did it include the costs of hospitalisation as, with the possible exception of standard care, this cost would be identical across the different treatment alternatives. However, this model was used to inform the model that did consider the diagnosis of preterm labour in women with suspected preterm labour and intact membranes (see 16.2.1.1). That model found that treatment remained cost-effective even when including diagnostic costs, hospitalisation costs and the treatment of false positives. This finding is consistent with the net mean benefits observed for calcium channel blockers (see Figure 66).

Sensitivity analysis suggests that the model is not sensitive to the QALY values used for RDS and IVH outcomes. Unlike mortality the model does not vary the QALY loss from RDS and IVH with gestational age partly because of the difficulty with estimating how these QALY losses would vary with gestational age. Nevertheless from a clinical perspective, the prognosis from these outcomes is related to gestational age with lower gestational age at birth being associated with worse progress. However, this sensitivity analysis suggests that an approach which varied QALY loss by age for these outcomes would have a negligible impact on results.

It is also apparent from the differences in net mean costs that the model's results would not be particularly sensitive to differences in treatment costs, especially at the lower gestational ages, as the treatment cost represents only a small proportion of the difference in net mean benefit between the various treatment alternatives.

One sensitivity analysis which did have a huge bearing on the model's results was basing the relative treatment effect on mortality on the perinatal mortality NMA rather than the neonatal mortality network meta-analysis. This difference was driven by the NMA on perinatal mortality which found a very high probability that nitrates were the most effective treatment for this outcome. This was then reflected in this sensitivity analysis where nitrates were estimated to have an 89% probability of being the most cost-effective treatment. However, the Committee discussed that this benefit of nitrates needs to be balanced against the potential harm to the fetus. They noted that the number of trials including nitrates was small

and that that therefore the results needed be interpreted with caution. In addition, there was no NMA data for nitrates for the outcome of respiratory distress syndrome.

In terms of modelling the cost-effectiveness of treatment, mortality across the entire perinatal/neonatal mortality period is the real outcome of interest and that is reflected in how the baseline mortality risk was calculated. However, due to overlaps in definition it was not possible to generate a single NMA for the relative treatment effect covering the broader period due to issues of double counting. The recommendations of the Guideline Committee reflect that neonatal mortality was selected a priori as the more important outcome measure. Nevertheless, the fact that the 2 analyses produced different results should be considered as a limitation of the analysis and may reduce confidence in the result.

The cost-effectiveness results were driven largely by the outcomes of the NMA. In the methods section the choice of outcomes was explained as more than 3 outcomes were evaluated with network-meta analysis. One NMA that was not included was delay in preterm birth. In some respects this might be considered the best measure of tocolysis as the benefits of tocolysis are predicated on them achieving such a delay. In that NMA prostaglandin inhibitors came out as the most likely to be the effective treatment, had the largest point estimate of relative treatment effect and was significantly better than placebo for this outcome. However, it wasn't included in the economic model as it wasn't considered a 'hard' outcome come and it would be anticipated that a treatment that was more successful in delaying birth would also be the most successful at reducing adverse outcomes. Nevertheless, it is worth noting that even for the outcome of delay in preterm birth, prostaglandin inhibitors were not significantly better than calcium channel blockers, which were also significantly better than placebo for delay in preterm labour.

There are some limitations with this analysis. Baseline risks are based on populations who are may be in receipt of treatment including tocolysis to improve preterm outcomes and therefore may underestimate the risk of not providing treatment and may therefore similarly underestimate the absolute treatment effect which would tend to cause cost-effectiveness to be under-estimated.

The studies included in the network analyses did not go below 26 weeks but the Guideline Committee considered it would be reasonable to extrapolate relative treatment effects to women with a gestational age between 24-26 weeks. The model did not make a similar extrapolation for women over 34 weeks as the Committee did not think that any delay after this age would be considered sufficiently worthwhile. Furthermore, whilst not a gestational age considered explicitly in the model considering the diagnosis of preterm labour the evidence at 33 and 34 weeks was that treatment was less obviously cost-effective.

16.4.5 Conclusion

This model provides reasonably strong evidence that calcium channel blockers can be considered as a cost-effective tocolytic treatment for women with diagnosed or suspected preterm labour between a gestational age of 24-34 weeks. They are additionally a relatively cheap tocolytic and the guideline Committee thought that they were often a first line treatment already.

17 Terms, glossary and abbreviations

17.1 Terms

Term	Definition
Symptoms of preterm labour	A woman has presented before 37^{+0} weeks of pregnancy reporting symptoms that might be indicative of preterm labour (such as abdominal pain), but no clinical assessment (including speculum or digital vaginal examination) has taken place.
Suspected preterm labour	A woman is in suspected preterm labour if she has reported symptoms of preterm labour and has had a clinical assessment (including a speculum or digital vaginal examination) that confirms the possibility of preterm labour but rules out established labour.
Diagnosed preterm labour	A woman is in diagnosed preterm labour if she is in suspected preterm labour and has had a positive diagnostic test for preterm labour.
Established preterm labour	A woman is in established preterm labour if she has progressive cervical dilatation from 4 cm with regular contractions (see recommendation 1.3.1 in the NICE guideline on intrapartum care).
Preterm prelabour rupture of membranes (P-PROM)	A woman is described as having P-PROM if she has ruptured membranes before 37^{+0} weeks of pregnancy but is not in established labour.
'Rescue' cervical cerclage	Cervical cerclage performed as an emergency procedure in a woman with premature cervical dilatation and often with exposed fetal membranes.

17.2 Glossary

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper
Accelerations (fetal monitoring)	An abrupt increase in fetal heart rate above baseline with onset to peak of the acceleration less than < 30 seconds and less than 2 minutes in duration.
Acidosis	An increased acidity in the blood and other body tissue.
Active management of the third stage	A package of care comprising the following components: • routine use of drugs to cause contraction of the uterus • clamping and cutting of the cord • controlled cord traction after signs of separation of the placenta
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Amniotic fluid	The protective liquid surrounding the baby within the amniotic sac of a pregnant woman.
Antenatal antibiotic prophylaxis	The use of antibiotics to prevent infections in antenatal care.
Antepartum haemorrhage	Bleeding from or in to the genital tract, occurring from 24^{+0} weeks of pregnancy and prior to the birth of the baby.
Apgar score	A measure of the physical condition of a newborn infant.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.

Term	Definition
Arm (of a clinical study)	Subsection of individuals within a study who receive 1 particular intervention, for example placebo arm
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.
Attrition bias	Systematic differences between comparison groups in withdrawals or exclusion of participants from a study.
Available case analysis	Analysis of data that is available for participants at the end of follow-up.
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Baseline variability (fetal monitoring)	Fluctuations in the fetal heart rate of more than 2 cycles per minute
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Bishop score	A pre-labour scoring system based on clinical examination of the cervix, to assist in predicting whether induction of labour will be required
Bradycardia (fetal monitoring)	Slow heart rate; for the term fetus, this is defined as a heart rate of less than 110 beats per minute.
Breech (presentation)	a baby which is so positioned in the womb that the buttocks or feet are delivered first
Bronchopulmonary dysplasia	A chronic lung disorder of infants and children
Bulging membranes	Amniotic membranes bulging through the opening of the cervix.
Caesarean section	A surgical operation for delivering a baby by cutting through the wall of the mother's abdomen. This may be an elective (planned) or emergency procedure
Cardiotocography	Electronic recording of the fetal heart rate using either a Doppler ultrasound transducer strapped to the woman's abdomen, or an electrode attached to the fetal scalp, plus a second toco transducer strapped to the woman's abdomen to record uterine contractions.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition. For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.

Term	Definition
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Cephalic (presentation)	A baby so positioned in the womb that the head is delivered first
Cerebral palsy	The general term for a number of neurological conditions that affect movement and co-ordination
Cervical cerclage	A surgical treatment for cervical incompetence or insufficiency
Cervical shortening	A condition in pregnant women where the cervix becomes softer and weaker than normal
Cervical trauma	Physical injury to the cervix
Chorioamnionitis	An inflammation of the fetal membranes (amnion and chorion) due to a bacterial infection
Chronic lung disease	A general term for long-term respiratory problems in premature babies
Clinical audit	A systematic process for setting and monitoring standards of clinical care. Whereas 'guidelines' define what the best clinical practice should be, 'audit' investigates whether best practice is being carried out. Clinical audit can be described as a cycle or spiral. Within the cycle there are stages that follow a systematic process of establishing best practice, measuring care against specific criteria, taking action to improve care and monitoring to sustain improvement. The spiral suggests that as the process continues, each cycle aspires to a higher level of quality.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cognitive dysfunction	The loss of intellectual functions such as thinking, remembering, and reasoning of sufficient severity to interfere with daily functioning
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	The process used to ensure that the person deciding to enter a participant into a randomised controlled trial does not know the comparison group into which that individual will be allocated. This is distinct from blinding, and is aimed at preventing selection bias. Some attempts at concealing allocation are more prone to manipulation than others, and the method of allocation concealment is used as an assessment of the quality of a trial.
Confidence interval (CI)	There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how certain we

Term	Definition
	are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that 'based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110'. In such a case the 95% CI would be 110 to 150. A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment - often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).
Confounding factor	Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore age is a confounding factor.
Consensus methods	Techniques used to reach agreement on a particular issue. Consensus methods may be used to develop NICE guidance if there is not enough good quality research evidence to give a clear answer to a question. Formal consensus methods include Delphi and nominal group techniques.
Continuous outcome	Data with a potentially infinite number of possible values within a given range. Height, weight and blood pressure are examples of continuous variables.
Control group	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.
Cord milking	After delivery, the caregiver holds the cord and squeezes blood down the cord into the baby
Cord prolapse	When the umbilical cord comes out of the uterus with or before the presenting part of the fetus
Corticosteroids	An anti-inflammatory medicine
Cost–benefit analysis (CBA)	Cost-benefit analysis is 1 of the tools used to carry out an economic evaluation. The costs and benefits are measured using the same monetary units (for example, pounds sterling) to see whether the benefits exceed the costs.
Cost–consequences analysis (CCA)	Cost-consequence analysis is 1 of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost-benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is

Term	Definition
	left to decision-makers to determine whether, overall, the treatment is worth carrying out
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is 1 of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost–utility analysis (CUA)	Cost-utility analysis is 1 of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
COX proportional hazard Model	In survival analysis, a statistical model that asserts that the effect of the study factors (for example the intervention of interest) on the hazard rate (the risk of occurrence of an event) in the study population is multiplicative and does not change over time.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decelerations (fetal monitoring)	A decrease in the fetal heart rate below the baseline rate
Decision analysis	An explicit quantitative approach to decision-making under uncertainty based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Decision to delivery interval	The time taken between the decision to expedite a birth and the birth.
Delayed cord clamping	A birth practice where the umbilical cord is not clamped or cut until after pulsations have ceased, or until after the placenta is delivered
Dichotomous outcomes	Outcome that can take 1 of 2 possible values, such as dead/alive, smoker/non-smoker, present/not present (also called binary data).
Dilated cervix	Open cervix
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be 'dominated' by the Alternative
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits - health effects - relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals.

Term	Definition
	There are several types of economic evaluation: cost-benefit analysis, cost consequence analysis, cost-effectiveness analysis, cost-minimisation analysis and cost-utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Early cord clamping	Clamping carried out in the first 60 seconds after birth.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in 1 group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just hap
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Established labour	Labour is established when there is both of: • regular painful contractions, and • progressive cervical effacement (thinning) and dilatation beyond 4 cm.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality-of-life. It provides a single index value for health status.
Equivalence study	A trial designed to determine whether the response to 2 or more treatments differs by an amount that is clinically unimportant. This is usually demonstrated by showing that the true treatment difference is likely to lie between a lower and an upper equivalence level of clinically acceptable differences.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients)
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extended dominance	If Option A is both more clinically effective than Option B and has a lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore more cost effective and should be preferred, other things remaining equal.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Expectant management	Awaiting events to take their natural course. This would usually include observation of the woman and/or baby's condition

Term	Definition
Exposed membranes	When the cervix opens, the membranes are exposed'
Fetal blood sampling	A technique to measure the level of acid-base status of the baby's blood. A sample of blood is taken from the baby's scalp and either the pH or lactate value is measured. It is used as an adjunct to cardiotocography to help to clarify whether the baby is developing an acidosis when may cause additional interventions to be required.
Fetal monitoring	Method used to monitor the fetal heartbeat during labour
Fetal fibronectin	A fibronectin protein produced by fetal cells
Fetal growth restriction	A condition where growth of the fetus slows or ceases when it is in the uterus
Fixed-effect model	In meta-analysis, a model that calculates a pooled effect estimate using the assumption that all observed variation between studies is caused by the play of chance. Studies are assumed to be measuring the same overall effect.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Forest plot	A graphical representation of the individual results of each study included in a meta-analysis together with the combined meta-analysis result. The plot also allows readers to see the heterogeneity among the results of the studies. The results of individual studies are shown as squares centred on each study's point estimate. A horizontal line runs through each square to show each study's confidence interval. The overall estimate from the meta-analysis and its confidence interval are shown at the bottom, represented as a diamond. The centre of the diamond represents the pooled point estimate, and its horizontal tips represent the confidence interval.
Gestation	The period of development in the uterus from conception until birth
Gestational age	A term used during pregnancy to describe how far along the pregnancy is, measured in weeks
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Gross motor dysfunction	Dysfunction in the movement of the large muscles of the body
Harms	Adverse effects of an intervention.
Hazard ratio	A graphical representation of the individual results of each study included in a meta-analysis together with the combined meta-analysis result. The plot also allows readers to see the heterogeneity among the results of the studies. The results of individual studies are shown as squares centred on each study's point estimate. A horizontal line runs through each square to show each study's confidence interval. The overall estimate from the meta-analysis and its confidence interval are shown at the bottom, represented as a diamond. The centre of the diamond represents the pooled point estimate, and its horizontal tips represent the confidence interval.

Term	Definition
Health economics	Study or analysis of the cost of using and distributing healthcare resources.
Health-related quality-of-life (HRQoL)	A measure of the effects of an illness to see how it affects someone's day to-day life.
Heterogeneity	The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental cost	The extra cost linked to using 1 test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for 1 treatment compared with another
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: (£20,000 x QALYs gained) – Incremental cost.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Induction of labour	A procedure where the midwife or doctor starts labour artificially using a membrane sweep, pessary, or a hormone drip.
Infant death	The death of a child less than 1 year of age
Insulin-like growth factor binding protein-1	A protein that in humans is encoded by the IGFBP1 gene
Instrumental birth	Birth in which the use of instruments is required
Intellectual delay	A disability characterised by significant limitations both in intellectual functioning (reasoning, learning, problem solving) and in adaptive behaviour, which covers a range of everyday social and practical skills
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Intermittent auscultation	Intermittent measurement of the fetal heart rate using a Doppler ultrasound or a Pinard stethoscope.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Intracranial haemorrhage	Bleeding within the skull cavity or brain

Term	Definition
Intraventricular haemorrhage	Bleeding into the brain's ventricular system, where the cerebrospinal fluid is produced and circulates
Haematocrit	The volume percentage of red blood cells in blood
Hyperbilirubinaemia	A condition in which there is too much bilirubin in the blood
Hypoxia	A condition in which the body or a region of the body is deprived of adequate oxygen supply
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance
Labour	The process of delivering a baby and the placenta, membranes, and umbilical cord from the uterus to the vagina to the outside world
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).
Liquor	The protective liquid contained by the amniotic sac of a pregnant woman
Long-term infant morbidity	The rate of illness and disease in children
Loss to follow-up	Patients who have withdrawn from the clinical trial at the point of follow-up
Low birth weight	A birth weight of a live born infant of less than 2,500 g (5 pounds 8 ounces)
Magnesium sulfate	An inorganic salt containing magnesium, sulfur and oxygen, with the formula MgSO4
Markov model	A method for estimating long-term costs and effects for recurrent or chronic conditions, based on health states and the probability of transition between them within a given time period (cycle).
McDonald suture	A purse-string stitch used to cinch the cervix shut
Mean	An average value, calculated by adding all the observations and dividing by the number of observations.
Mean difference	In meta-analysis, a method used to combine measures on continuous scales (such as weight), where the mean, standard deviation and sample size in each group are known. The weight given to the difference in means from each study (e.g. how much influence each study has on the overall results of the meta-analysis) is determined by the precision of its estimate of effect.
Mechanical ventilation	A technique in which gas is moved toward and from the lungs through an external device connected directly to the patient.
Median	The value of the observation that comes half way when the observations are ranked in order.
Meta-analysis	A method often used in systematic reviews. Results from several studies of the same test or treatment are combined to estimate the overall effect of the treatment.
Mid-trimester loss	The death of a fetus in the second trimester (3-6 months of pregnancy)
Minimal important difference (MID)	Thresholds for clinical importance, which represent minimal important differences for benefit or for harm; e.g. the threshold at which drug A is less effective than drug B by an amount that is clinically important to patients.
Multiple pregnancy	A pregnancy in which there is more than 1 fetus
Multivariate model	A statistical model for analysis of the relationship between 2 or more

Term	Definition
	predictor (independent) variables and the outcome (dependent) variable.
Necrotising enterocolitis	A medical condition primarily seen in preterm infants, where portions of the bowel undergo necrosis
Neonatal death	The death of a baby within the first 28 days of life
Neonatal intensive care	Intensive-care for ill or preterm newborn infants
Net monetary benefit (NMB)	The value (usually in monetary terms) of an intervention net of its cost. The NMB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the NMB is calculated as: (£20,000 x QALYs gained) – cost.
Network meta-analysis	Meta-analysis in which multiple treatments (that is, 3 or more) are being compared using both direct comparisons of interventions within randomised controlled trials and indirect comparisons across trials based on a common comparator.
Neurodevelopmental delay	Disabilities in the functioning of the brain that affect a child's behaviour, memory or ability to learn
Neonatal morbidity	Health disorders in neonates occurring the first 4 weeks of life
Nitrazine	A pH indicator dye
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in 1 characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio (OR)	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in 1 group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups - in this case, 1 of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, relative risk, risk ratio.
Odds ratio (OR)	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in 1

Term	Definition
	group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups - in this case, 1 of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, relative risk, risk ratio.
Opportunity cost	The loss of other health care programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that 1 seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Performance bias	Systematic differences between intervention groups in care provided apart from the intervention being evaluated. Blinding of study participants (both the recipients and providers of care) is used to protect against performance bias.
Perinatal death	Death occurring after 24 completed weeks of pregnancy and within 7 days after birth
Periventricular leucomalacia	A form of white-matter brain injury, characterised by the necrosis of white matter near the lateral ventricles
Physiological management of the third stage	A package of care comprising the following components: • no routine use of uterotonic drugs

Term	Definition
	• no clamping of the cord until pulsation has stopped • delivery of the placenta by maternal effort.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had - over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Placebo effect	A beneficial (or adverse) effect produced by a placebo and not due to any property of the placebo itself.
Placental abruption	A complication of pregnancy where the placenta has separated from the uterus of the mother
Placental alpha microglobulin	A human protein that was first isolated in amniotic fluid
Planned preterm birth	The planned birth of an infant before 37 weeks of pregnancy due to medical complications
Post-hoc analysis	Statistical analyses that are not specified in the trial protocol, and are generally suggested by the data.
Postpartum haemorrhage	Blood loss over 500 ml from the vagina following labour.
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Preterm birth	The birth of an infant before 37 weeks of pregnancy
Preterm labour	Regular contractions of the uterus resulting in changes in the cervix that start before 37 weeks of pregnancy
Preterm pre-labour rupture of membranes	Rupture of the membranes before 37 weeks of pregnancy, occurring before the onset of labour
Pre-eclampsia	A disorder of pregnancy characterised by high blood pressure and a large amount of protein in the urine
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Product licence	An authorisation from the MHRA to market a medicinal product.
Progesterone	A steroid hormone released by the corpus luteum that stimulates the uterus to prepare for pregnancy
Prognosis	A probable course or outcome of a disease. Prognostic factors are patient or disease characteristics that influence the course. Good prognosis is associated with low rate of undesirable outcomes; poor prognosis is associated with a high rate of undesirable outcomes.
Prophylactic antibiotics	Antibiotics used for the prevention of infection complications
Prophylactic cervical cerclage	A treatment for cervical weakness (also termed cervical incompetence or insufficiency) to prevent preterm birth and miscarriage
Prophylactic progesterone	Progesterone by vaginal suppository to reduce the incidence of spontaneous preterm birth

Term	Definition
Prospective study	A research study in which the health or other characteristic of participants is monitored (or 'followed up') for a period of time, with events recorded as they happen. This contrasts with retrospective studies.
Publication bias	Publication bias occurs when researchers publish the results of studies showing that a treatment works well and don't publish those showing it did not have any effect. If this happens, analysis of the published results will not give an accurate idea of how well the treatment works. This type of bias can be assessed by a funnel plot.
Puerperal sepsis	Serious infection affecting the mother after giving birth
Pyrexia	A fever
'Rescue' cervical cerclage	An emergency procedure to prevent preterm delivery when there is cervical dilation
Respiratory distress syndrome	A syndrome in premature infants caused by developmental insufficiency of surfactant production and structural immaturity in the lungs
Sepsis	A whole-body inflammation caused by an infection
Shirodkar suture	A non-absorbable stitch that is inserted and put around the cervix to hold it closed
Special care baby unit	A unit taking premature and term babies who do not require intensive care, but are unable to be cared for on a normal ward
Speculum examination	A method for visualizing the cervix (the opening of the uterus), and the interior walls of the vagina, using an instrument
Spontaneous preterm labour	Regular contractions of the uterus resulting in changes in the cervix that start before 37 weeks of pregnancy that occur with no intervention
Stakeholder	An organisation with an interest in a topic that NICE is developing a clinical guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: □ manufacturers of drugs or equipment □ national patient and carer organisations □ NHS organisations □ organisations representing healthcare professionals.
Standard deviation (SD)	A measure of the spread or dispersion of a set of observations, calculated as the average difference from the mean value in the sample.
Still birth	The death of a baby after 24 weeks of pregnancy but before birth.
Subgroup analysis	An analysis in which the intervention effect is evaluated in a defined subset of the participants in a trial, or in complementary subsets.
Systematic review (SR)	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Tachycardia (fetal monitoring)	Rapid heart rate; for the term fetus, this is defined as a heart rate of over 160 beats per minute.
Third stage of labour	The interval from the birth of the baby to the expulsion of the placenta and membranes.
Time horizon	The time span over which costs and health outcomes are considered in adecision analysis or economic evaluation.
Tocolytic	A drug used to prevent or lessen uterine contractions.
Transfer	This term indicates where responsibility for the woman's care passes from one healthcare professional to another. This may or may not also involve a physical transfer of the woman from one birth setting to another.

Term	Definition
Transvaginal ultrasound	An internal ultrasound scan to look at a women's reproductive system
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis which separately explores each variable in a data set.
Upper uterine segment	It is the portion of the uterus above the bladder edge. The lower segment is the portion of the uterus normally covered anteriorly by the bladder; the lower segment is not well formed until the last trimester.
Uterotonic	A drug used to induce uterine contractions.
Utility	In health economics, a 'utility' is the measure of the preference or value that an individual or society places upon a particular health state. It is generally a number between 0 (representing death) and 1 (perfect health). The most widely used measure of benefit in cost-utility analysis is the quality-adjusted life year, but other measures include disability-adjusted life years (DALYs) and healthy year equivalents (HYEs).
Vaginal birth	The birth of a baby through the vagina.

17.3 Abbreviations

Abbreviation	Definition
CS	Caesarean section
CTG	Cardiotocography
FHR	Fetal heart rate
FBS	Fetal Blood Sampling
FEM	Fetal Electronic Monitoring
P-PROM	Preterm Premature Rupture of Membranes
PTLB	Preterm Labour and Birth
PLB	Preterm Labour

18 References

Alfirevic 2012

Alfirevic,Z., Stampalija,T., Roberts,D., Jorgensen,A.L., Cervical stitch (cerclage) for preventing preterm birth in singleton pregnancy, Cochrane Database of Systematic Reviews, 4, CD008991-, 2012

Alfirevic 2013

Alfirevic,Zarko, Milan,Stephen J., Livio,Stefania, Caesarean section versus vaginal delivery for preterm birth in singletons, Cochrane Database of Systematic Reviews, -, 2013

Althaus 2005

Althaus,J.E., Petersen,S.M., Fox,H.E., Holcroft,C.J., Graham,E.M., Can electronic fetal monitoring identify preterm neonates with cerebral white matter injury?, Obstetrics and Gynecology, 105, 458-465, 2005

Althuisius 2003

Althuisius,S.M., Dekker,G.A., Hummel,P., van Geijn,H.P., Cervical inc, Cervical incompetence prevention randomized cerclage trial: emergency cerclage with bed rest versus bed rest alone, American Journal of Obstetrics and Gynecology, 189, 907-910, 2003

Aoki 2014

Aoki,S., Ohnuma,E., Kurasawa,K., Okuda,M., Takahashi,T., Hirahara,F., Emergency cerclage versus expectant management for prolapsed fetal membranes: a retrospective, comparative study, Journal of Obstetrics and Gynaecology Research, 40, 381-386, 2014

Asztalos 2013

Asztalos, E. V., Murphy, K. E., Willan, A. R., Matthews, S. G., Ohlsson, A., Saigal, S., Armson, B. A., Kelly, E. N., Delisle, M. F., Gafni, A., Lee, S. K., Sananes, R., Rovet, J., Guselle, P., Amankwah, K., Saleem, M., Sanchez, J., Macs- Collaborative Group, Multiple courses of antenatal corticosteroids for preterm birth study: outcomes in children at 5 years of age (MACS-5), JAMA Pediatrics, 167, 1102-10, 2013

Atarod 2014

Atarod,Z., Taghipour,M., Roohanizadeh,H., Fadavi,S., Taghavipour,M., Effects of single course and multicourse betamethasone prior to birth in the prognosis of the preterm neonates: A randomized, double-blind placebo-control clinical trial study, Journal of Research in Medical Sciences, 19, 715-719, 2014

Azlin 2010

Azlin,M.I., Bang,H.K., An,L.J., Mohamad,S.N., Mansor,N.A., Yee,B.S., Zulkifli,N.H., Tamil,A.M., Role of phIGFBP-1 and ultrasound cervical length in predicting pre-term labour, Journal of Obstetrics and Gynaecology, 30, 456-459, 2010

Bagga 2010

Bagga,R., Takhtani,M., Suri,V., Adhikari,K., Arora,S., Bhardwaj,S., Cervical length and cervicovaginal HCG for prediction of pre-term birth in women with signs and symptoms of pre-term labour, Journal of Obstetrics and Gynaecology, 30, 451-455, 2010

Bartnicki 1996

Bartnicki,J., Casal,D., Kreaden,U.S., Saling,E., Vetter,K., Fetal fibronectin in vaginal specimens predicts preterm delivery and very-low-birth-weight infants, American Journal of Obstetrics and Gynecology, 174, 971-974, 1996

Benattar 1997

Benattar,C., Taieb,J., Fernandez,H., Lindendaum,A., Frydman,R., Ville,Y., Rapid fetal fibronectin swab-test in preterm labor patients treated by betamimetics, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 72, 131-135, 1997

Berghella 2011

Berghella,V., Rafael,T.J., Szychowski,J.M., Rust,O.A., Owen,J., Cerclage for short cervix on ultrasonography in women with singleton gestations and previous preterm birth: a meta-analysis, Obstetrics and Gynecology, 117, 663-671, 2011

Bickford 2013

Bickford,C.D., Magee,L.A., Mitton,C., Kruse,M., Synnes,A.R., Sawchuck,D., Basso,M., Senikas,V.M., von,Dadelszen P., MAG-CP Working Group., Magnesium sulphate for fetal neuroprotection: a cost-effectiveness analysis, BMC Health Services Research, 13, 527-, 2013

Blencowe 2013

Blencowe H, Lee AC, Cousens S, Bahalim A, Narwal R, Zhong N, Chou D, Say L, Modi N, Katz J, Vos T, Marlow N, Lawn JE. Preterm birth-associated neurodevelopmental impairment estimates at regional and global levels for 2010. Pediatric research. 2013 Dec;74 Suppl 1:17-34.

Botsis 2006

Botsis,D., Makrakis,E., Papagianni,V., Kouskouni,E., Grigoriou,O., Dendrinos,S., Creatsas,G., The value of cervical length and plasma proMMP-9 levels for the prediction of preterm delivery in pregnant women presenting with threatened preterm labor, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 128, 108-112, 2006

Bowes 1980

Bowes,W.A.,Jr., Gabre,S.G., Bowes,C., Fetal heart rate monitoring in premature infants weighing 1,500 grams or less, American Journal of Obstetrics and Gynecology, 137, 791-796, 1980

Braithwaite 1986

Braithwaite,N.D.J., Milligan,J.E., Shennan,A.T., Fetal heart rate monitoring and neonatal mortality in the very preterm infant, American Journal of Obstetrics and Gynecology, 154, 250-254, 1986

Brik 2010

Brik,M., Hernandez,A.I., Pedraz,C.C., Perales,A., Phosphorylated insulin-like growth factor binding protein-1 and cervical measurement in women with threatening preterm birth, Acta Obstetricia et Gynecologica Scandinavica, 89, 268-274, 2010

British Association of Perinatal Medicine 2008

British Association of Perinatal Medicine, The Management of Babies born Extremely Preterm at less than 26 weeks of gestation, A Framework for Clinical Practice at the time of Birth, 2008

Report of a Working Group

Burrus 1994

Burrus,D.R., O'Shea,T.M.,Jr., Veille,J.C., Mueller-Heubach,E., The predictive value of intrapartum fetal heart rate abnormalities in the extremely premature infant, American Journal of Obstetrics and Gynecology, 171, 1128-1132, 1994

Burwick 2011

Burwick,R.M., Zork,N.M., Lee,G.T., Ross,M.G., Kjos,S.L., Cervilenz assessment of cervical length compared to fetal fibronectin in the prediction of preterm delivery in women with threatened preterm labor, Journal of Maternal-Fetal and Neonatal Medicine, 24, 127-131, 2011

Cahill 2010

Cahill,A.G., Odibo,A.O., Caughey,A.B., Stamilio,D.M., Hassan,S.S., Macones,G.A., Romero,R., Universal cervical length screening and treatment with vaginal progesterone to prevent preterm birth: a decision and economic analysis, American Journal of Obstetrics and Gynecology, 202, 548-548, 2010

Cahill 2011

Cahill,A.G., Odibo,A.O., Stout,M.J., Grobman,W.A., Macones,G.A., Caughey,A.B., Magnesium sulfate therapy for the prevention of cerebral palsy in preterm infants: a decision-analytic and economic analysis, American Journal of Obstetrics and Gynecology, 205, 542-547, 2011

Carroll 1995

Carroll,S.G., Papaioannou,S., Nicolaides,K.H., Assessment of fetal activity and amniotic fluid volume in the prediction of intrauterine infection in preterm prelabor amniorrhexis, American Journal of Obstetrics and Gynecology, 172, 1427-1435, 1995

Cazan-London 2005

Cazan-London,G., Mozurkewich,E.L., Xu,X., Ransom,S.B., Willingness or unwillingness to perform cesarean section for impending preterm delivery at 24 weeks' gestation: a cost-effectiveness analysis, American Journal of Obstetrics and Gynecology, 193, 1187-1192, 2005

Colbourn 2007

Colbourn,T., Asseburg,C., Bojke,L., Philips,Z., Claxton,K., Ades,A.E., Gilbert,R.E., Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: Cost-effectiveness and expected value of information analyses, Health Technology Assessment, 11, 21-108, 2007

Costeloe 2012

Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: comparison of 2 birth cohorts in 1995 and 2006 (the EPICure studies). BMJ (Clinical research ed. 2012;345:e7976.

MacKay 2010

MacKay DF, Smith GC, Dobbie R, Pell JP. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. PLoS medicine. 2010;7(6):e1000289.

Crowther 2003

Crowther,C.A., Hiller,J.E., Doyle,L.W., Haslam,R.R., Australasian Collaborative Trial of Magnesium Sulphate (ACTOMg SO, Effect of magnesium sulfate given for neuroprotection before preterm birth: a randomized controlled trial, JAMA, 290, 2669-2676, 2003

Crowther 2013

Crowther,Caroline A., McKinlay,JD Christopher, Middleton,Philippa, Harding,Jane E., Repeat doses of prenatal corticosteroids for women at risk of preterm birth for improving neonatal health outcomes, Cochrane Database of Systematic Reviews, -, 2013

Curti 2012

Curti,A., Simonazzi,G., Farina,A., Mehmeti,H., Facchinetti,F., Rizzo,N., Exam-indicated cerclage in patients with fetal membranes at or beyond external os: a retrospective evaluation, Journal of Obstetrics and Gynaecology Research, 38, 1352-1357, 2012

Danti 2011

Danti,L., Prefumo,F., Lojacono,A., Corini,S., Testori,A., Frusca,T., The combination of short cervical length and phIGFBP-1 in the prediction of preterm delivery in symptomatic women, Journal of Maternal-Fetal and Neonatal Medicine, 24, 1262-1266, 2011

Daskalakis 2006

Daskalakis,G., Papantoniou,N., Mesogitis,S., Antsaklis,A., Management of cervical insufficiency and bulging fetal membranes, Obstetrics and Gynecology, 107, 221-226, 2006

Del Valle 1992

Del Valle,G.O., Joffe,G.M., Izquierdo,L.A., Smith,J.F., Gilson,G.J., Curet,L.B., The biophysical profile and the nonstress test: poor predictors of chorioamnionitis and fetal infection in prolonged preterm premature rupture of membranes, Obstetrics and Gynecology, 80, 106-110, 1992

Demirci 2011

Demirci,O., Unal,A., Demirci,E., Sozen,H., Akdemir,Y., Boybek,E., Ertekin,A., Sonographic measurement of cervical length and risk of preterm delivery, Journal of Obstetrics and Gynaecology Research, 37, 809-814, 2011

Diaz 2009

Diaz,J., Chedraui,P., Hidalgo,L., Medina,M., The clinical utility of fetal fibronectin in the prediction of pre-term birth in a low socio-economic setting hospital in Ecuador, Journal of Maternal-Fetal and Neonatal Medicine, 22, 89-93, 2009

Dodd 2013

Dodd,Jodie M., Jones,Leanne, Flenady,Vicki, Cincotta,Robert, Crowther,Caroline A., Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth, Cochrane Database of Systematic Reviews, -, 2013

Douvas 1984

Douvas,S.G., Meeks,G.R., Graves,G., Intrapartum fetal heart rate monitoring as a predictor of fetal distress and immediate neonatal condition in low-birth weight (<1,800 grams) infants, American Journal of Obstetrics and Gynecology, 148, 300-302, 1984

Doyle 2014

Doyle,L.W., Anderson,P.J., Haslam,R., Lee,K.J., Crowther,C., Australasian Collaborative Trial of Magnesium Sulphate (ACTOMgSO, School-age outcomes of very preterm infants after antenatal treatment with magnesium sulfate vs placebo, JAMA, 312, 1105-1113, 2014

Eleje 2014

Eleje,G.U., Ezugwu,E.C., Ogunyemi,D., Eleje,L.I., Ikechebelu,J.I., Igwegbe,A.O., Okonkwo,J.E., Ikpeze,O.C., Udigwe,G.O., Onah,H.E., Nwosu,B.O., Ezeama,C.O., Ezenkwele,E.P., Accuracy and cost-analysis of placental alpha-microglobulin-1 test in the diagnosis of premature rupture of fetal membranes in resource-limited community settings, Journal of Obstetrics and Gynaecology Research, 41, 29-38, 2014

Elimian 2014

Elimian,A., Goodman,J., Escobedo,M., Nightingale,L., Knudtson,E., Williams,M., Immediate compared with delayed cord clamping in the preterm neonate: a randomized controlled trial, Obstetrics and Gynecology, 124, 1075-1079, 2014

Eroglu 2007

Eroglu,D., Yanik,F., Oktem,M., Zeyneloglu,H.B., Kuscu,E., Prediction of preterm delivery among women with threatened preterm labor, Gynecologic and Obstetric Investigation, 64, 109-116, 2007

Farb 1983

Farb,H.F., Arnesen,M., Geistler,P., Knox,G.E., C-reactive protein with premature rupture of membranes and premature labor, Obstetrics and Gynecology, 62, 49-51, 1983

Fisk 1987

Fisk,N.M., Fysh,J., Child,A.G., Gatenby,P.A., Jeffery,H., Bradfield,A.H., Is C-reactive protein really useful in preterm premature rupture of the membranes?, British Journal of Obstetrics and Gynaecology, 94, 1159-1164, 1987

Fleming 2004

Fleming,A., Bonebrake,R., Istwan,N., Rhea,D., Coleman,S., Stanziano,G., Pregnancy and economic outcomes in patients treated for recurrent preterm labor, Journal of Perinatology, 24, 223-227, 2004

Garite & Freeman 1982

Garite,T.J., Freeman,R.K., Chorioamnionitis in the preterm gestation, Obstetrics and Gynecology, 59, 539-545, 1982

Gaucher & Payot 2011

Gaucher,N., Payot,A., From powerlessness to empowerment: Mothers expect more than information from the prenatal consultation for preterm labour, Paediatrics and Child Health, 16, 638-642, 2011

Giles 2000

Giles,W., Bisits,A., Knox,M., Madsen,G., Smith,R., The effect of fetal fibronectin testing on admissions to a tertiary maternal-fetal medicine unit and cost savings, American Journal of Obstetrics and Gynecology, 182, 439-442, 2000

Gomez 2005

Gomez,R., Romero,R., Medina,L., Nien,J.K., Chaiworapongsa,T., Carstens,M., Gonzalez,R., Espinoza,J., Iams,J.D., Edwin,S., Rojas,I., Cervicovaginal fibronectin improves the prediction

of preterm delivery based on sonographic cervical length in patients with preterm uterine contractions and intact membranes.[Erratum appears in Am J Obstet Gynecol. 2005 Jul;193(1):308-9], American Journal of Obstetrics and Gynecology, 192, 350-359, 2005

Gramellini 2007

Gramellini,D., Fieni,S., Kaihura,C., Modena,A.B., Cervical length as a predictor of preterm delivery: gestational age-related percentiles vs fixed cut-offs, Acta Bio-Medica de l Ateneo Parmense, 78, 220-224, 2007

Griffin 1997

Griffin,T., Kavanaugh,K., Soto,C.F., White,M., Parental evaluation of a tour of the neonatal intensive care unit during a high-risk pregnancy, JOGNN - Journal of Obstetric, Gynecologic, and Neonatal Nursing, 26, 59-65, 1997

Gupton & Heaman 1994

Gupton,A., Heaman,M., Learning needs of hospitalized women at risk for preterm birth, Applied Nursing Research, 7, 118-124, 1994

Haas 2012

Haas,D.M., Caldwell,D.M., Kirkpatrick,P., McIntosh,J.J., Welton,N.J., Tocolytic therapy for preterm delivery: systematic review and network meta-analysis, BMJ, 345, e6226-, 2012

Hawrylyshyn 1983

Hawrylyshyn,P., Bernstein,P., Milligan,J.E., Soldin,S., Pollard,A., Papsin,F.R., Premature rupture of membranes: the role of C-reactive protein in the prediction of chorioamnionitis, American Journal of Obstetrics and Gynecology, 147, 240-246, 1983

Hodnett 2010

Hodnett,Ellen D., Fredericks,Suzanne, Weston,Julie, Support during pregnancy for women at increased risk of low birthweight babies, Cochrane Database of Systematic Reviews, -, 2010

Holmes 2001

Holmes,P., Oppenheimer,L.W., Gravelle,A., Walker,M., Blayney,M., The effect of variable heart rate decelerations on intraventricular hemorrhage and other perinatal outcomes in preterm infants, Journal of Maternal-Fetal Medicine, 10, 264-268, 2001

Holst 2006

Holst,R.M., Jacobsson,B., Hagberg,H., Wennerholm,U.B., Cervical length in women in preterm labor with intact membranes: relationship to intra-amniotic inflammation/microbial invasion, cervical inflammation and preterm delivery, Ultrasound in Obstetrics and Gynecology, 28, 768-774, 2006

HoutzagerBA 2006

Houtzager BA,Hogendoorn SM,Papatsonis DN,Samsom JF,van Geijn HP,Bleker OP,van Wassenaer AG, Long-term follow up of children exposed in utero to nifedipine or ritodrine for the management of preterm labour, BJOG : an international journal of obstetrics and gynaecology, -, 2006

Iams 1995

Iams,J.D., Casal,D., McGregor,J.A., Goodwin,T.M., Kreaden,U.S., Lowensohn,R., Lockitch,G., Fetal fibronectin improves the accuracy of diagnosis of preterm labor, American Journal of Obstetrics and Gynecology, 173, 141-145, 1995

Aina-Mumuney 2007

Aina-Mumuney,A.J., Althaus,J.E., Henderson,J.L., Blakemore,M.C., Johnson,E.A., Graham,E.M., Intrapartum electronic fetal monitoring and the identification of systemic fetal inflammation, Journal of Reproductive Medicine, 52, 762-768, 2007

Ismail 1985

Ismail,M.A., Zinaman,M.J., Lowensohn,R.I., Moawad,A.H., The significance of C-reactive protein levels in women with premature rupture of membranes, American Journal of Obstetrics and Gynecology, 151, 541-544, 1985

Jain & Morris 1998

Jain,K., Morris,P.G., A clinical study to evaluate the usefulness of the MAST test in diagnosing pre-labour rupture of membranes, Journal of Obstetrics and Gynaecology, 18, 33-36, 1998

Jaju & Dhabadi 2011

Jaju,P.B., Dhabadi,V.B., Nifedipine versus ritodrine for suppression of preterm labor and analysis of side effects, Journal of Obstetrics and Gynaecology of India, 61, 534-537, 2011

Kariniemi 1984

Kariniemi,V., Jarvenpaa,A.L., Teramo,K., Fetal heart rate patterns and perinatal outcome of very-low-birthweight infants, British Journal of Obstetrics and Gynaecology, 91, 18-22, 1984

Kashanian 2014

Kashanian,M., Zamen,Z., Sheikhansari,N., Comparison between nitroglycerin dermal patch and nifedipine for treatment of preterm labor: a randomized clinical trial, Journal of Perinatology, 34, 683-687, 2014

Kenyon 2008

Kenyon,S., Pike,K., Jones,D.R., Brocklehurst,P., Marlow,N., Salt,A., Taylor,D.J., Childhood outcomes after prescription of antibiotics to pregnant women with preterm rupture of the membranes: 7-year follow-up of the ORACLE I trial, Lancet, 372, 1310-1318, 2008

Kenyon 2013

Kenyon,Sara, Boulvain,Michel, Neilson,James P., Antibiotics for preterm rupture of membranes, Cochrane Database of Systematic Reviews, -, 2013

Klauser 2012

Klauser,C.K., Briery,C.M., Keiser,S.D., Martin,R.W., Kosek,M.A., Morrison,J.C., Effect of antenatal tocolysis on neonatal outcomes, Journal of Maternal-Fetal and Neonatal Medicine, 25, 2778-2781, 2012

Klauser 2014

Klauser,C.K., Briery,C.M., Martin,R.W., Langston,L., Magann,E.F., Morrison,J.C., A comparison of three tocolytics for preterm labor: a randomized clinical trial, Journal of Maternal-Fetal and Neonatal Medicine, 27, 801-806, 2014

Kurki 1990

Kurki,T., Teramo,K., Ylikorkala,O., Paavonen,J., C-reactive protein in preterm premature rupture of the membranes.[Erratum appears in Arch Gynecol Obstet 1990;247(2):106], Archives of Gynecology and Obstetrics, 247, 31-37, 1990

Kwek 2004

Kwek,K., Khi,C., Ting,H.S., Yeo,G.S., Evaluation of a bedside test for phosphorylated insulin-like growth factor binding protein-1 in preterm labour, Annals of the Academy of Medicine, Singapore, 33, 780-783, 2004

Lam 2003

Lam,F., Istwan,N.B., Jacques,D., Coleman,S.K., Stanziano,G.J., Managing perinatal outcomes: the clinical benefit and cost-effectiveness of pharmacologic treatment of recurrent preterm labor, Managed Care, 12, 39-46, 2003

LaShay 2000

LaShay,N., Gilson,G., Joffe,G., Qualls,C., Curet,L., Will cervicovaginal interleukin-6 combined with fetal fibronectin testing improve the prediction of preterm delivery?, Journal of Maternal-Fetal Medicine, 9, 336-341, 2000

Lembet 2002

Lembet,A., Eroglu,D., Ergin,T., Kuscu,E., Zeyneloglu,H., Batioglu,S., Haberal,A., New rapid bed-side test to predict preterm delivery: phosphorylated insulin-like growth factor binding protein-1 in cervical secretions, Acta Obstetricia et Gynecologica Scandinavica, 81, 706-712, 2002

Lewis 1999

Lewis,D.F., Adair,C.D., Weeks,J.W., Barrilleaux,P.S., Edwards,M.S., Garite,T.J., A randomized clinical trial of daily nonstress testing versus biophysical profile in the management of preterm premature rupture of membranes, American Journal of Obstetrics and Gynecology, 181, 1495-1499, 1999

Lukes 1997

Lukes,A.S., Thorp,J.M.,Jr., Eucker,B., Pahel-Short,L., Predictors of positivity for fetal fibronectin in patients with symptoms of preterm labor, American Journal of Obstetrics and Gynecology, 176, 639-641, 1997

Luthy 1987

Luthy,D.A., Shy,K.K., van,Belle G., Larson,E.B., Hughes,J.P., Benedetti,T.J., Brown,Z.A., Effer,S., King,J.F., Stenchever,M.A., A randomized trial of electronic fetal monitoring in preterm labor, Obstetrics and Gynecology, 69, 687-695, 1987

Malak 1996

Malak,T.M., Sizmur,F., Bell,S.C., Taylor,D.J., Fetal fibronectin in cervicovaginal secretions as a predictor of preterm birth, British Journal of Obstetrics and Gynaecology, 103, 648-653, 1996

March 2011

March,M., De,VecianaM, Parson,A., The efficacy of umbilical cord milking on the reduction of red blood cell transfusion rates in infants born between 24 and 28 6/7 weeks gestation - A randomized controlled trial, American Journal of Obstetrics and Gynecology, 204, S204-, 2011

Marret 2007

Marret,S., Marpeau,L., Zupan-Simunek,V., Eurin,D., Leveque,C., Hellot,M.F., Benichou,J., PREMAG trial group., Magnesium sulphate given before very-preterm birth to protect infant

brain: the randomised controlled PREMAG trial*, BJOG: An International Journal of Obstetrics and Gynaecology, 114, 310-318, 2007

Marret 2008

Marret,S., Marpeau,L., Benichou,J., Benefit of magnesium sulfate given before very preterm birth to protect infant brain, Pediatrics, 121, 225-226, 2008

Martin 1974

Martin,Jr, Siassi,B., Hon,E.H., Fetal heart rate patterns and neonatal death in low birthweight infants, Obstetrics and Gynecology, 44, 503-510, 1974

Matsuda 2003

Matsuda,Y., Maeda,T., Kouno,S., The critical period of non-reassuring fetal heart rate patterns in preterm gestation, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 106, 36-39, 2003

McKenna 1999

McKenna,D.S., Chung,K., Iams,J.D., Effect of digital cervical examination on the expression of fetal fibronectin, Journal of Reproductive Medicine, 44, 796-800, 1999

Mercer 2012

Mercer,B., Antibiotics in the management of PROM and preterm labor, Obstetrics and Gynecology Clinics of North America, 39, 65-76, 2012

Mittendorf 2002

Mittendorf,R., Dambrosia,J., Pryde,P.G., Lee,K.S., Gianopoulos,J.G., Besinger,R.E., Tomich,P.G., Association between the use of antenatal magnesium sulfate in preterm labor and adverse health outcomes in infants, American Journal of Obstetrics and Gynecology, 186, 1111-1118, 2002

Mugford 1991

Mugford,M., Piercy,J., Chalmers,I., Cost implications of different approaches to the prevention of respiratory distress syndrome, Archives of Disease in Childhood, 66, 757-764, 1991

Nankali 2014

Nankali,A., Jamshidi,P.K., Rezaei,M., The Effects of Glyceryl Trinitrate Patch on the Treatment of Preterm Labor: A Single-blind Randomized Clinical Trial, Journal of Reproduction and Infertility, 15, 71-77, 2014

Nikbakht 2014

Nikbakht,R., Taheri,Moghadam M., Ghane'ee,H., Nifedipine compared to magnesium sulfate for treating preterm labor: A randomized clinical trial, Iranian Journal of Reproductive Medicine, 12, 145-150, 2014

Nisenblat 2006

Nisenblat,V., Alon,E., Barak,S., Gonen,R., Bader,D., Ohel,G., Fetal heart rate patterns and neurodevelopmental outcome in very low birth weight infants, Acta Obstetricia et Gynecologica Scandinavica, 85, 792-796, 2006

Olatunbosun 1995

Olatunbosun,O.A., al-Nuaim,L., Turnell,R.W., Emergency cerclage compared with bed rest for advanced cervical dilatation in pregnancy, International Surgery, 80, 170-174, 1995

Palacio 2007

Palacio,M., Sanin-Blair,J., Sanchez,M., Crispi,F., Gomez,O., Carreras,E., Coll,O., Cararach,V., Gratacos,E., The use of a variable cut-off value of cervical length in women admitted for preterm labor before and after 32 weeks, Ultrasound in Obstetrics and Gynecology, 29, 421-426, 2007

Perrone 2012

Perrone,G., Anceschi,M.M., Capri,O., Galoppi,P., Pizzulo,S., Buccheri,M., Pascone,R., Nofroni,I., Brunelli,R., Maternal C-reactive protein at hospital admission is a simple predictor of funisitis in preterm premature rupture of membranes, Gynecologic and Obstetric Investigation, 74, 95-99, 2012

Pizzi 2014

Pizzi,L.T., Seligman,N.S., Baxter,J.K., Jutkowitz,E., Berghella,V., Cost and cost effectiveness of vaginal progesterone gel in reducing preterm birth: an economic analysis of the PREGNANT trial, Pharmacoeconomics, 32, 467-478, 2014

POPPY Steering Group 2009

Family-centred care in neonatal units. A summary of research results and recommendations from the POPPY Parents of premature babies) project. London: NCT; 2009

Porto 2011

Porto,A.M., Coutinho,I.C., Correia,J.B., Amorim,M.M., Effectiveness of antenatal corticosteroids in reducing respiratory disorders in late preterm infants: randomised clinical trial, BMJ (Clinical research ed.), Vol.342, pp.d1696, 2011., -

Rabe 2012

Rabe,H., az-Rossello,J.L., Duley,L., Dowswell,T., Effect of timing of umbilical cord clamping and other strategies to influence placental transfusion at preterm birth on maternal and infant outcomes. [Update of Cochrane Database Syst Rev. 2004;(4):CD003248; PMID: 15495045], Cochrane Database of Systematic Reviews, 8, CD003248-, 2012

Ranjit 2015

Ranjit,T., Nesargi,S., Rao,P.N., Sahoo,J.P., Ashok,C., Chandrakala,B.S., Bhat,S., Effect of early versus delayed cord clamping on hematological status of preterm infants at 6 wk of age, Indian Journal of Pediatrics, 82, 29-34, 2015

RCOG 2011

Cervical Cerclage, RCOG Green-top Guideline No. 60

RCOG 2015

RCOG, Clamping of the Umbilical Cord and Placental Transfusion (Scientific Impact Paper No. 14), 2015

Rayburn 1987

Rayburn,W.F., Johnson,M.Z., Hoffman,K.L., Donn,S.M., Nelson,R.M.,Jr., Intrapartum fetal heart rate patterns and neonatal intraventricular hemorrhage, American Journal of Perinatology, 4, 98-101, 1987

Roberts & Dalziel 2013

Roberts,Devender, Dalziel,Stuart R., Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth, Cochrane Database of Systematic Reviews, -, 2013

Romem & Artal 1984

Romem,Y., Artal,R., C-reactive protein as a predictor for chorioamnionitis in cases of premature rupture of the membranes, American Journal of Obstetrics and Gynecology, 150, 546-550, 1984

Romero 2012

Romero,R., Nicolaides,K., Conde-Agudelo,A., Tabor,A., O'Brien,J.M., Cetingoz,E., Da,Fonseca E., Creasy,G.W., Klein,K., Rode,L., Soma-Pillay,P., Fusey,S., Cam,C., Alfirevic,Z., Hassan,S.S., Vaginal progesterone in women with an asymptomatic sonographic short cervix in the midtrimester decreases preterm delivery and neonatal morbidity: a systematic review and metaanalysis of individual patient data, American Journal of Obstetrics and Gynecology, 206, 124-19, 2012

Rouse 2008

Rouse,D.J., Hirtz,D.G., Thom,E., Varner,M.W., Spong,C.Y., Mercer,B.M., Iams,J.D., Wapner,R.J., Sorokin,Y., Alexander,J.M., Harper,M., Thorp,J.M.,Jr., Ramin,S.M., Malone,F.D., Carpenter,M., Miodovnik,M., Moawad,A., O'Sullivan,M.J., Peaceman,A.M., Hankins,G.D., Langer,O., Caritis,S.N., Roberts,J.M., Eunice Kennedy Shriver NICHD Maternal-Fetal Medicine Units Network., A randomized, controlled trial of magnesium sulfate for the prevention of cerebral palsy, New England Journal of Medicine, 359, 895-905, 2008

Sakai 2003

Sakai,M., Sasaki,Y., Yamagishi,N., Tanebe,K., Yoneda,S., Saito,S., The preterm labor index and fetal fibronectin for prediction of preterm delivery with intact membranes, Obstetrics and Gynecology, 101, 123-128, 2003

Salim 2012

Salim,R., Garmi,G., Nachum,Z., Zafran,N., Baram,S., Shalev,E., Nifedipine compared with atosiban for treating preterm labor: a randomized controlled trial, Obstetrics and Gynecology, 120, 1323-1331, 2012

Sawyer 2013

Sawyer,A., Rabe,H., Abbott,J., Gyte,G., Duley,L., Ayers,S., Parents' experiences and satisfaction with care during the birth of their very preterm baby: A qualitative study, BJOG: An International Journal of Obstetrics and Gynaecology, 120, 637-643, 2013

Schmitz 2006

Schmitz,T., Maillard,F., Bessard-Bacquaert,S., Kayem,G., Fulla,Y., Cabrol,D., Goffinet,F., Selective use of fetal fibronectin detection after cervical length measurement to predict spontaneous preterm delivery in women with preterm labor, American Journal of Obstetrics and Gynecology, 194, 138-143, 2006

Schmitz 2008

Schmitz,T., Kayem,G., Maillard,F., Lebret,M.T., Cabrol,D., Goffinet,F., Selective use of sonographic cervical length measurement for predicting imminent preterm delivery in women with preterm labor and intact membranes, Ultrasound in Obstetrics and Gynecology, 31, 421-426, 2008

Schreyer 1989

Schreyer,P., Caspi,E., Bar,NatanN, Tal,E., Weinraub,Z., The predictive value of fetal breathing movement and Bishop score in the diagnosis of 'true' preterm labor, American Journal of Obstetrics and Gynecology, 161, 886-889, 1989

Senden & Owen 1996

Senden,I.P., Owen,P., Comparison of cervical assessment, fetal fibronectin and fetal breathing in the diagnosis of preterm labour, Clinical and Experimental Obstetrics and Gynecology, 23, 5-9, 1996

Shy 1988

Shy,K.K., Olshan,A.F., Hickok,D.E., Luthy,D.A., Electronic fetal monitoring during premature labor and the occurrence of perinatal mortality in very low birthweight infants, Birth, 15, 14-18, 1988

Siassakos 2009

Siassakos,D., O'Brien,K., Draycott,T., Healthcare evaluation of the use of atosiban and fibronectin for the management of pre-term labour, Journal of Obstetrics and Gynaecology, 29, 507-511, 2009

Skoll 2006

Skoll,A., St,Louis P., Amiri,N., Delisle,M.F., Lalji,S., The evaluation of the fetal fibronectin test for prediction of preterm delivery in symptomatic patients, Journal of Obstetrics and Gynaecology Canada: JOGC, 28, 206-213, 2006

Smith 2012

Smith,E.J., Muller,C.L., Sartorius,J.A., White,D.R., Maslow,A.S., C-reactive protein as a predictor of chorioamnionitis, Journal of the American Osteopathic Association, 112, 660-664, 2012

Sotiriadis 2010

Sotiriadis,A., Kavvadias,A., Papatheodorou,S., Paraskevaidis,E., Makrydimas,G., The value of serial cervical length measurements for the prediction of threatened preterm labour, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 148, 17-20, 2010

Stupin 2008

Stupin,J.H., David,M., Siedentopf,J.P., Dudenhausen,J.W., Emergency cerclage versus bed rest for amniotic sac prolapse before 27 gestational weeks. A retrospective, comparative study of 161 women, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 139, 32-37, 2008

Swamy 2005

Swamy,G.K., Simhan,H.N., Gammill,H.S., Heine,R.P., Clinical utility of fetal fibronectin for predicting preterm birth, Journal of Reproductive Medicine, 50, 851-856, 2005

Tagore & Kwek 2010

Tagore,S., Kwek,K., Comparative analysis of insulin-like growth factor binding protein-1 (IGFBP-1), placental alpha-microglobulin-1 (PAMG-1) and nitrazine test to diagnose premature rupture of membranes in pregnancy, Journal of Perinatal Medicine, 38, 609-612, 2010

Tanir 2008

Tanir,H.M., Sener,T., Yildiz,Z., Cervicovaginal fetal fibronectin (FFN) for prediction of preterm delivery in symptomatic cases: a prospective study, Clinical and Experimental Obstetrics and Gynecology, 35, 61-64, 2008

Tanir 2009

Tanir,H.M., Sener,T., Yildiz,Z., Cervical phosphorylated insulin-like growth factor binding protein-1 for the prediction of preterm delivery in symptomatic cases with intact membranes, Journal of Obstetrics and Gynaecology Research, 35, 66-72, 2009

Tekesin 2005

Tekesin,I., Marek,S., Hellmeyer,L., Reitz,D., Schmidt,S., Assessment of rapid fetal fibronectin in predicting preterm delivery, Obstetrics and Gynecology, 105, 280-284, 2005

Ting 2007

Ting,H.S., Chin,P.S., Yeo,G.S., Kwek,K., Comparison of bedside test kits for prediction of preterm delivery: phosphorylated insulin-like growth factor binding protein-1 (pIGFBP-1) test and fetal fibronectin test, Annals of the Academy of Medicine, Singapore, 36, 399-402, 2007

Tsoi 2005

Tsoi,E., Fuchs,I.B., Rane,S., Geerts,L., Nicolaides,K.H., Sonographic measurement of cervical length in threatened preterm labor in singleton pregnancies with intact membranes, Ultrasound in Obstetrics and Gynecology, 25, 353-356, 2005

Tsoi 2006

Tsoi,E., Akmal,S., Geerts,L., Jeffery,B., Nicolaides,K.H., Sonographic measurement of cervical length and fetal fibronectin testing in threatened preterm labor, Ultrasound in Obstetrics and Gynecology, 27, 368-372, 2006

Valdes 2012

Valdes,E., Salinas,H., Toledo,V., Lattes,K., Cuellar,E., Perucca,E., Diaz,R., Montecinos,F., Reyes,A., Nifedipine versus fenoterol in the management of preterm labor: a randomized, multicenter clinical study, Gynecologic and Obstetric Investigation, 74, 109-115, 2012

van Baaren 2014

van Baaren,G.J., Vis,J.Y., Wilms,F.F., Oudijk,M.A., Kwee,A., Porath,M.M., Oei,G., Scheepers,H.C., Spaanderman,M.E., Bloemenkamp,K.W., Haak,M.C., Bolte,A.C., Bax,C.J., Cornette,J.M., Duvekot,J.J., Nij Bijvanck,B.W., van,Eyck J., Franssen,M.T., Sollie,K.M., Vandenbussche,F.P., Woiski,M., Grobman,W.A., van der Post,J.A., Bossuyt,P.M., Opmeer,B.C., Mol,B.W., Predictive value of cervical length measurement and fibronectin testing in threatened preterm labor, Obstetrics and Gynecology, 123, 1185-1192, 2014

Wex 2009

Wex,J., Connolly,M., Rath,W., Atosiban versus betamimetics in the treatment of preterm labour in Germany: an economic evaluation, BMC pregnancy and childbirth, 9, 23-, 2009

Wex 2011

Wex,J., bou-Setta,A.M., Clerici,G., Di Renzo,G.C., Atosiban versus betamimetics in the treatment of preterm labour in Italy: clinical and economic importance of side-effects, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 157, 128-135, 2011

Yoon 1996

Yoon,B.H., Jun,J.K., Park,K.H., Syn,H.C., Gomez,R., Romero,R., Serum C-reactive protein, white blood cell count, and amniotic fluid white blood cell count in women with preterm premature rupture of membranes, Obstetrics and Gynecology, 88, 1034-1040, 1996

Young 2012

Young,E., Tsai,E., O'Riordan,A., A qualitative study of predelivery counselling for extreme prematurity, Paediatrics and Child Health, 17, 432-436, 2012

Appendices

All appendices are contained in a separate file.

Appendix A: Scope

Appendix B: Stakeholders

Appendix C: Declarations of interest

Appendix D: Review protocols

Appendix E: Search stratagies

Appendix F: PRIMA flow diagrams

Appendix G: Excluded studies

Appendix H: Evidence tables

Appendix I: Forest plots

Appendix J: Network meta-analysis of tocolytics