NCC-WCH

Version 2.0

Preterm labour and birth

Appendices I & J

NICE Guideline 25

Methods, evidence and recommendations

November 2015, updated June 2022

Final

Commissioned by the National Institute for Health and care Excellence

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

Update information

In June 2022 this document was updated to redact some content that was now out of date as a result of the 2022 evidence review on the use of repeat courses of maternal corticosteroids. See the NICE website for the current recommendations at https://www.nice.org.uk/guidance/ng25.

Funding

Registered charity no. 213280

Contents

Appendix A: Scope

The scope is presented in a separate document

Appendix B: Stakeholders

The scope is presented in a separate document

Appendix C: Declarations of interest

The scope is presented in a separate document

Appendix D: Review protocols

The scope is presented in a separate document

Appendix E: Search strategies

The scope is presented in a separate document

Appendix F: PRISMA flow diagrams

The scope is presented in a separate document

Appendix G: Excluded studies

The scope is presented in a separate document

Appendix H: Evidence tables

The scope is presented in a separate document

Appendix I: Forest plots

I.1 Forest plots for review question: Information and support

No forest plots were generated for this review question

I.2 Prophylactic vaginal progesterone and prophylactic cervical cerclage

I.2.1 Prophylactic progesterone

This section was updated and replaced in 2019. Please see the nice website for the updated guideline.

I.2.2 Prophylactic cervical cerclage

Figure 1: Prophylactic cervical cerclage versus no cerclage - perinatal death

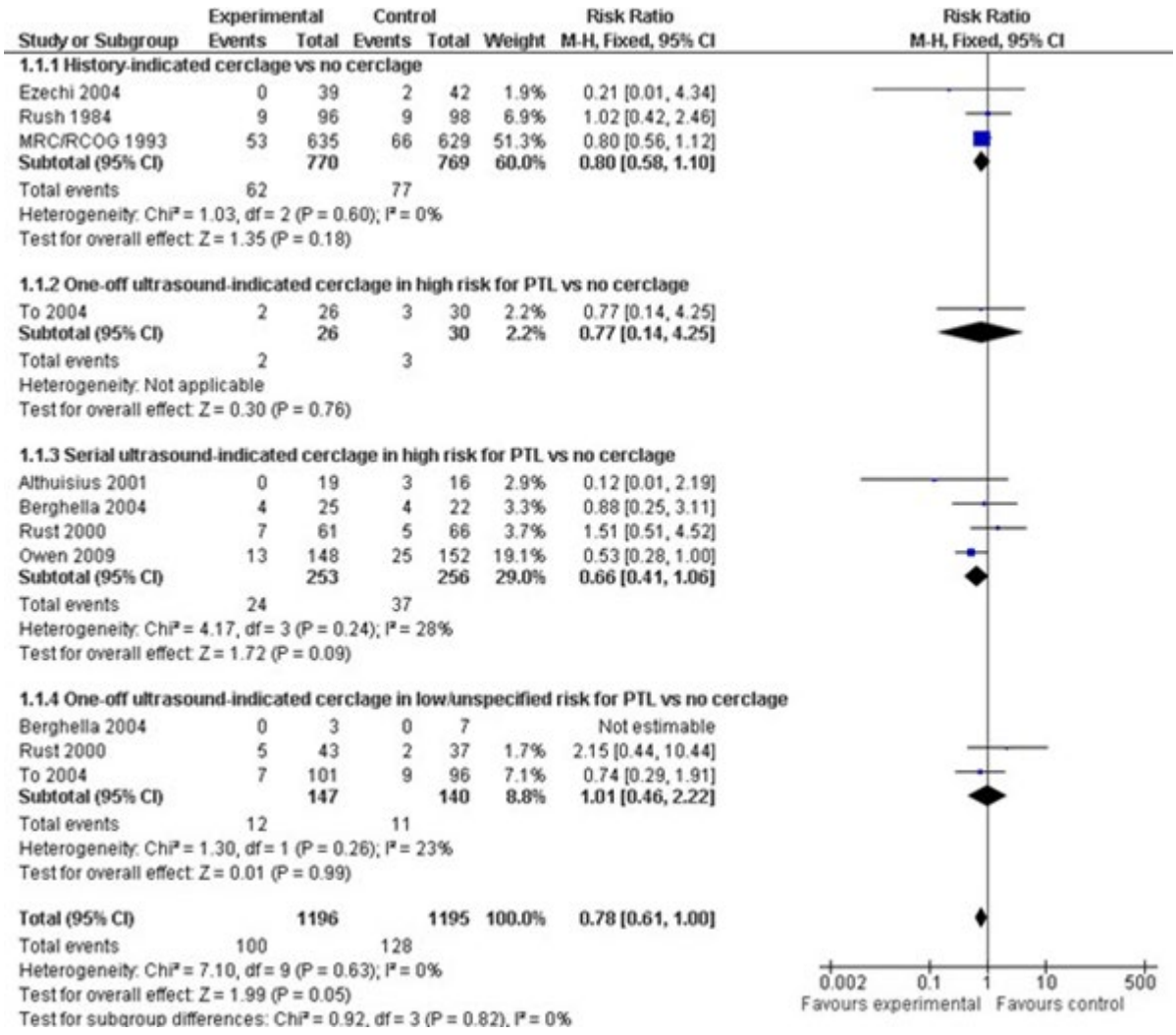


Figure 2: Prophylactic cervical cerclage versus no cerclage - Serious neonatal morbidity

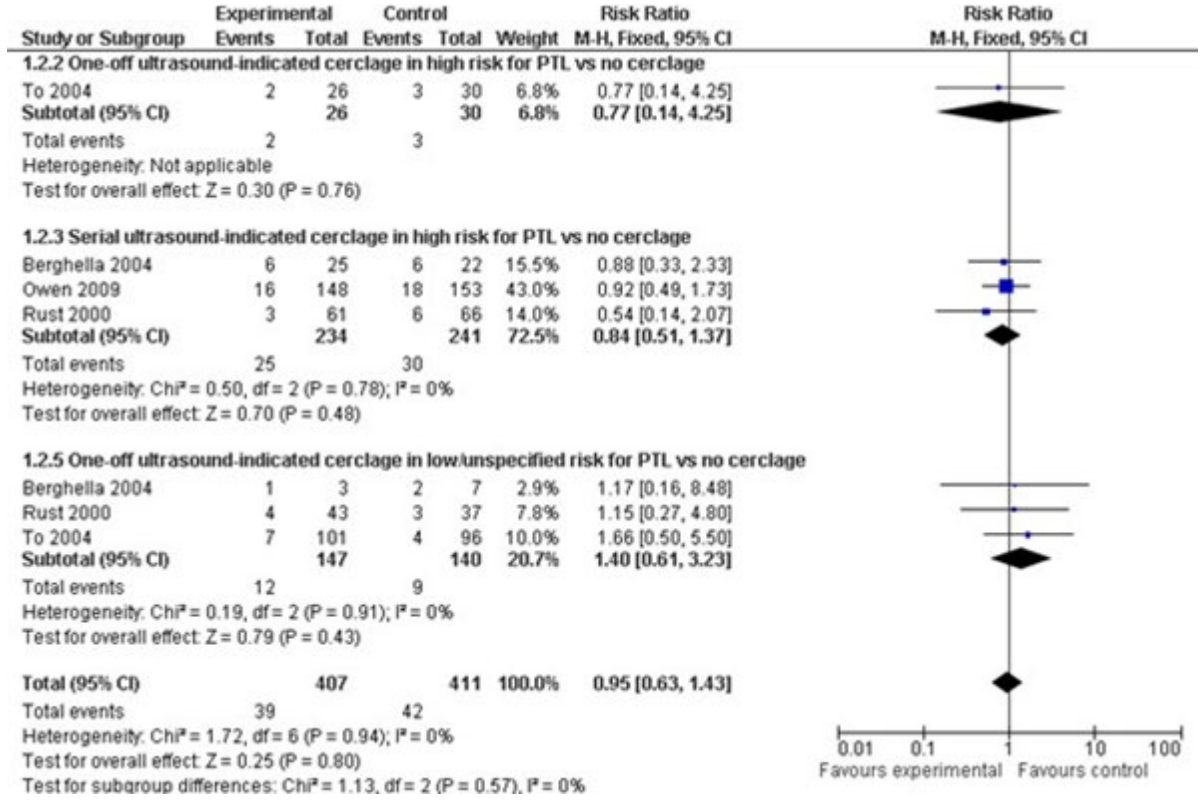


Figure 3: Prophylactic cervical cerclage versus no cerclage- Preterm birth before 37+0 weeks

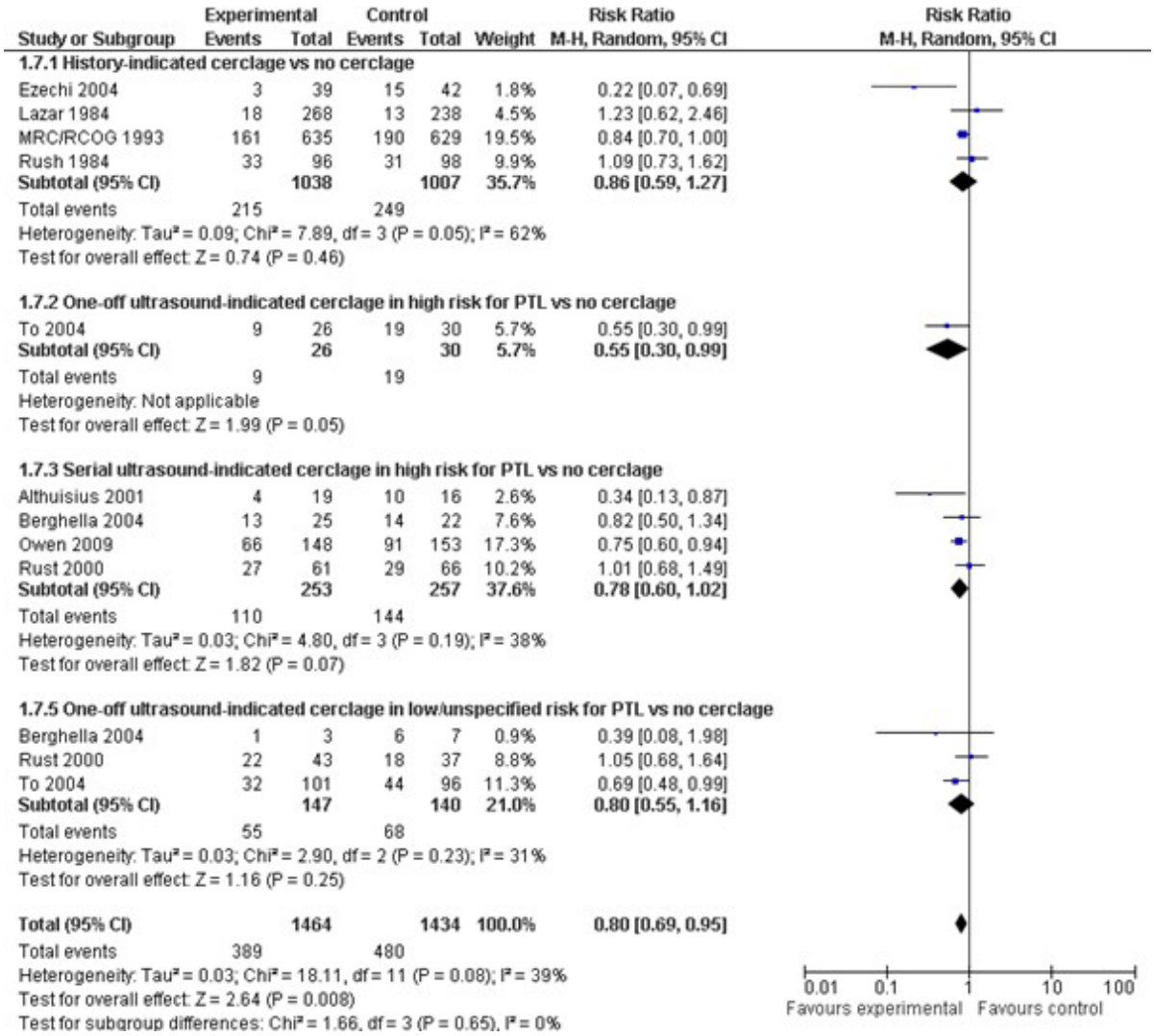


Figure 4: Prophylactic cervical cerclage versus no cerclage- Preterm birth before 34+0 weeks

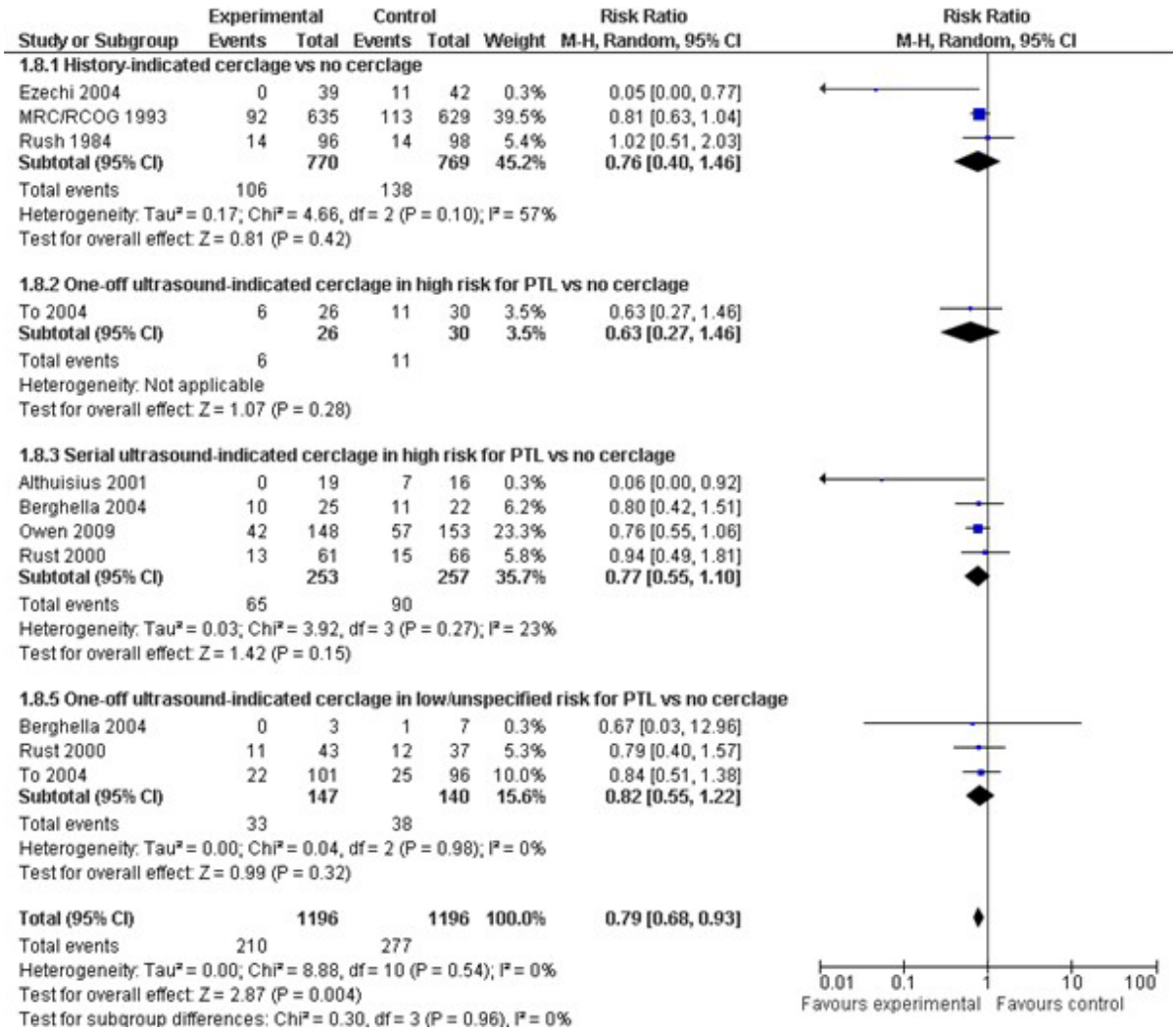


Figure 5: Prophylactic cervical cerclage versus no cerclage- Preterm birth before 38+0 weeks

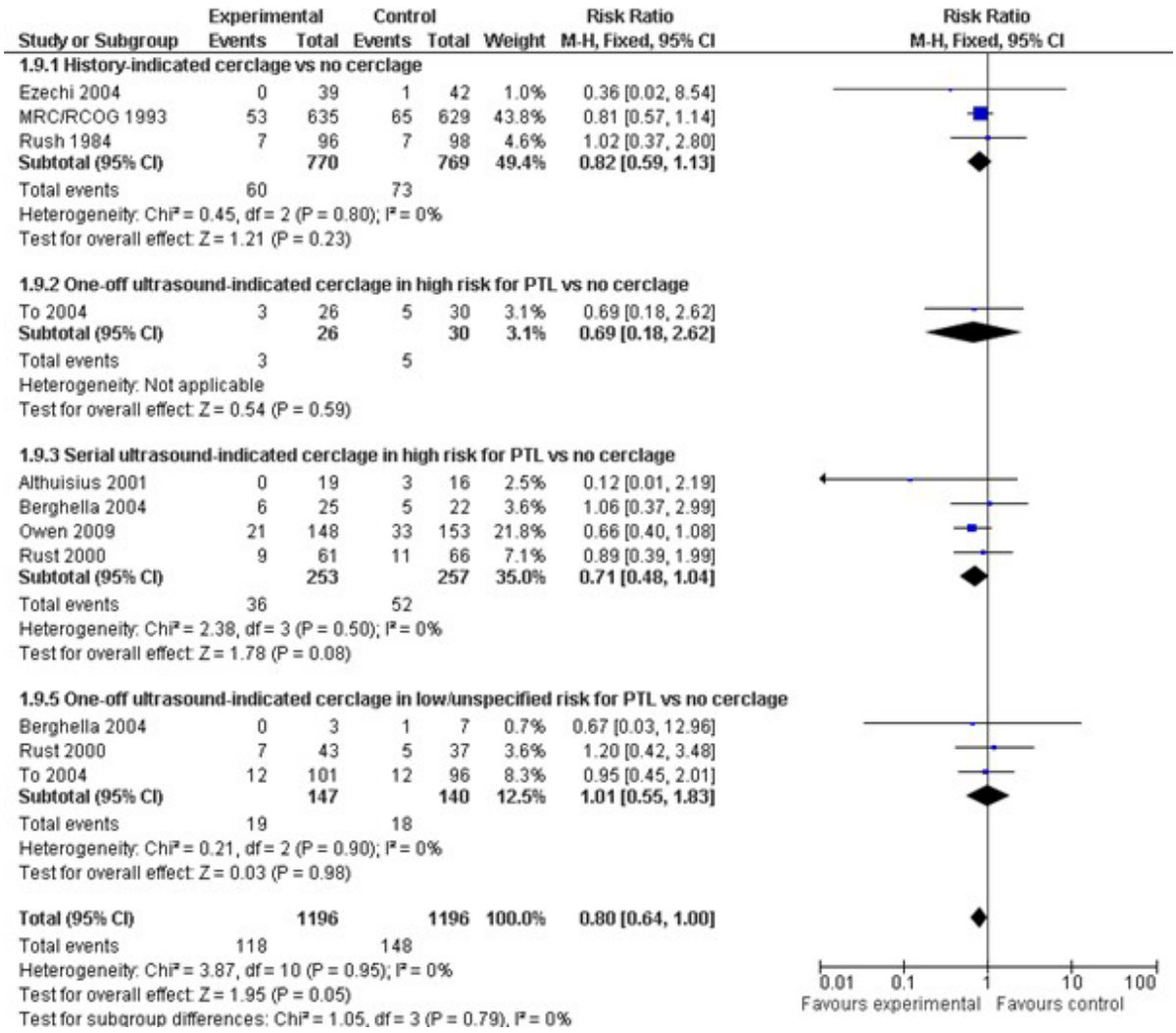


Figure 6: Prophylactic cervical cerclage versus no cerclage- maternal side effects

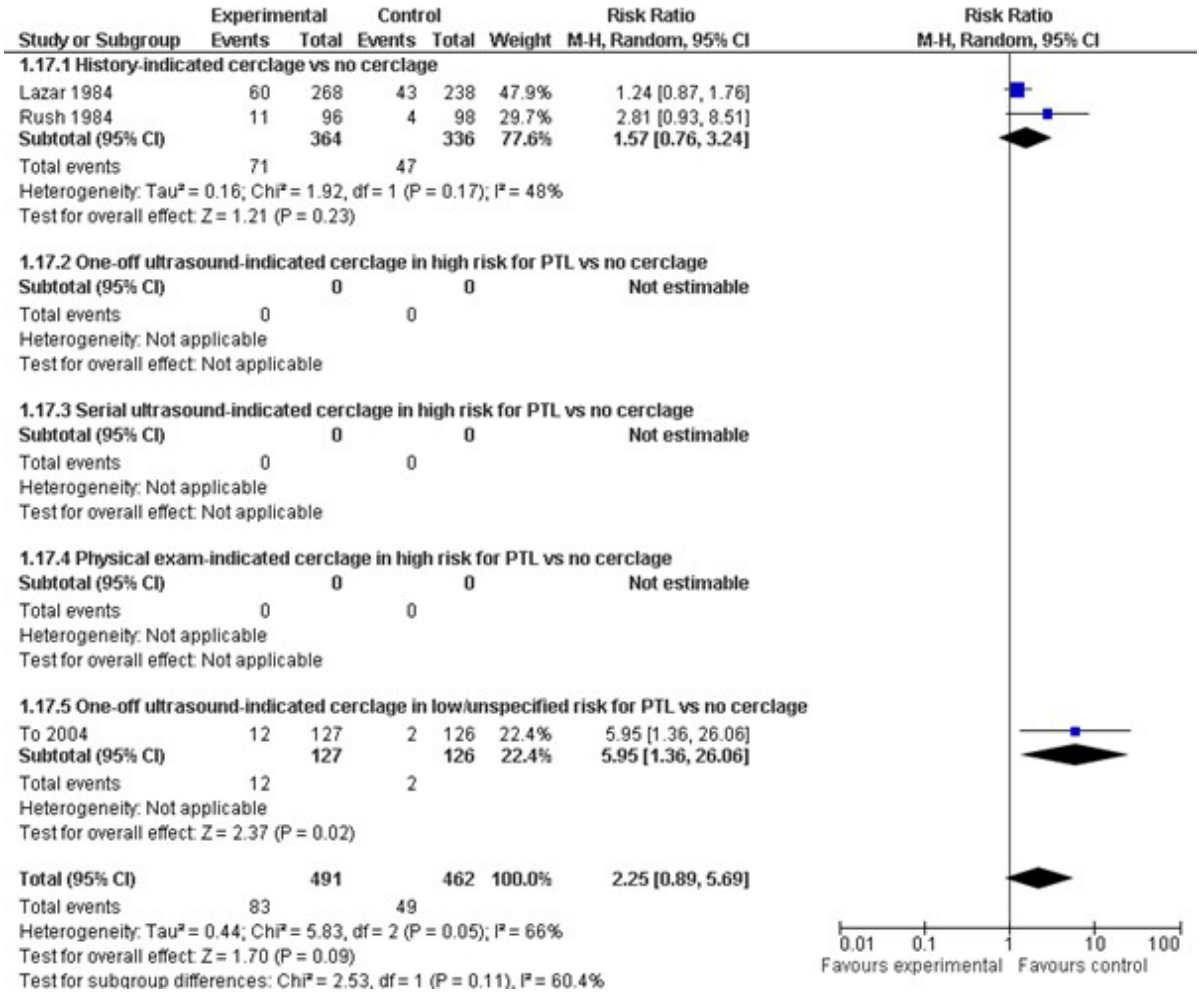
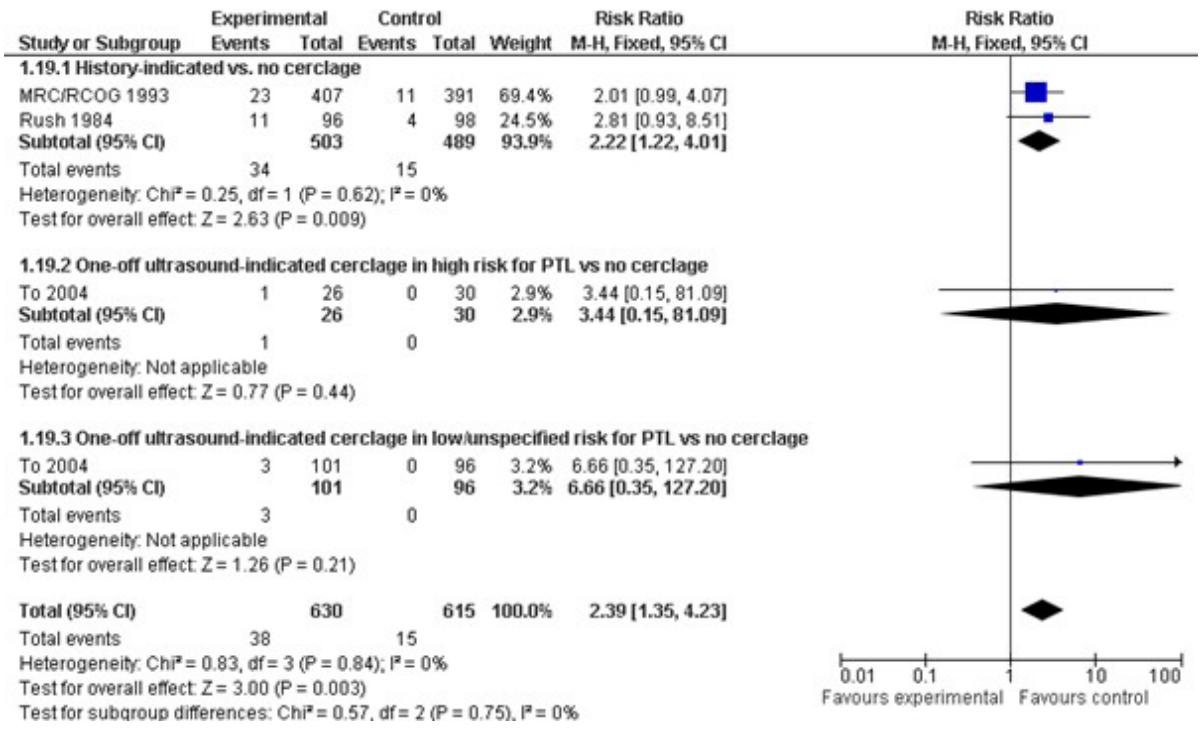
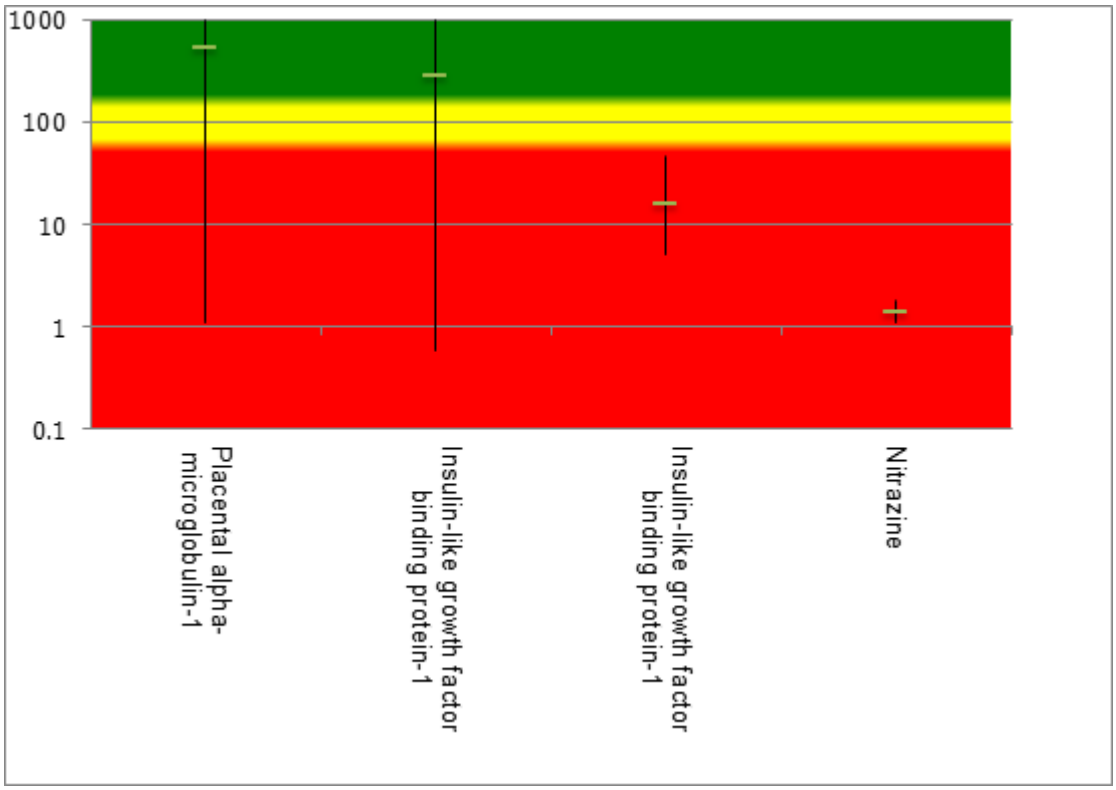


Figure 7: Prophylactic cervical cerclage versus no cerclage- pyrexia



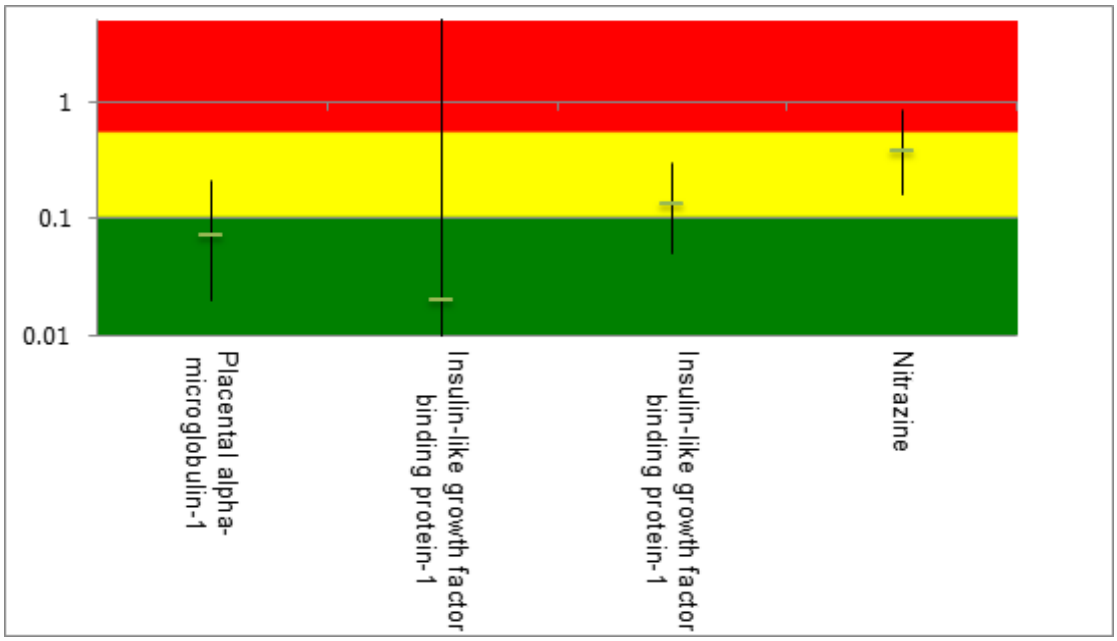
I.3 Diagnosing preterm prelabour rupture of membranes (P-PROM

Figure 8: Positive likelihood ratio for diagnosing preterm pre-labour rupture of membranes



Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 9: Negative likelihood ratio for diagnosing preterm pre-labour rupture of membranes



Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

I.4 Antenatal prophylactic antibiotics for women with P-PROM

I.4.1 Any antibiotic versus placebo

I.4.1.1 Neonatal outcomes

Figure 10: Perinatal death

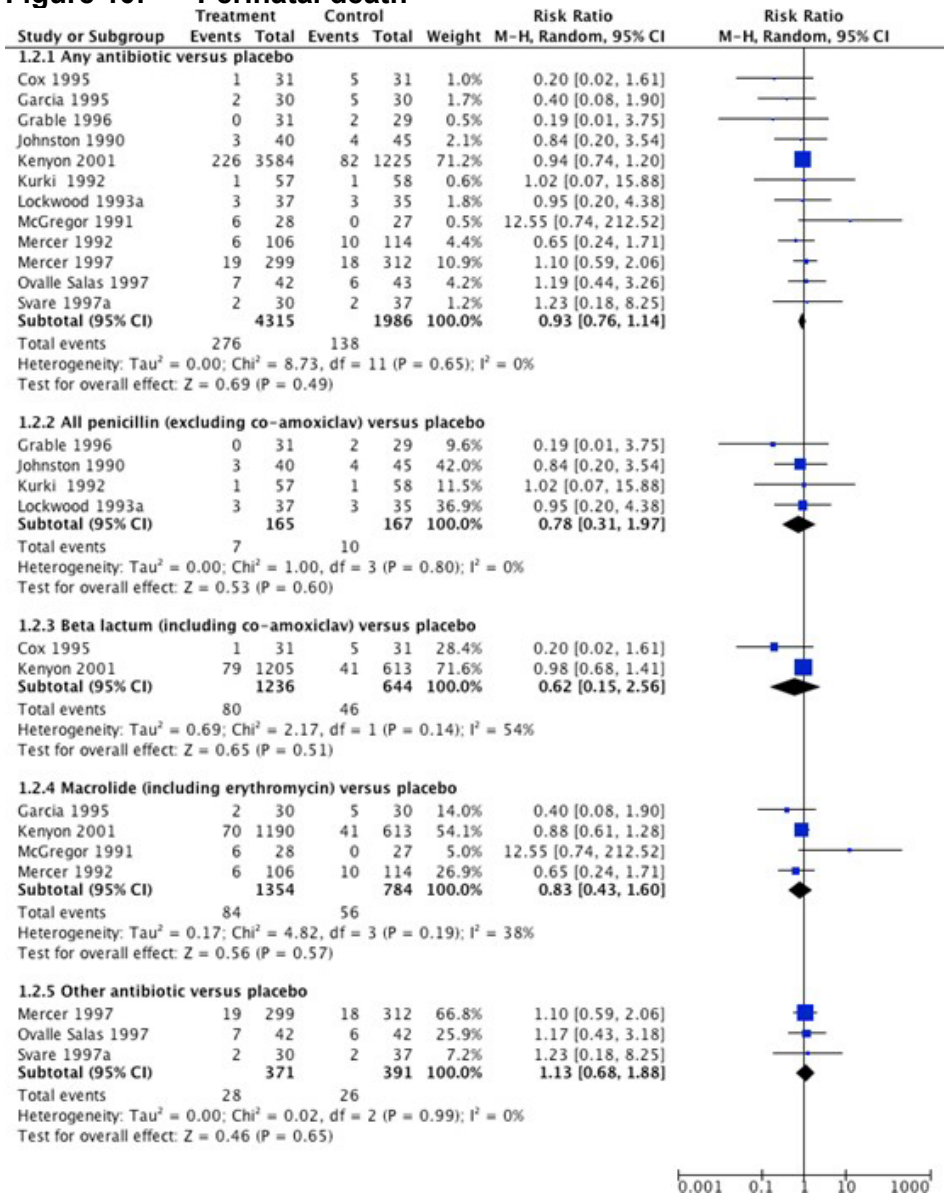


Figure 11: Neonatal necrotising enterocolitis

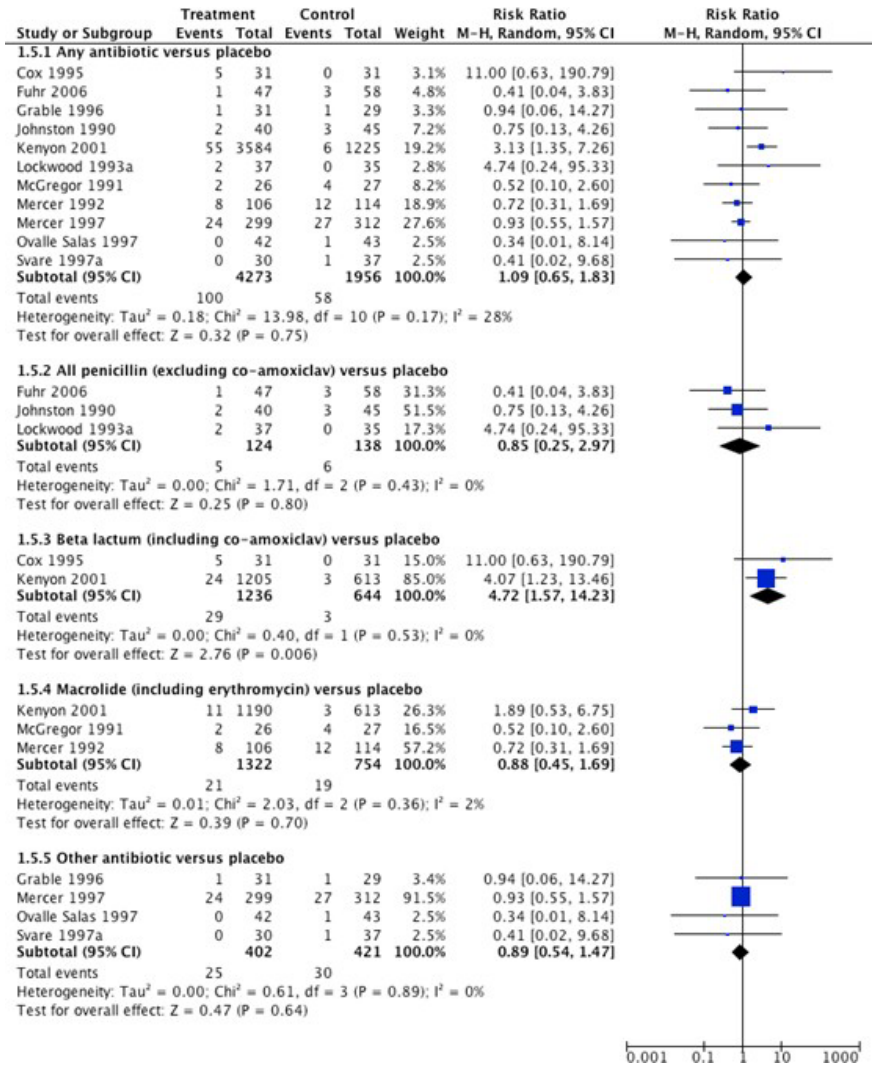


Figure 12: Neonatal necrotising enterocolitis

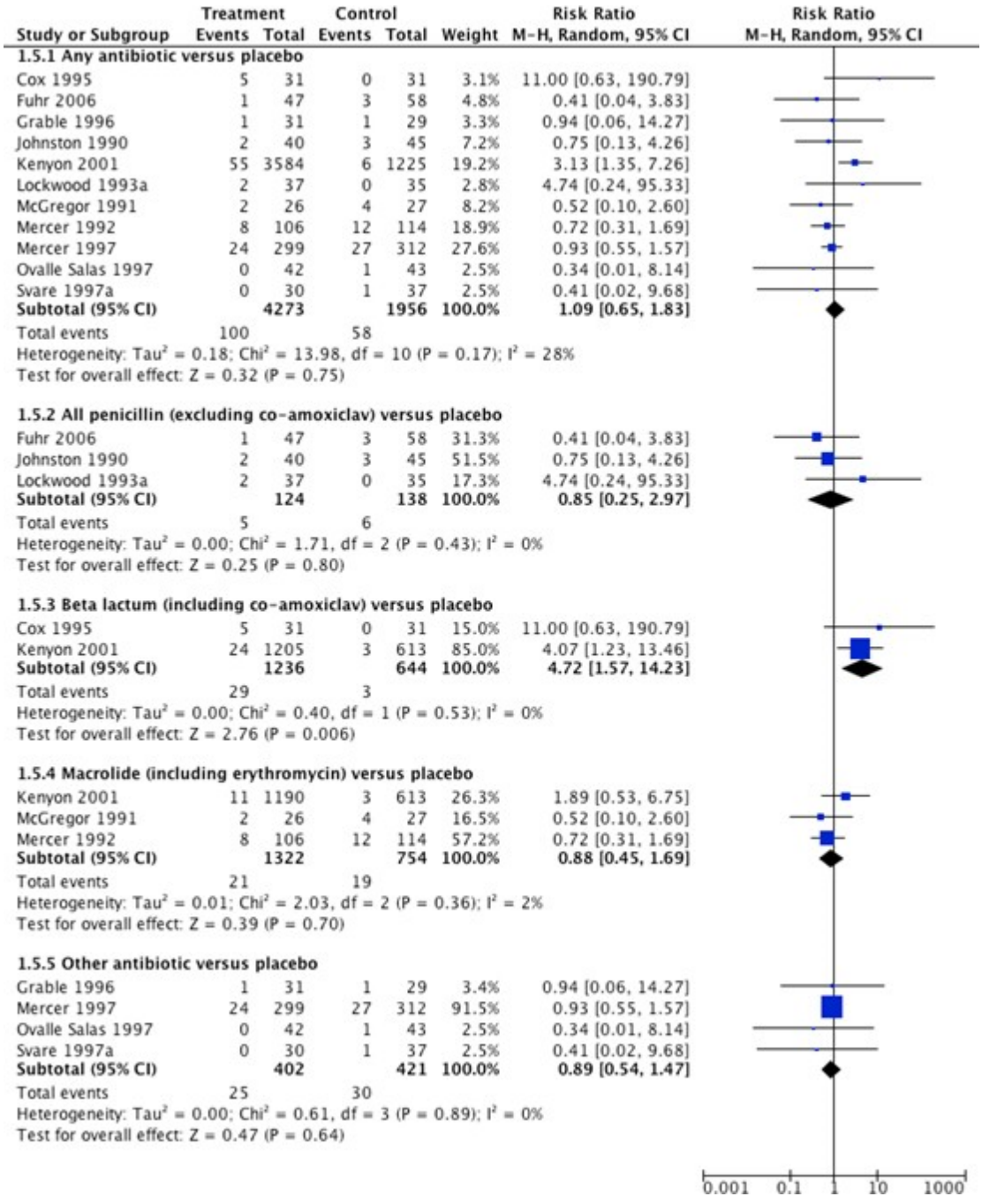


Figure 13: Birth before 37 weeks' gestation

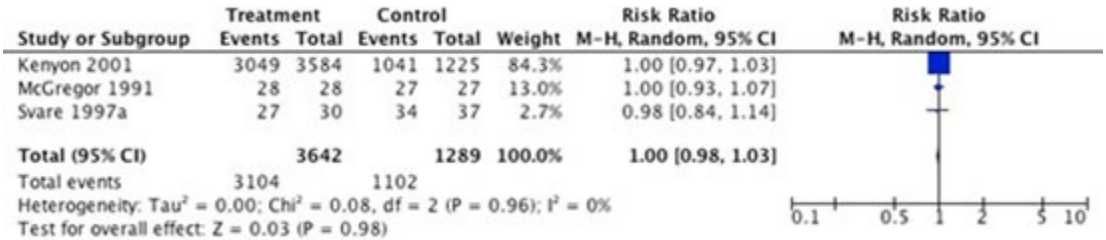
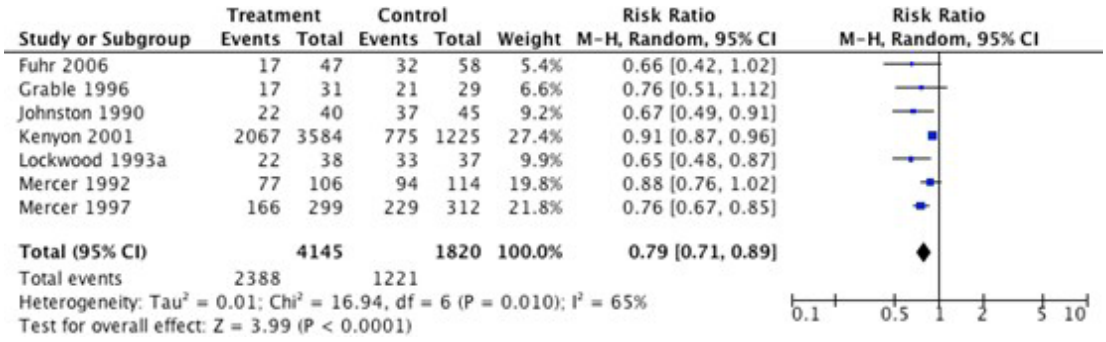


Figure 14: Birth within 7 days of randomisation



I.4.2 Maternal outcomes

Figure 15: Maternal death

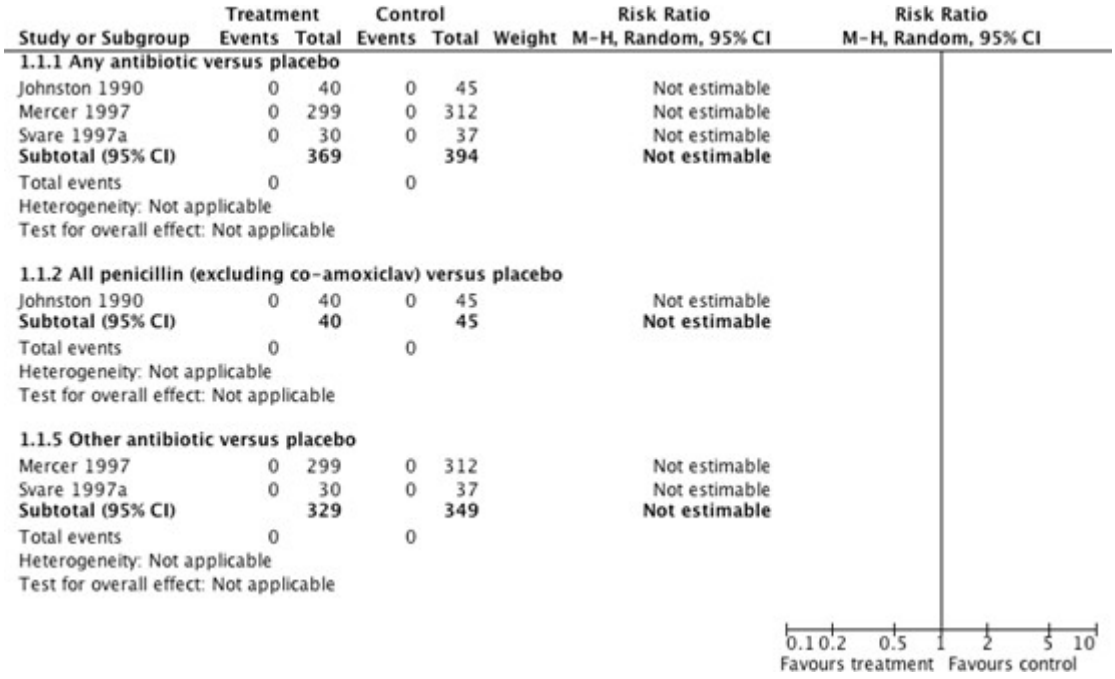


Figure 16: Maternal infection after delivery prior to discharge

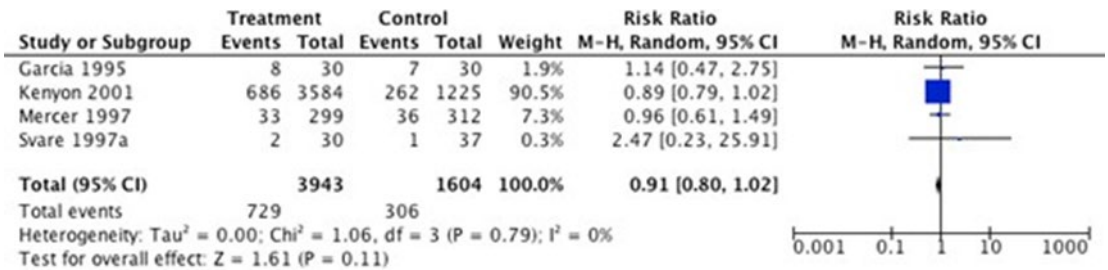


Figure 17: Chorioamnionitis

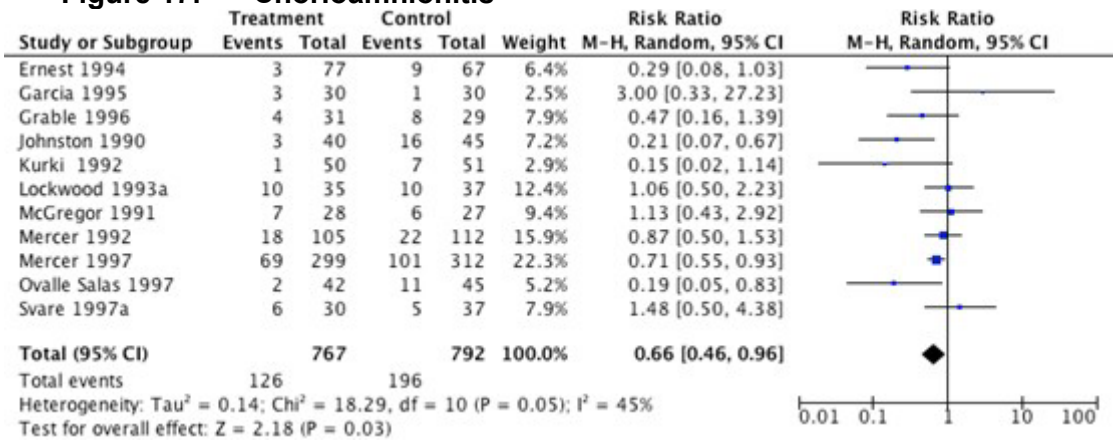


Figure 18: Major adverse drug reaction

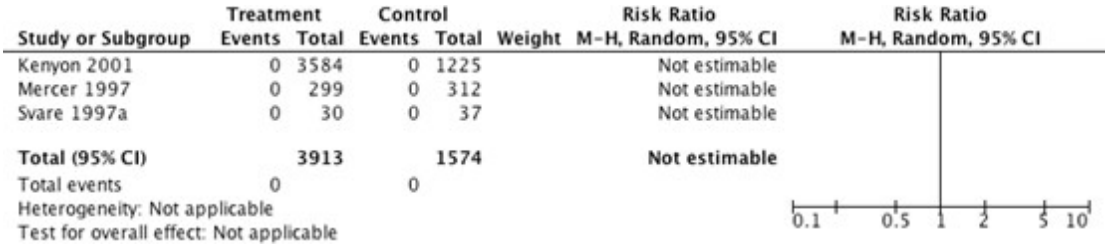


Figure 19: Antibiotics therapy versus either placebo or no antibiotics therapy

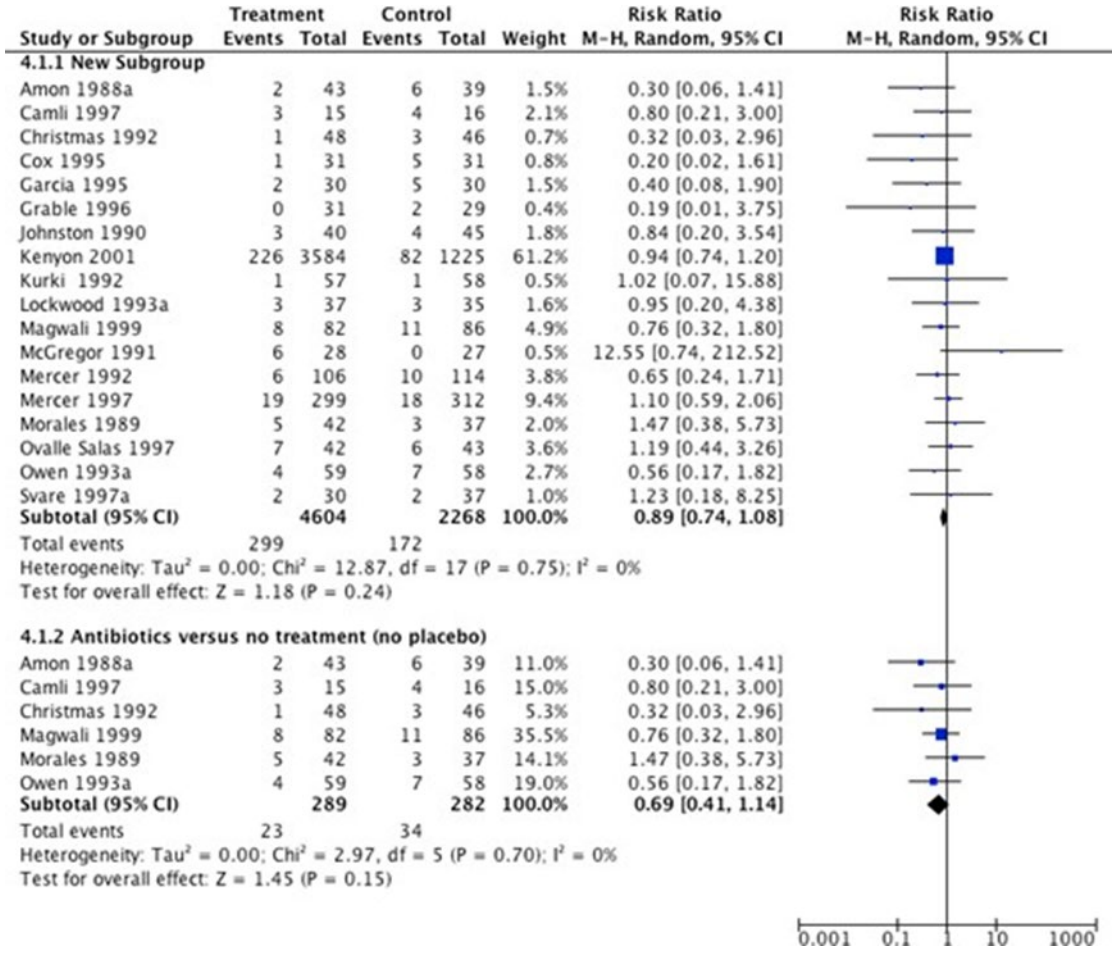


Figure 20: Intraventricular haemorrhage

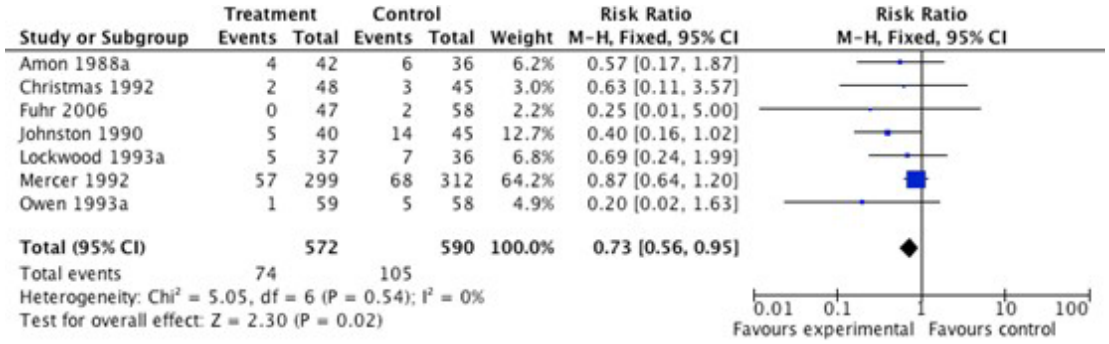


Figure 21: Sepsis

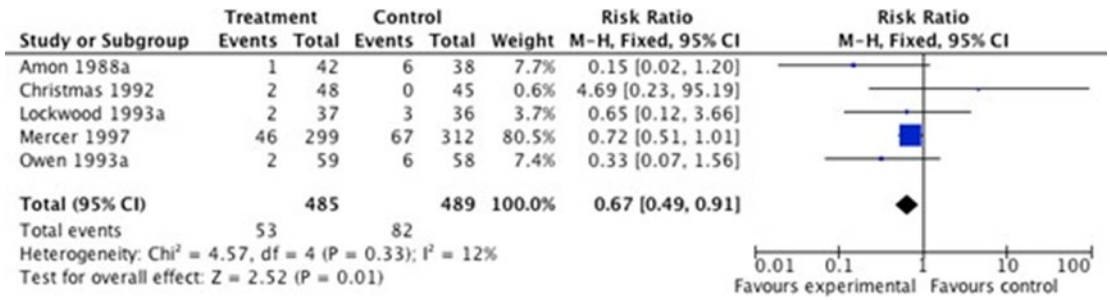
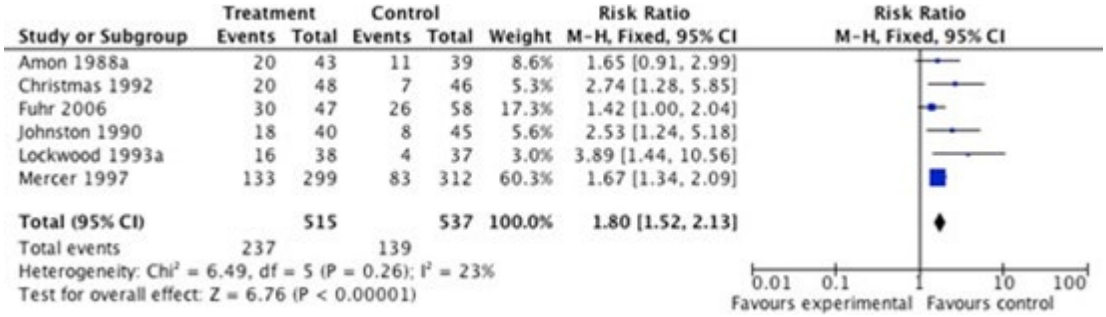
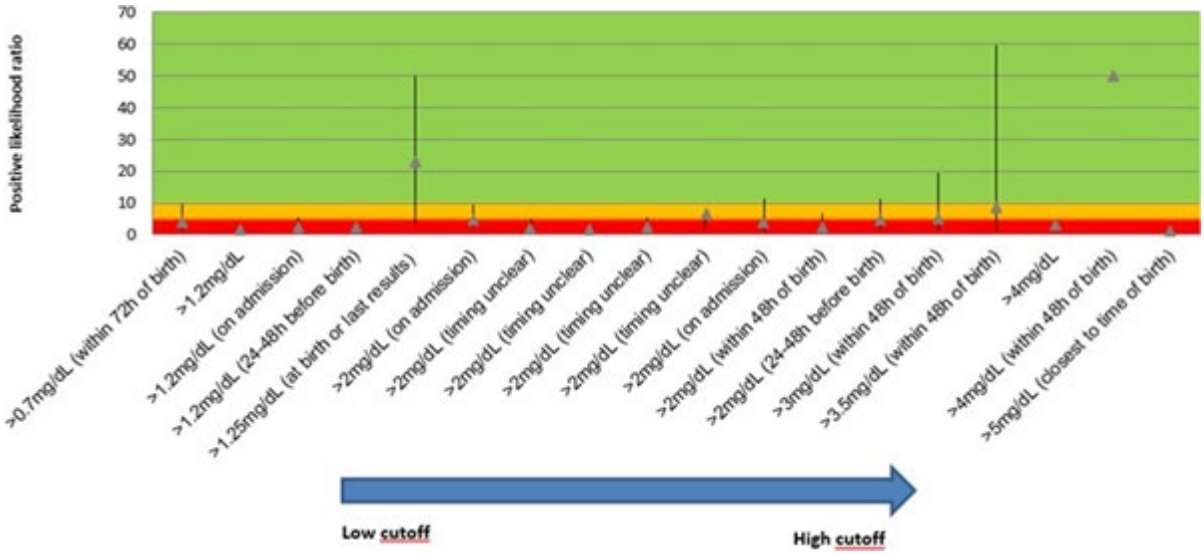


Figure 22: Delivery delayed ≥ 7 days



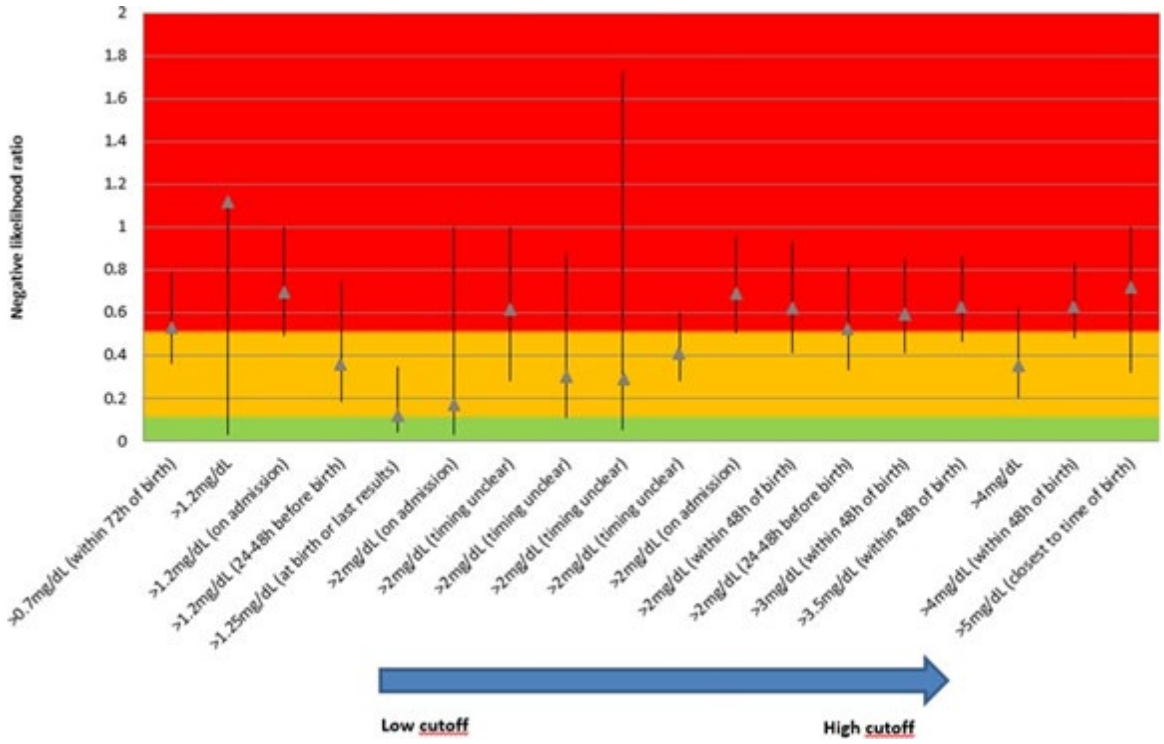
I.5 Identifying infection in women with P-PROM

Figure 23: Predictive value of monitoring women with preterm pre-labour rupture of membranes – Positive likelihood ratio for C-reactive protein



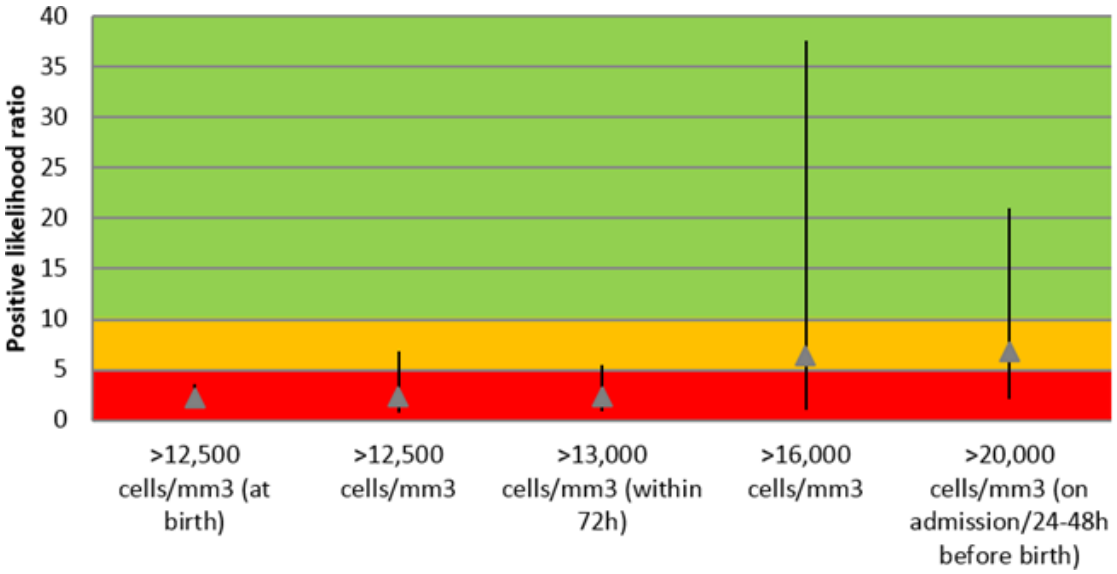
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 24: Predictive value of monitoring women with preterm pre-labour rupture of membranes – Negative likelihood ratio for C-reactive protein



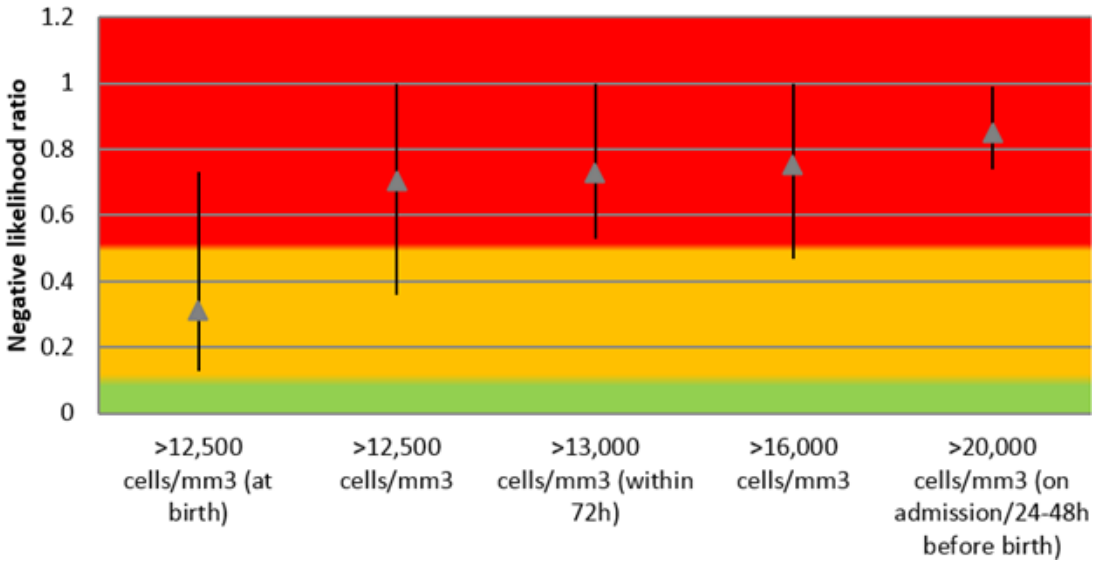
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 25: Predictive value of monitoring women with preterm pre-labour rupture of membranes – Positive likelihood ratio for white blood cell count



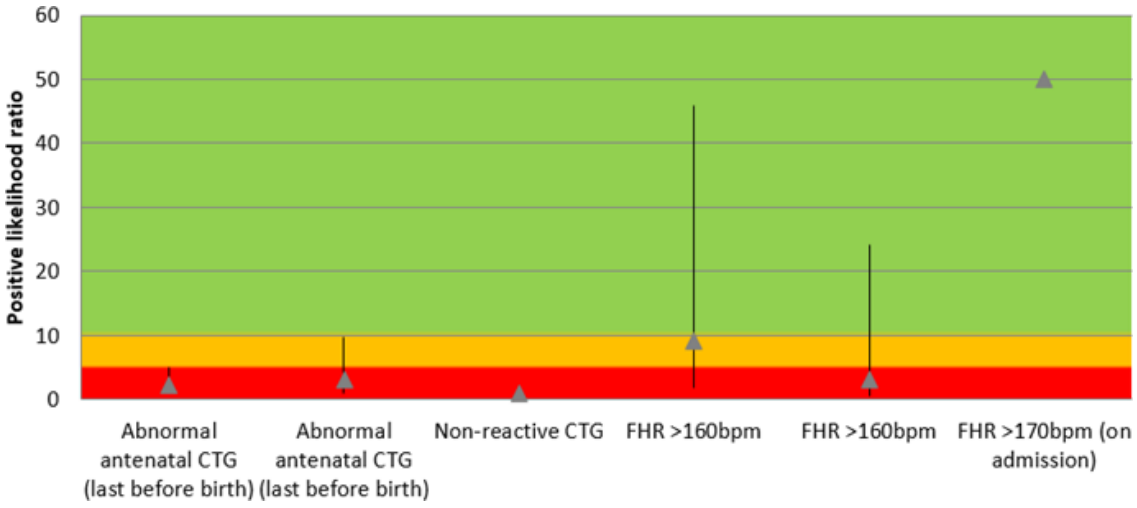
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 26: Predictive value of monitoring women with preterm pre-labour rupture of membranes – Negative likelihood ratio for white blood cell count



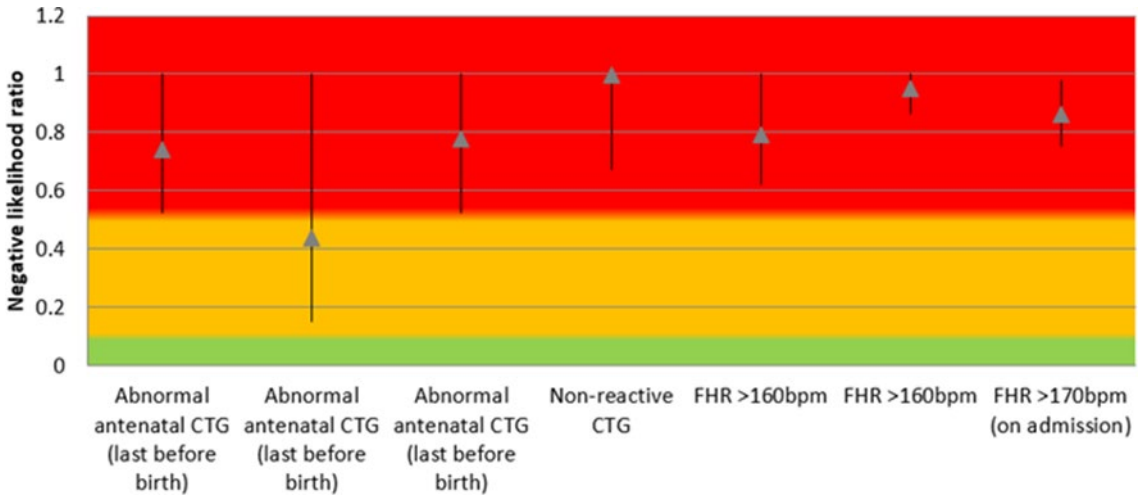
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 27: Predictive value of monitoring women with preterm pre-labour rupture of membranes – Positive likelihood ratio for fetal heart rate



Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 28: Predictive value of monitoring women with preterm pre-labour rupture of membranes – Negative likelihood ratio for fetal heart rate



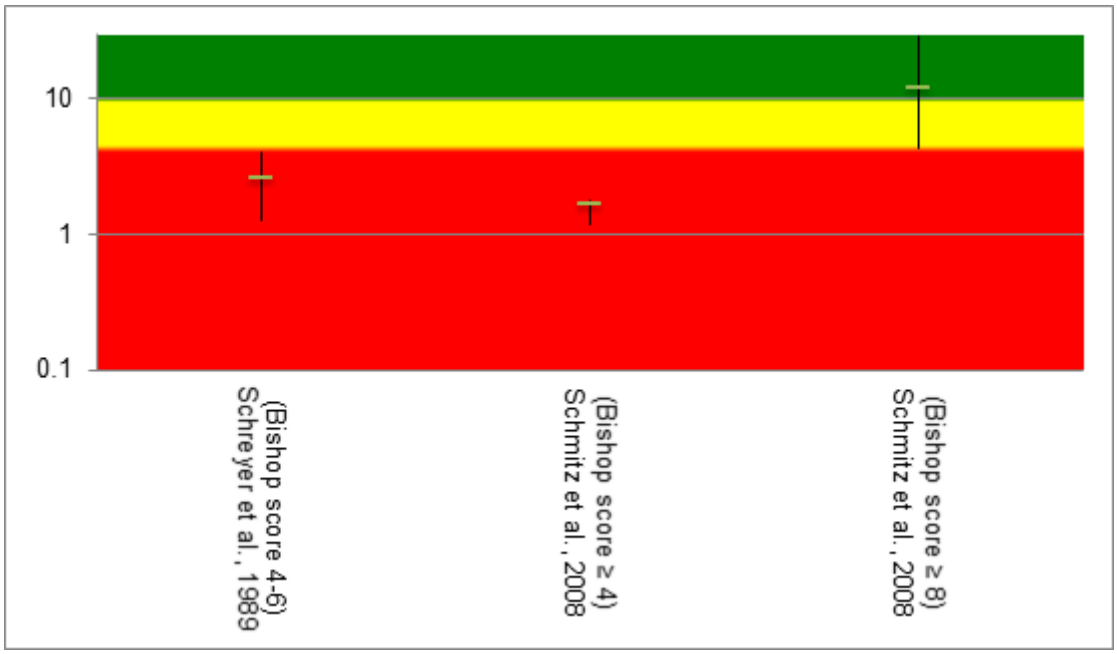
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

I.6 'Rescue' cervical cerclage

No forest plots were generated for this review question

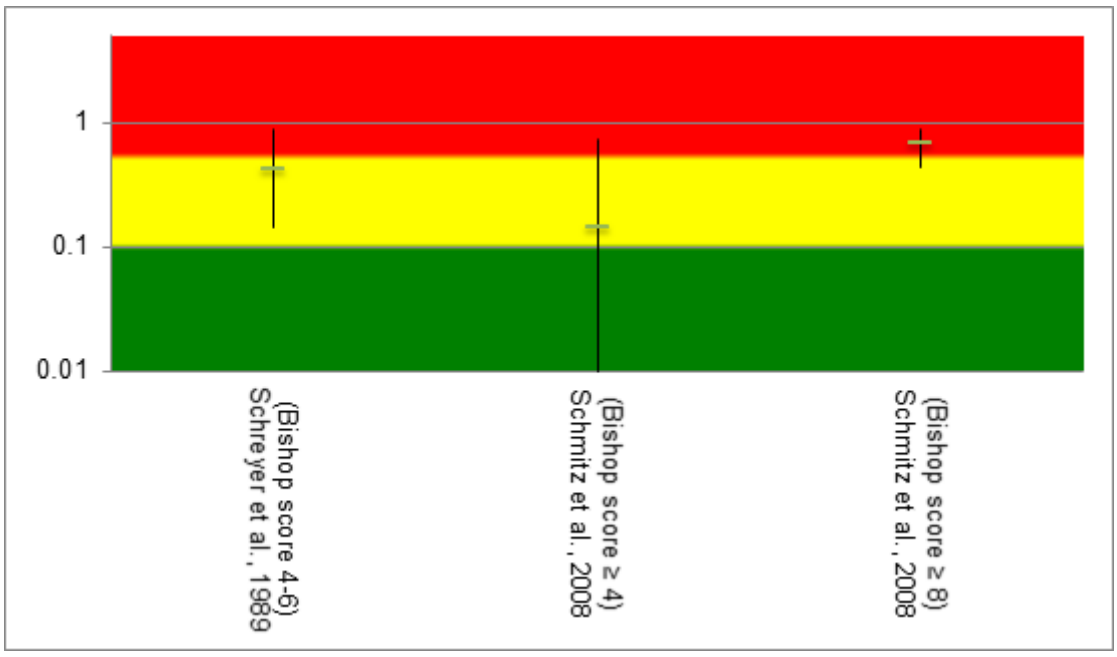
I.7 Diagnosing preterm labour for women with intact membranes

Figure 29: Positive likelihood ratio of Bishop score to diagnose pre-term birth within 48 hours



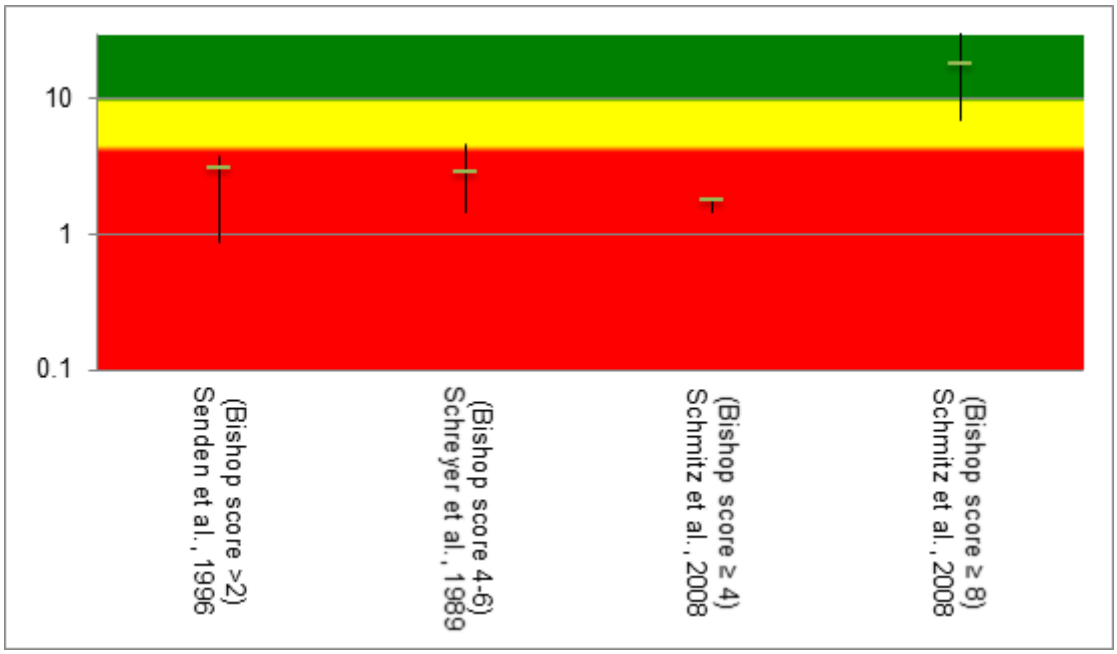
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 30: Negative likelihood ratio of Bishop score to diagnose pre-term birth within 48 hours



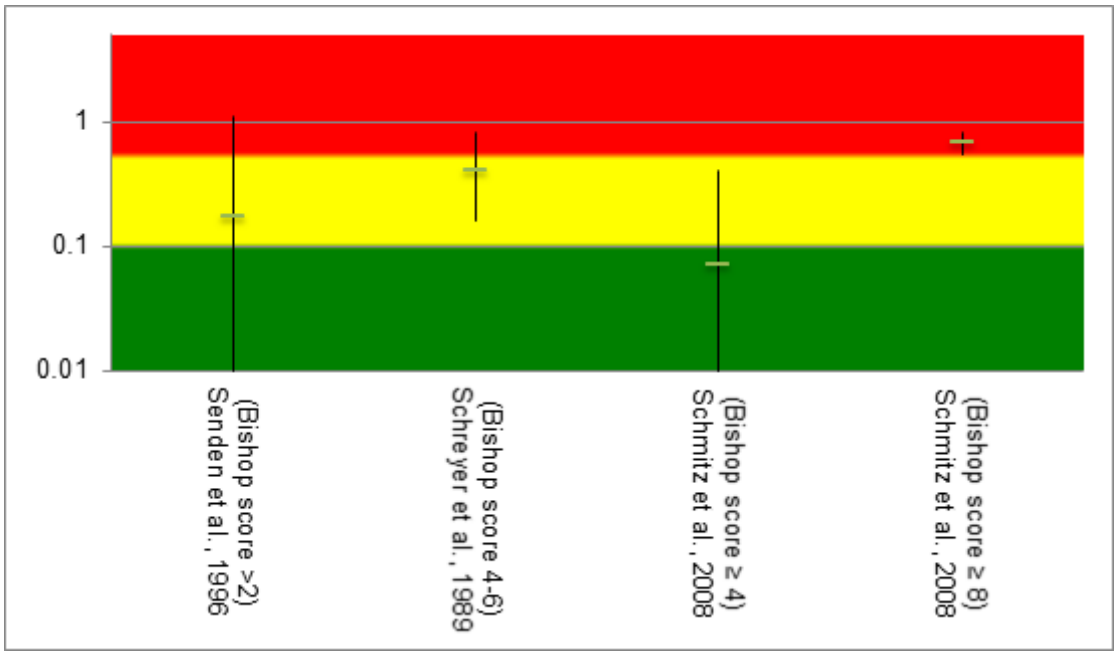
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 31: Positive likelihood ratio of Bishop score to diagnose pre-term birth within 7 days



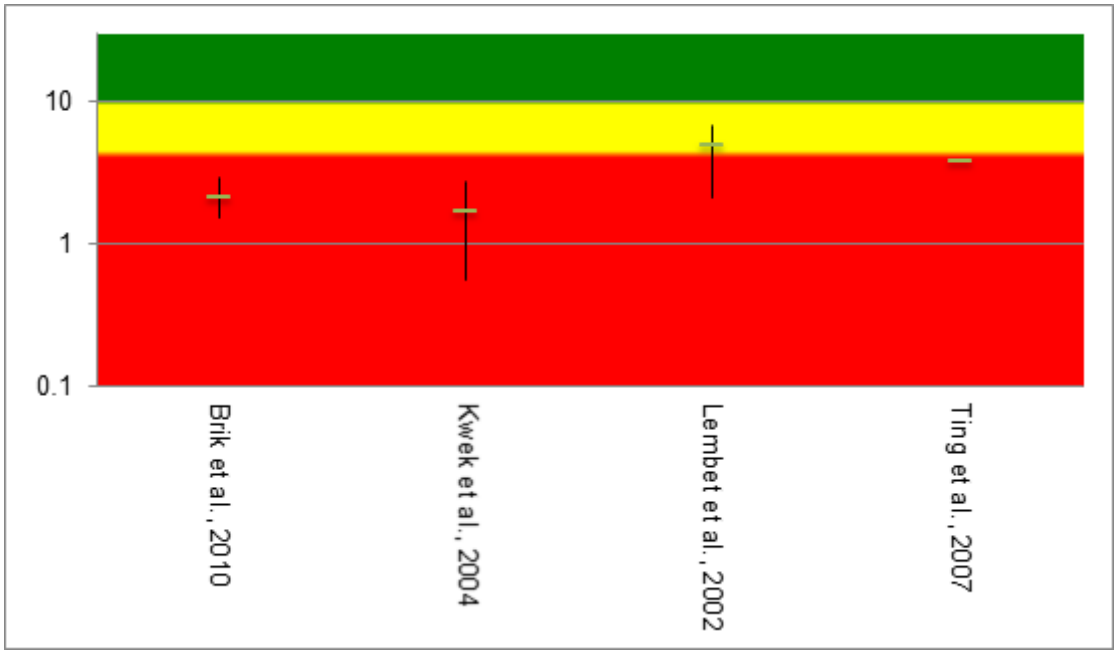
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 32: Negative likelihood ratio of Bishop score to diagnose pre-term birth within 7 days



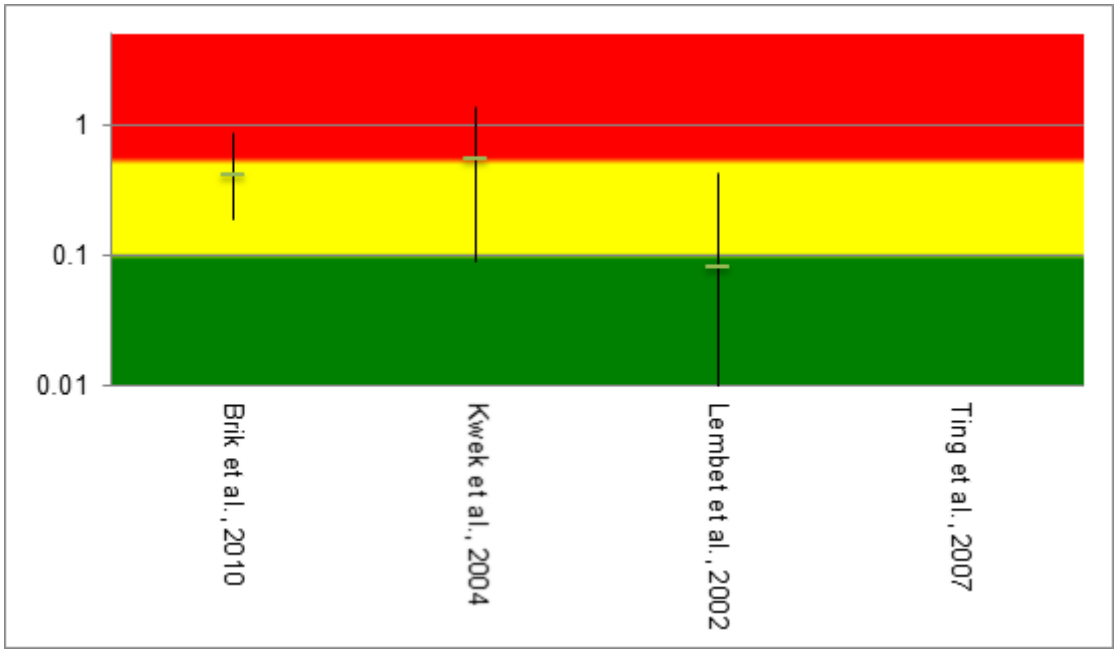
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 33: Positive likelihood ratio of plGFBP-1 to diagnose pre-term birth within 48 hours



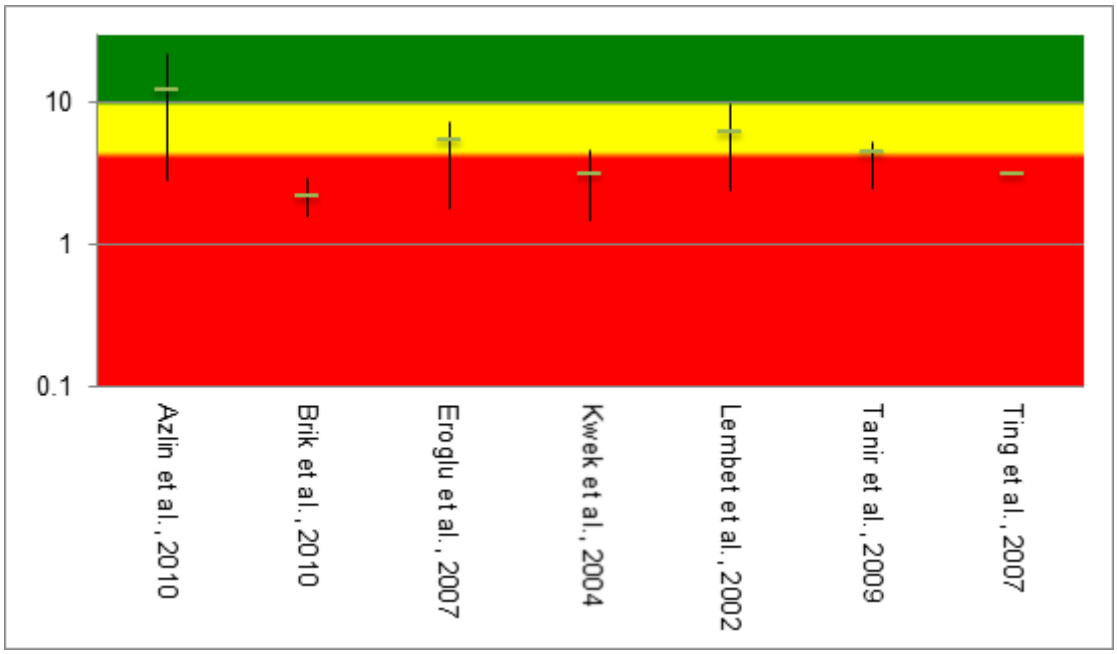
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 34: Negative likelihood ratio of plGFBP-1 to diagnose pre-term birth within 48 hours



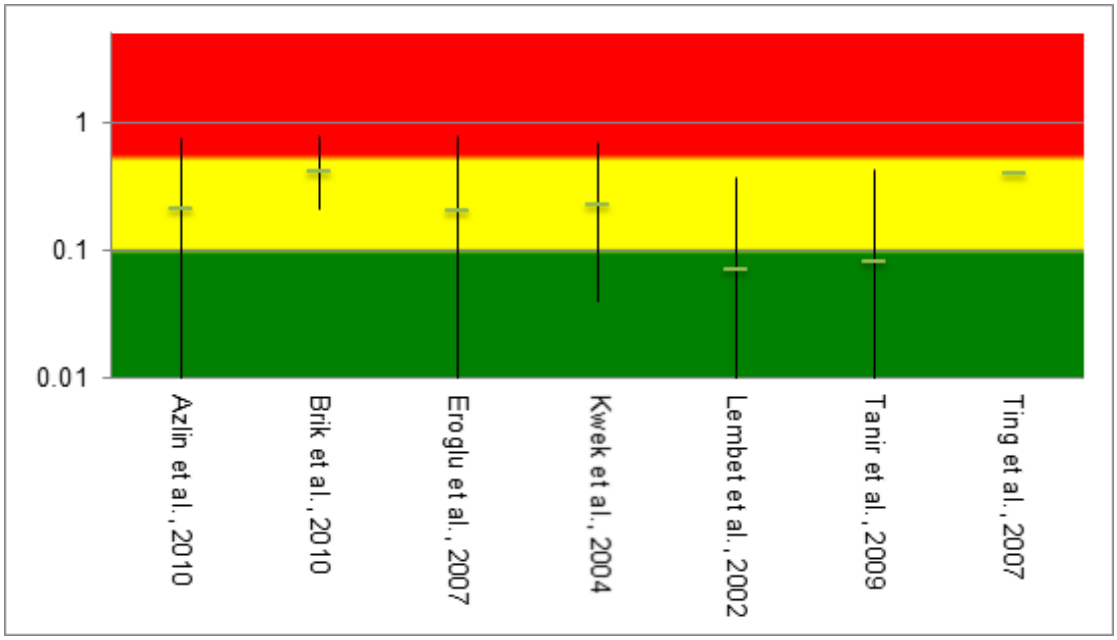
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 35: Positive likelihood ratio of plGFBP-1 to diagnose pre-term birth within 7 days



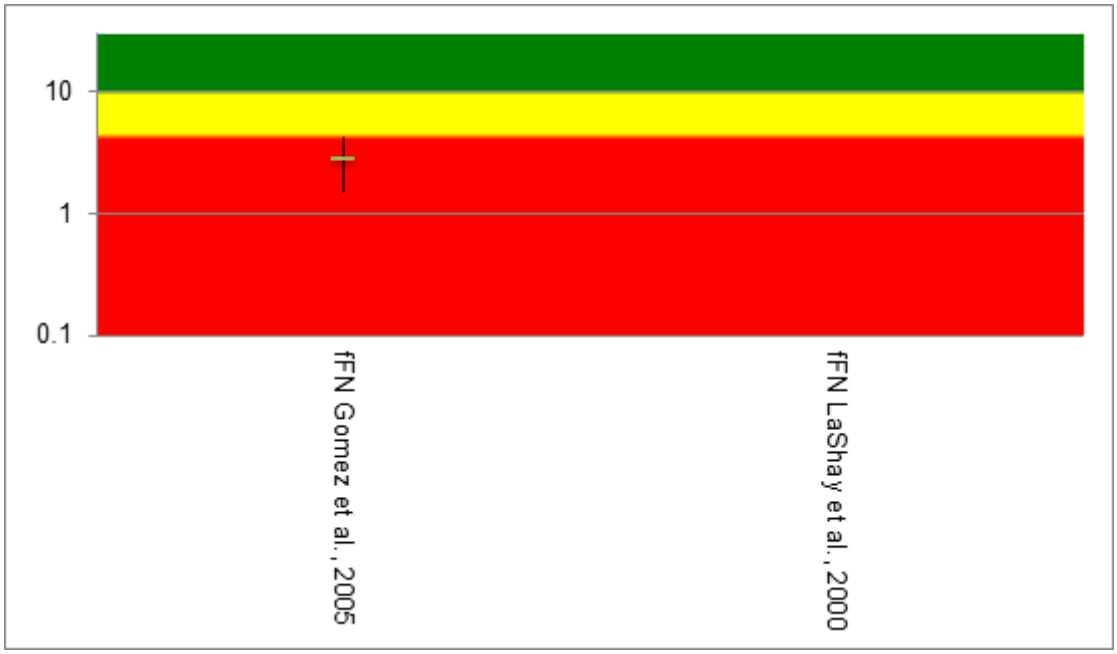
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 36: Negative likelihood ratio of plGFBP-1 to diagnose pre-term birth within 7 days



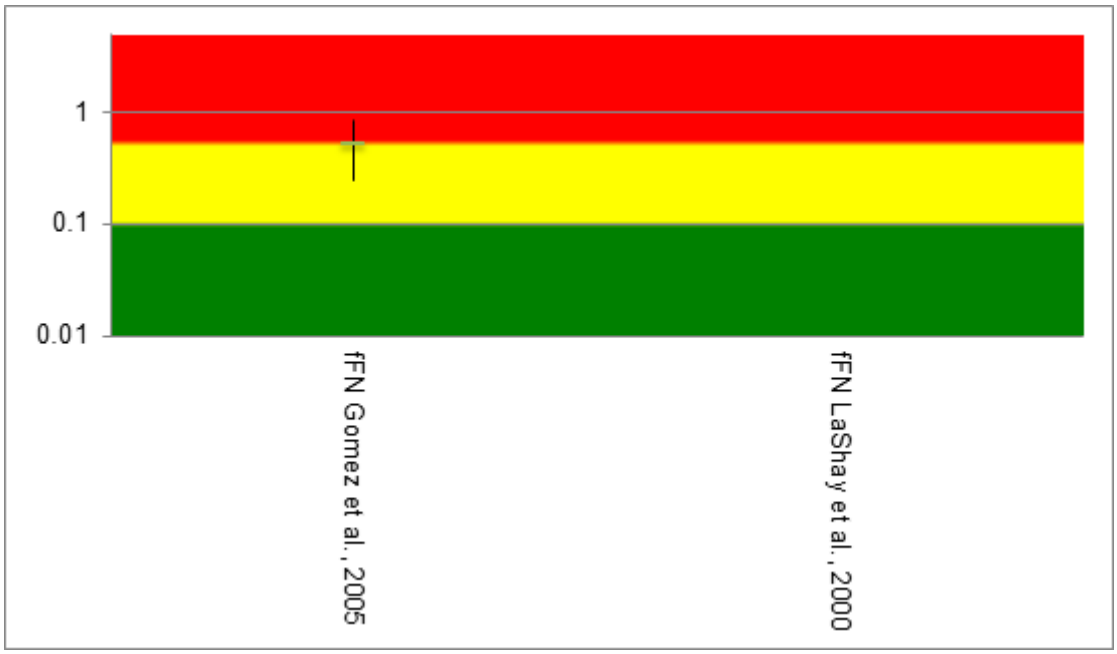
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 37: Positive likelihood ratio of fetal fibronectin to diagnose pre-term birth within 48 hours



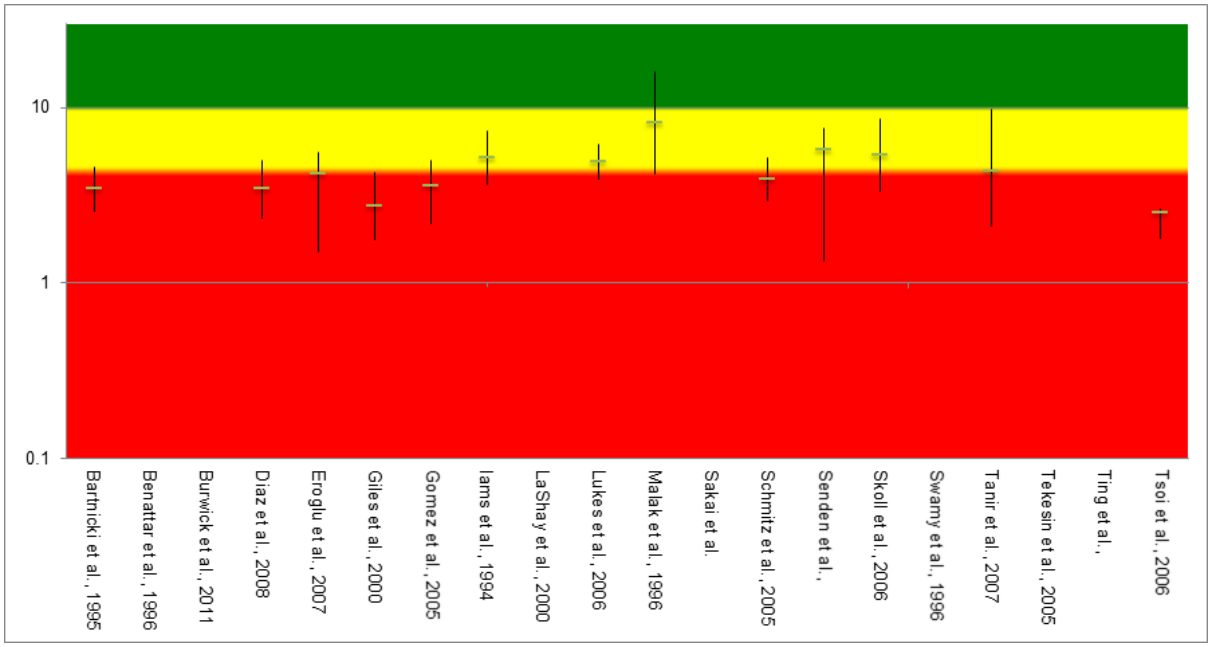
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 38: Negative likelihood ratio of fetal fibronectin to diagnose pre-term birth within 48 hours



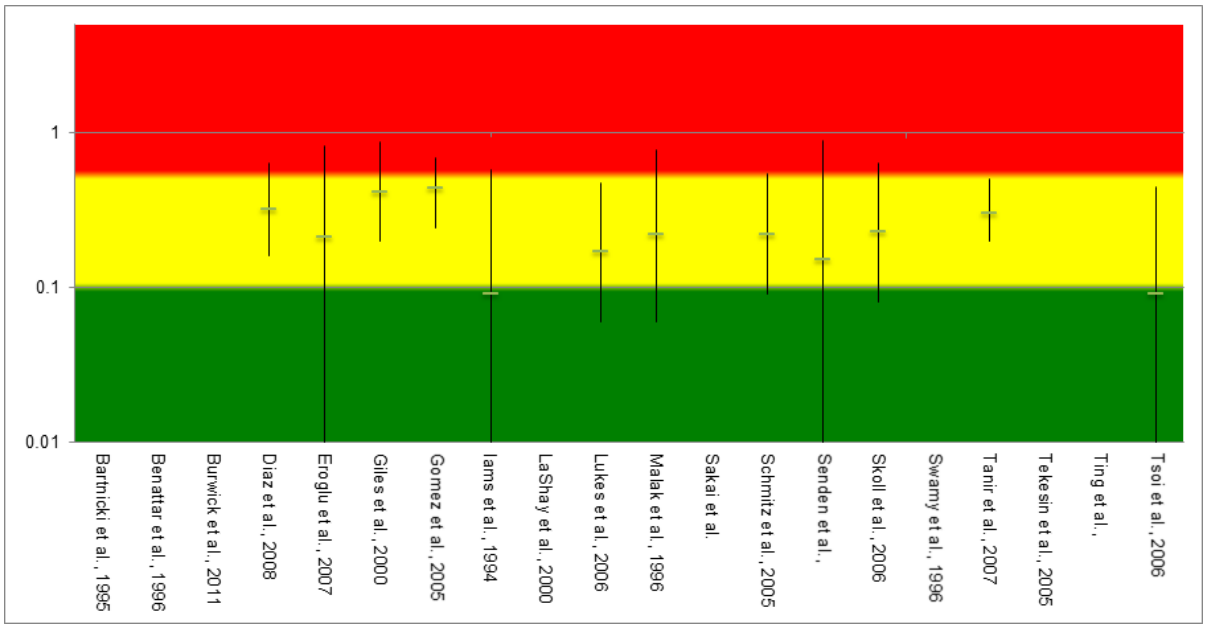
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 39: Positive likelihood ratio of fetal fibronectin to diagnose pre-term birth within 7 days



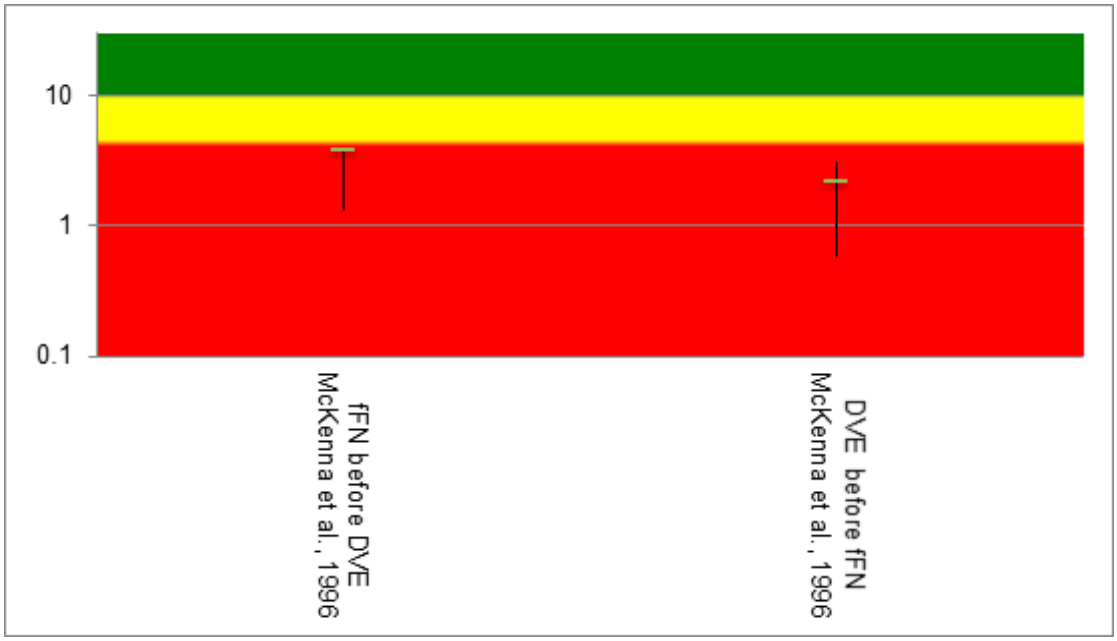
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 40: Negative likelihood ratio of fetal fibronectin to diagnose pre-term birth within 7 days



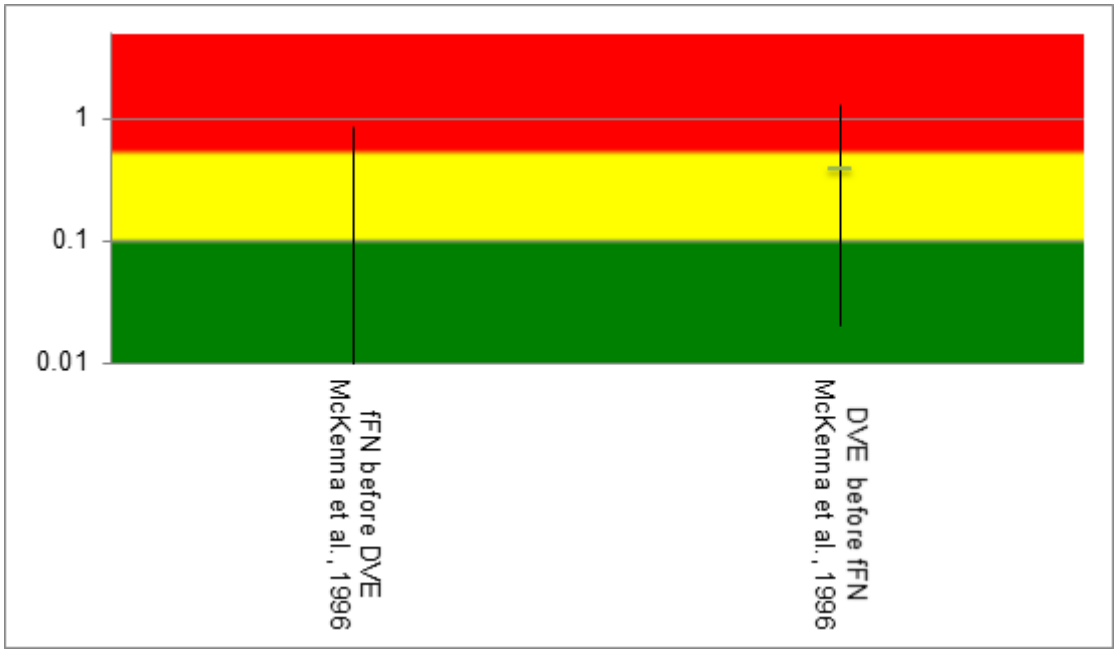
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 41: Positive likelihood ratio of fetal fibronectin and digital examination to diagnose pre-term birth within 7 days



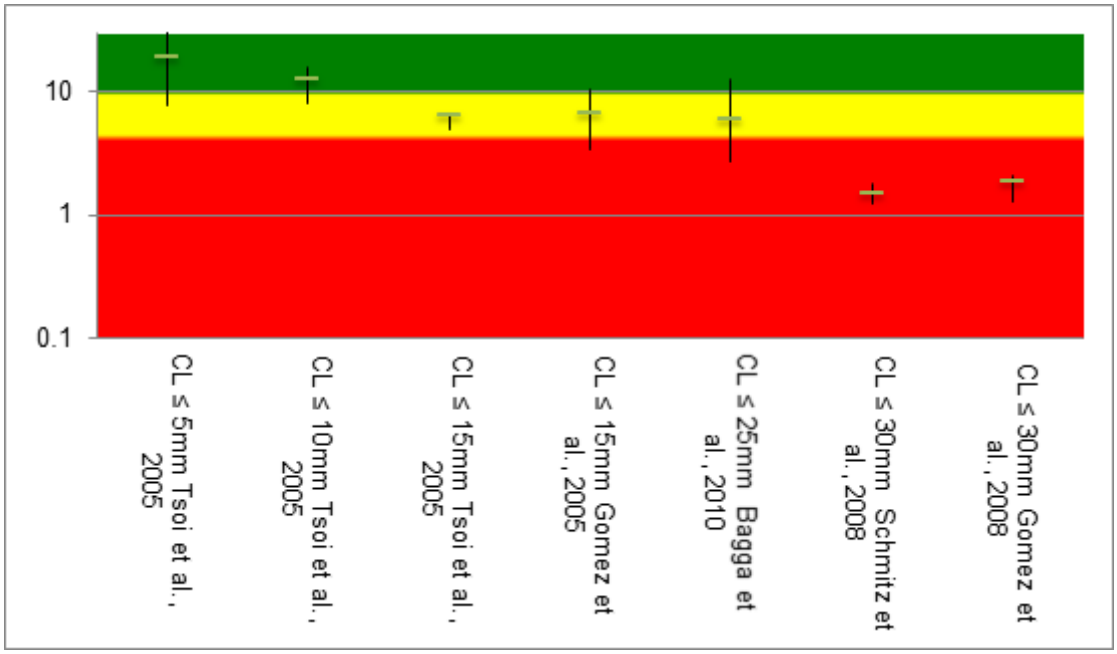
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 42: Negative likelihood ratio of fetal fibronectin and digital examination to diagnose pre-term birth within 7 days



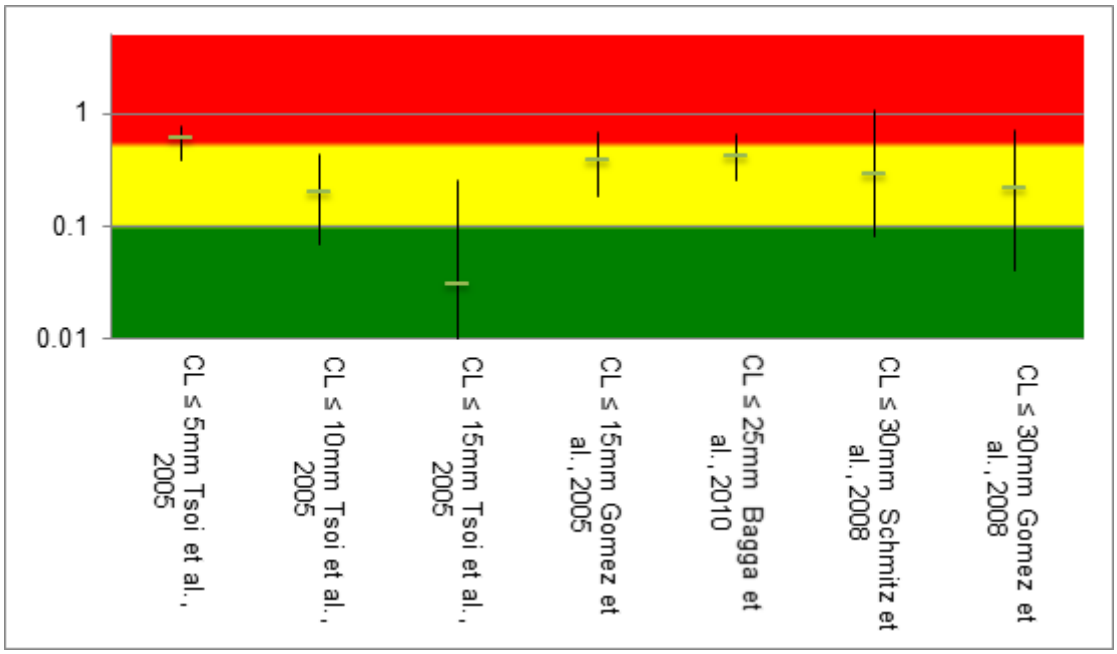
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 43: Positive likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 48 hours



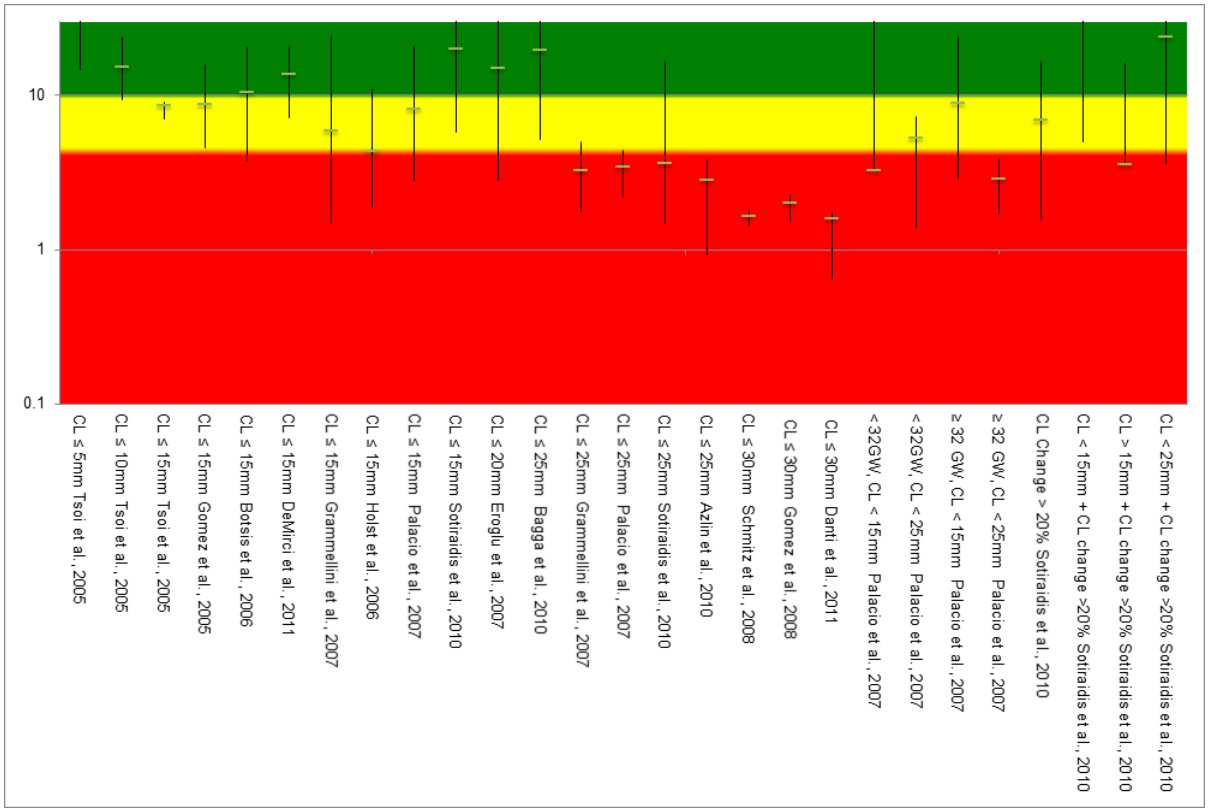
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 44: Negative likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 48 hours



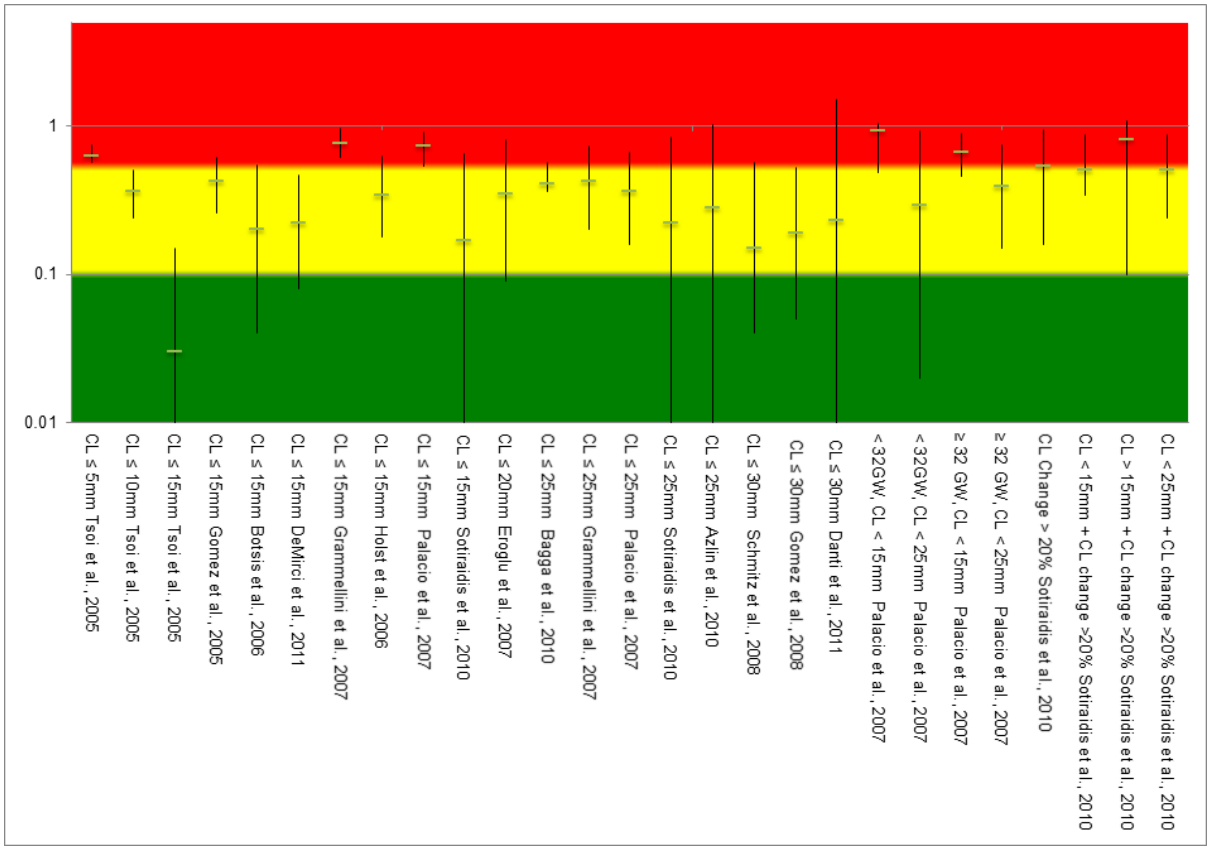
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 45: Positive likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 7 days



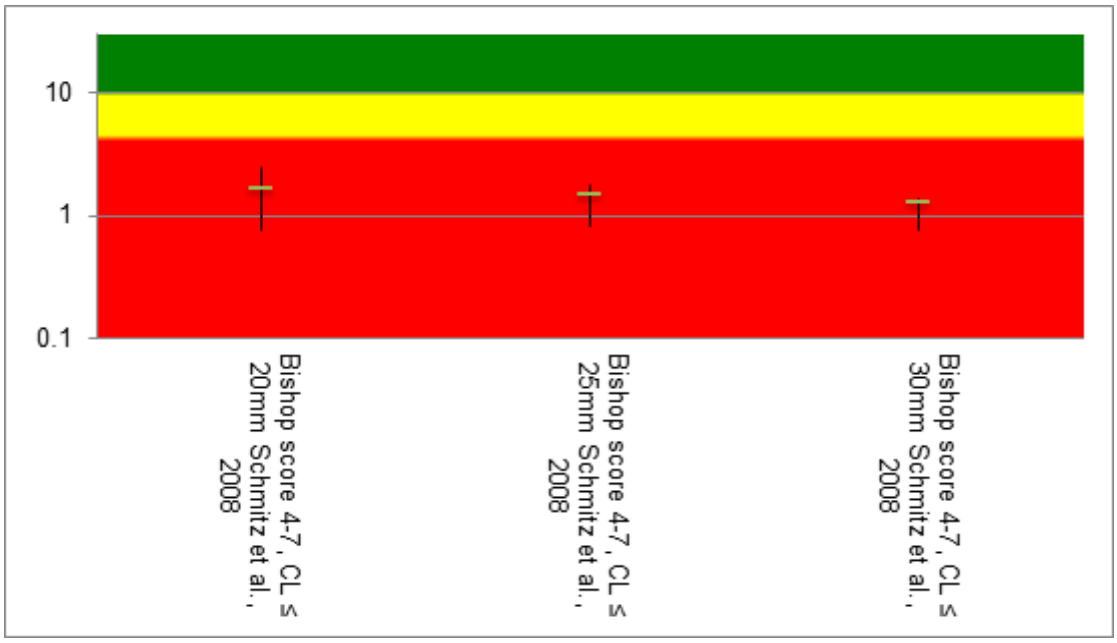
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 46: Negative likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 7 days



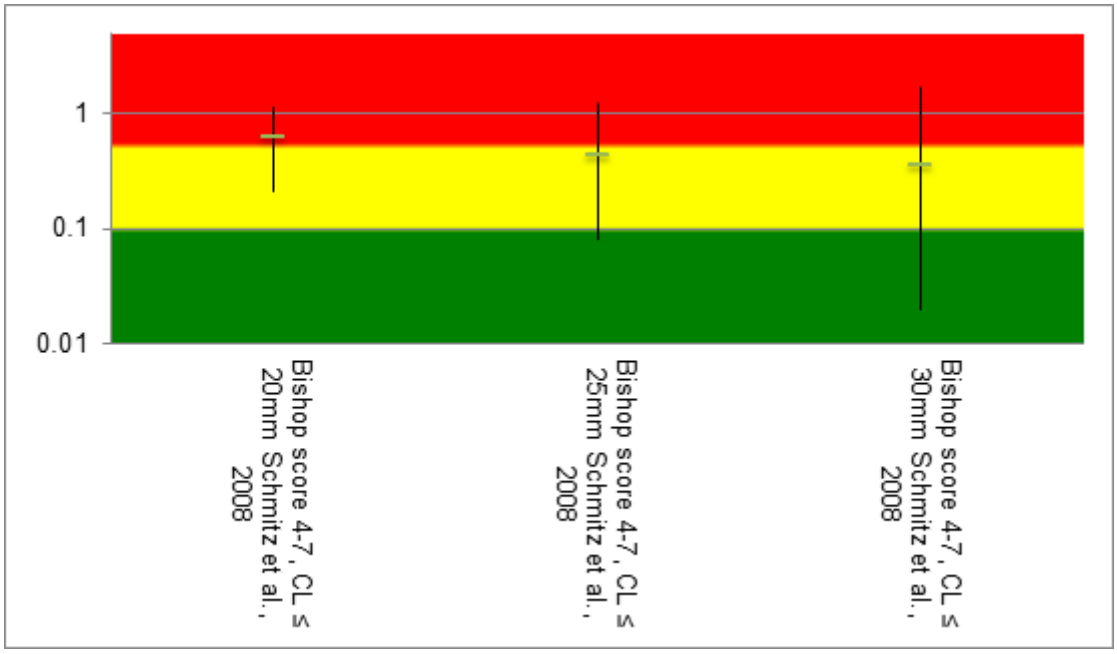
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 47: Positive likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 48 hours in women with a Bishop score of 4-7



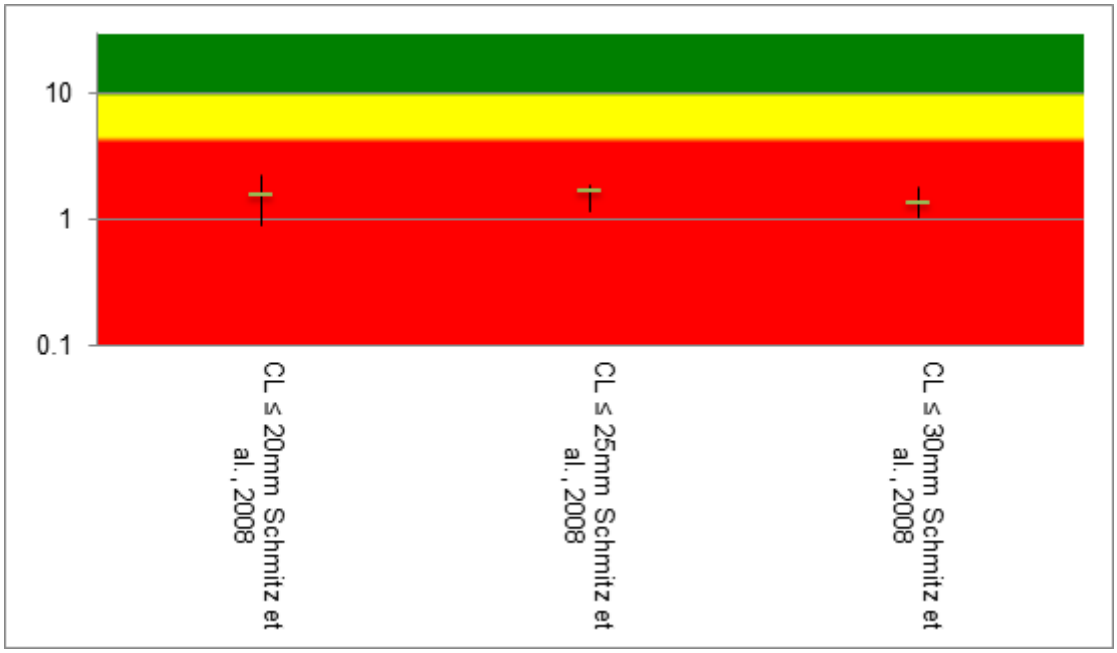
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 48: Negative likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 48 hours in women with a Bishop score of 4-7



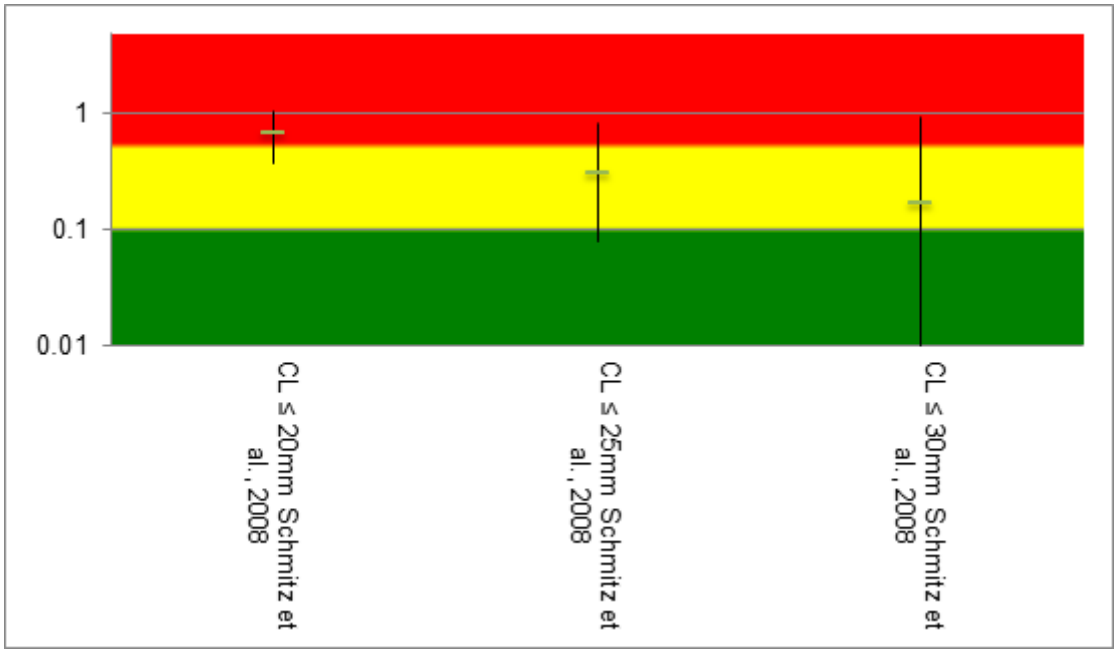
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 49: Positive likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 7 days in women with a Bishop score of 4-7



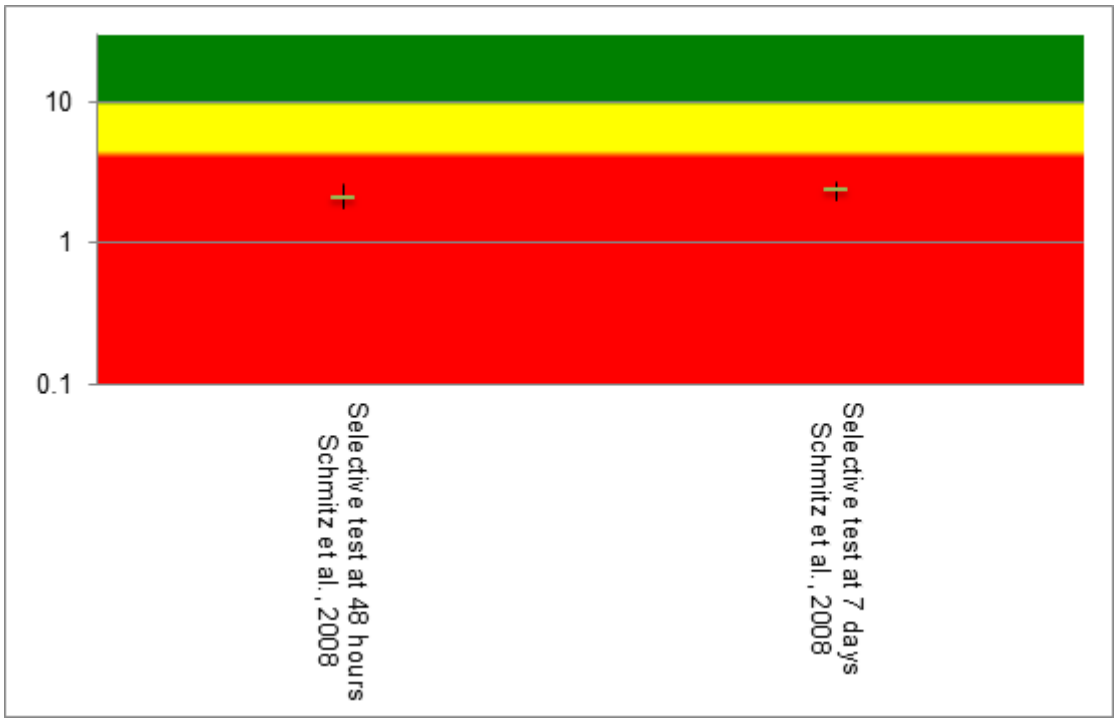
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 50: Negative likelihood ratio of cervical length (measured by transvaginal ultrasound) to diagnose pre-term birth within 7 days in women with a Bishop score of 4-7



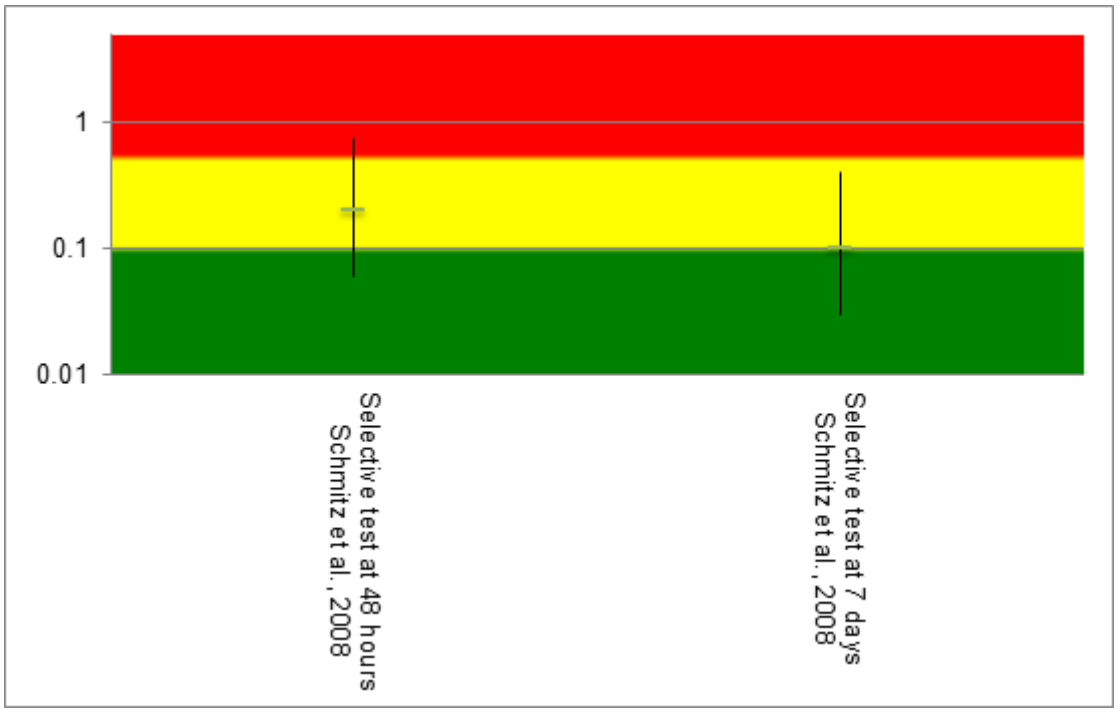
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 51: Positive likelihood ratio of a selective test (using cervical length measured by transvaginal ultrasound and a Bishop score) to diagnose pre- term birth within 48 hours and 7 days



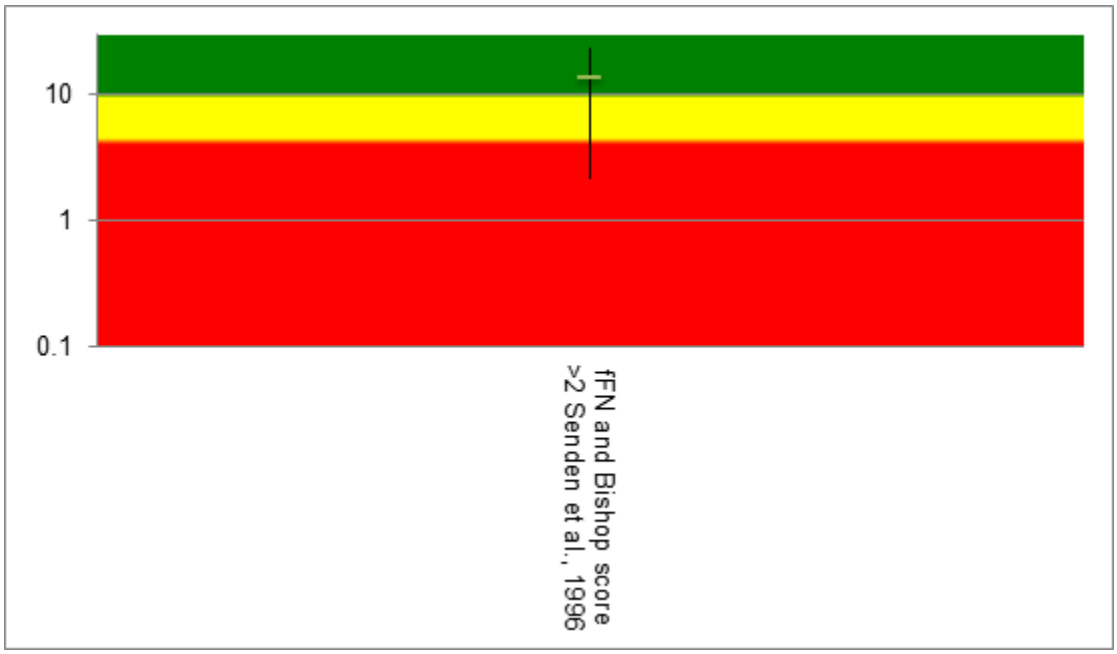
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 52: Negative likelihood ratio of a selective test (using cervical length measured by transvaginal ultrasound and a Bishop score) to diagnose pre- term birth within 48 hours and 7 days



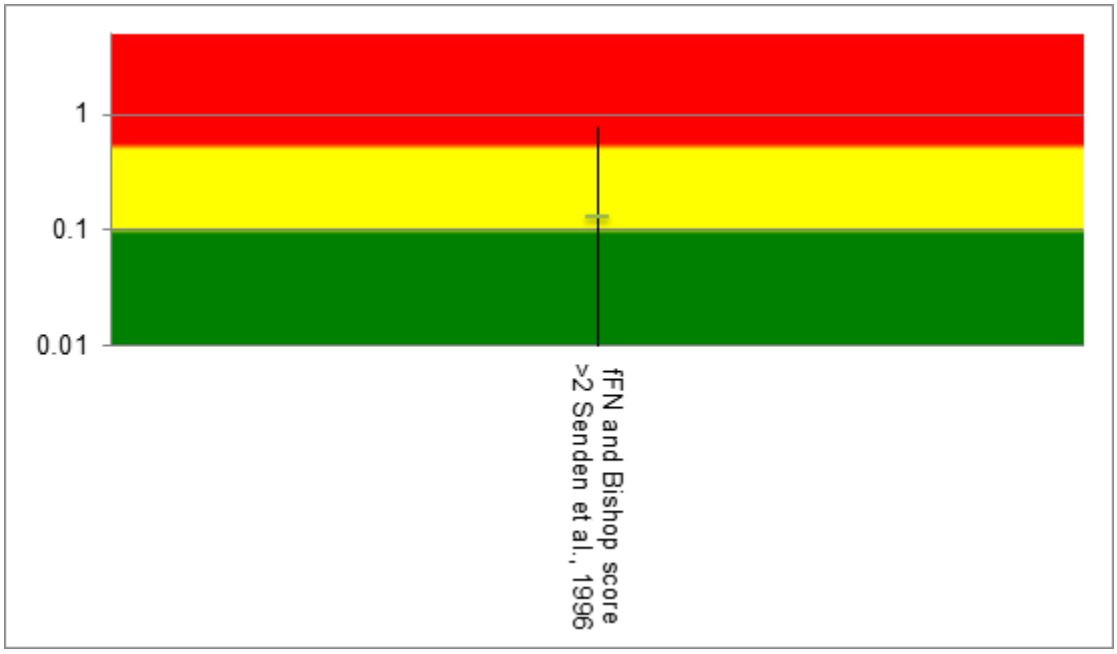
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 53: Positive likelihood ratio for fetal fibronectin score and Bishop score to diagnose pre-term birth within 7 days



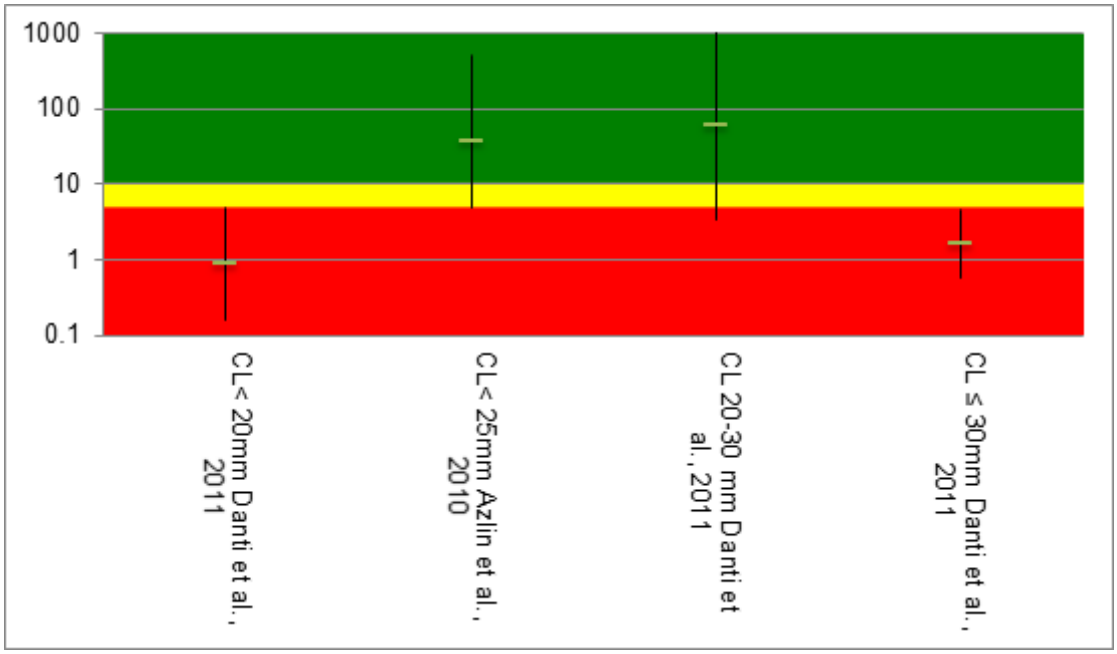
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 54: Negative likelihood ratio for fetal fibronectin score and Bishop score to diagnose pre-term birth within 7 days



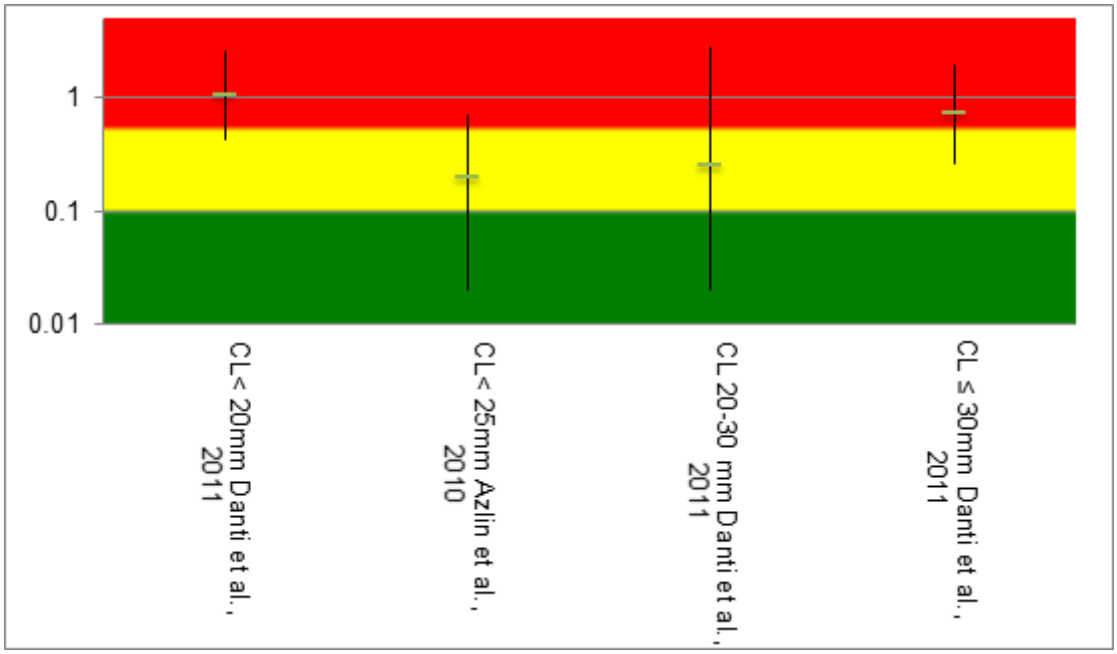
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 55: Positive likelihood ratio for plGFBP-1 to diagnose pre-term birth within 7 days in women with different cervical lengths



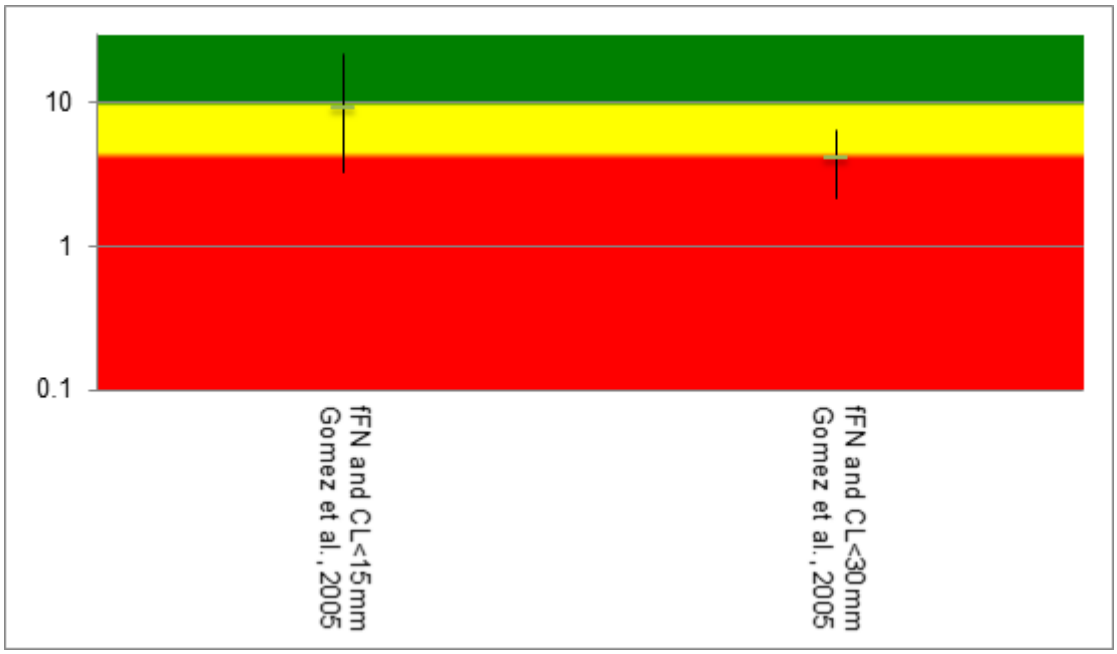
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 56: Negative likelihood ratio for plGFBP-1 to diagnose pre-term birth within 7 days in women with different cervical lengths



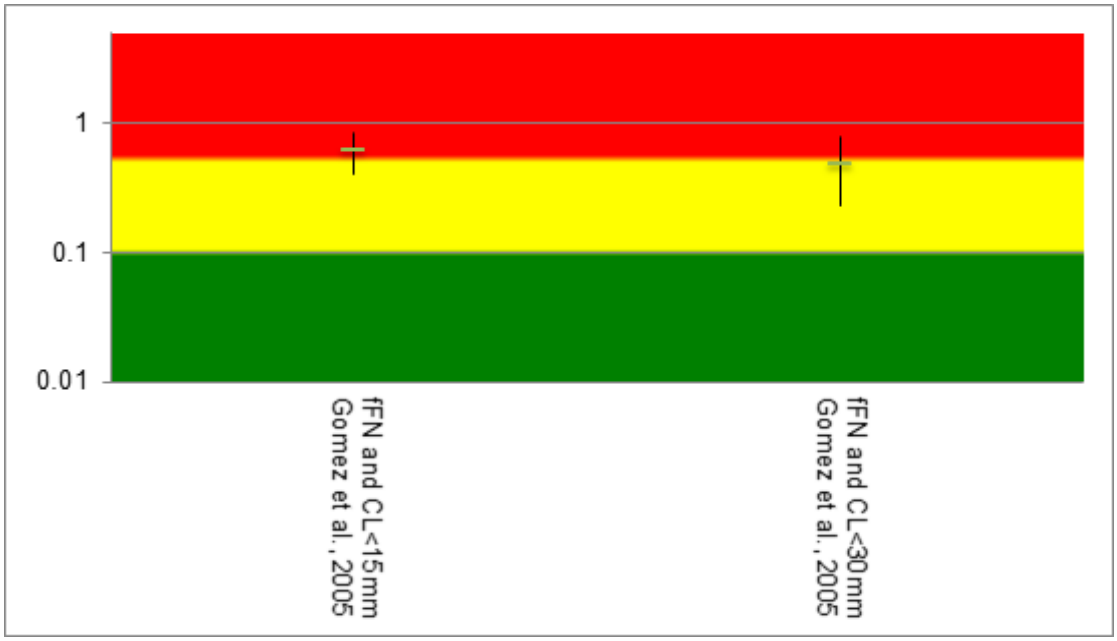
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 57: Positive likelihood ratio for fetal fibronectin to diagnose pre-term birth within 48 hours in women with different cervical lengths



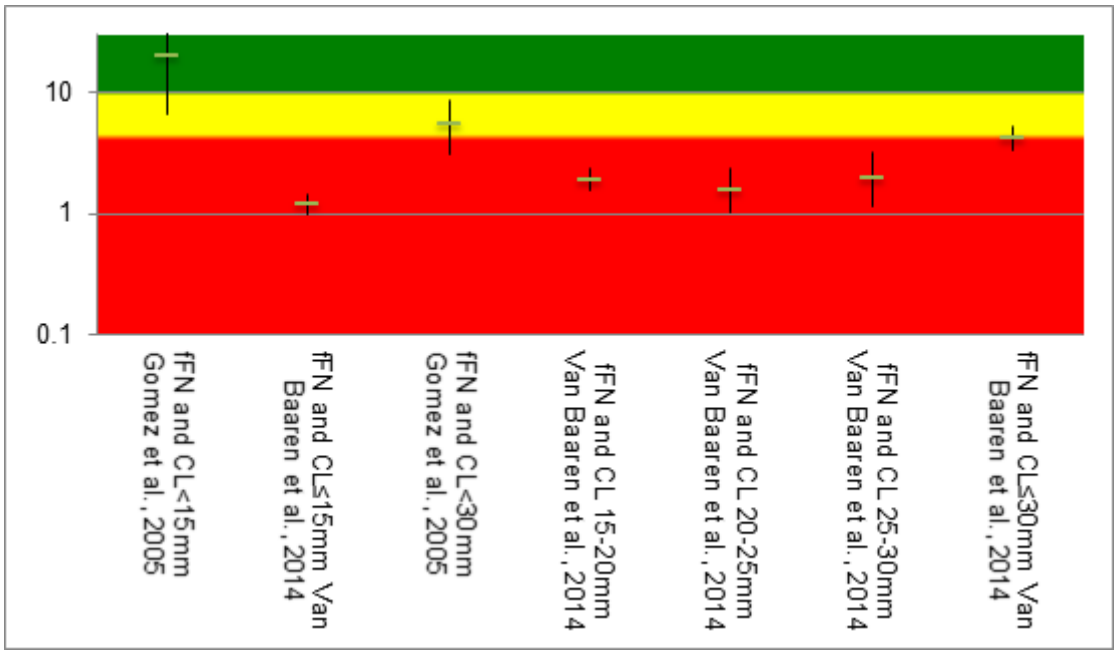
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 58: Negative likelihood ratio for fetal fibronectin to diagnose pre-term birth within 48 hours in women with different cervical lengths



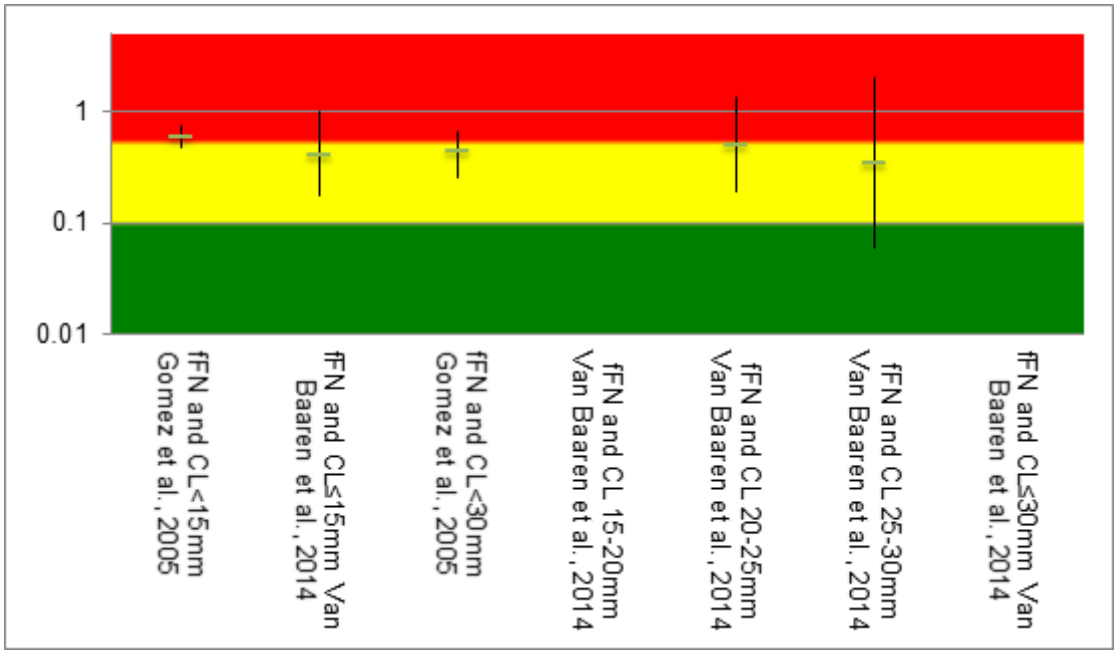
Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 59: Positive likelihood ratio for fetal fibronectin to diagnose pre-term birth within 7 days in women with different cervical lengths



Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

Figure 60: Negative likelihood ratio for fetal fibronectin to diagnose pre-term birth within 7 days in women with different cervical lengths



Colours indicate diagnostic thresholds – Green: very useful; Yellow: moderately useful; Red: not useful

I.8 A. 8 Maternal corticosteroids

I.8.1 Different gestations

Single-course corticosteroids versus placebo or expectant management

Figure 61: Fetal and neonatal mortality

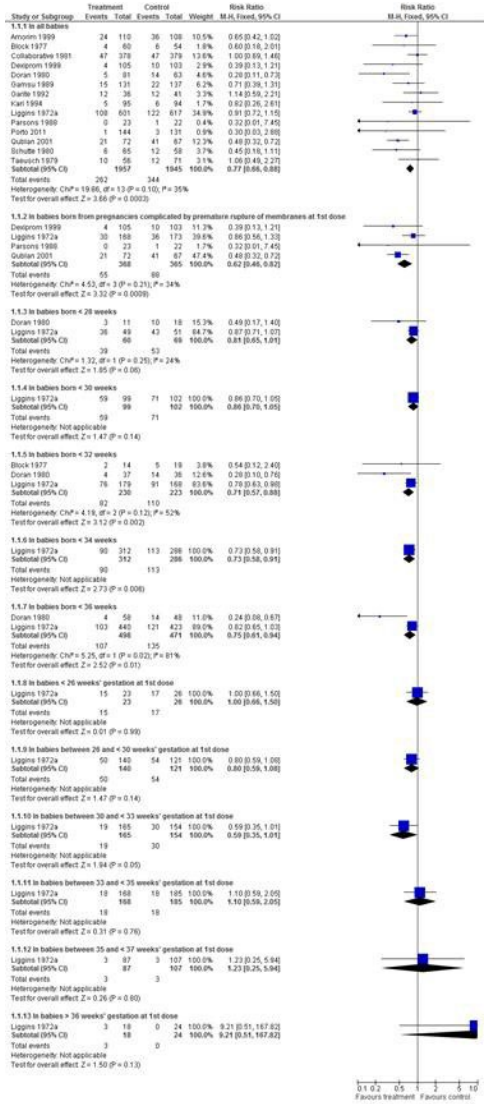


Figure 62: Cerebroventricular haemorrhage

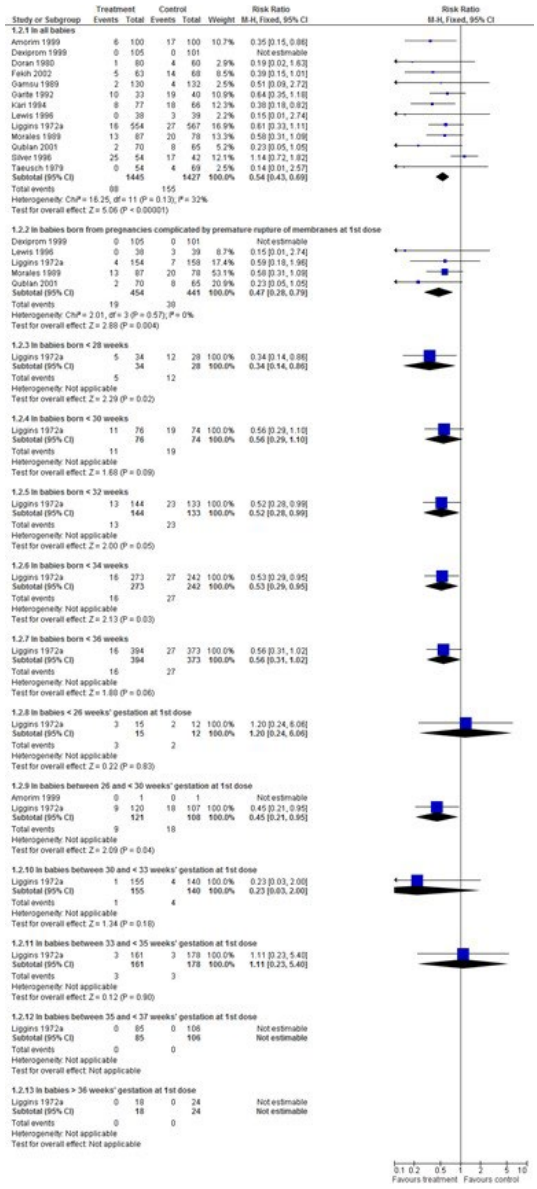


Figure 63: Intraventricular haemorrhage – grades 3 or 4

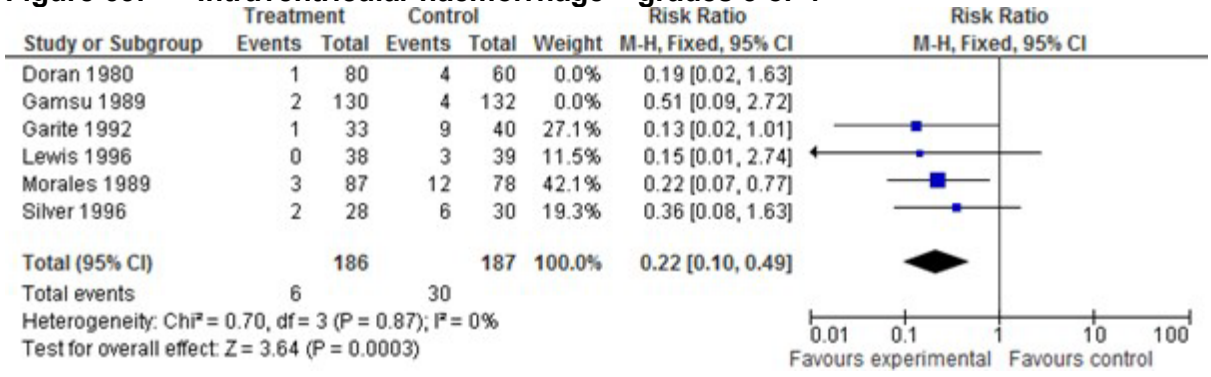


Figure 64: Chronic lung disease

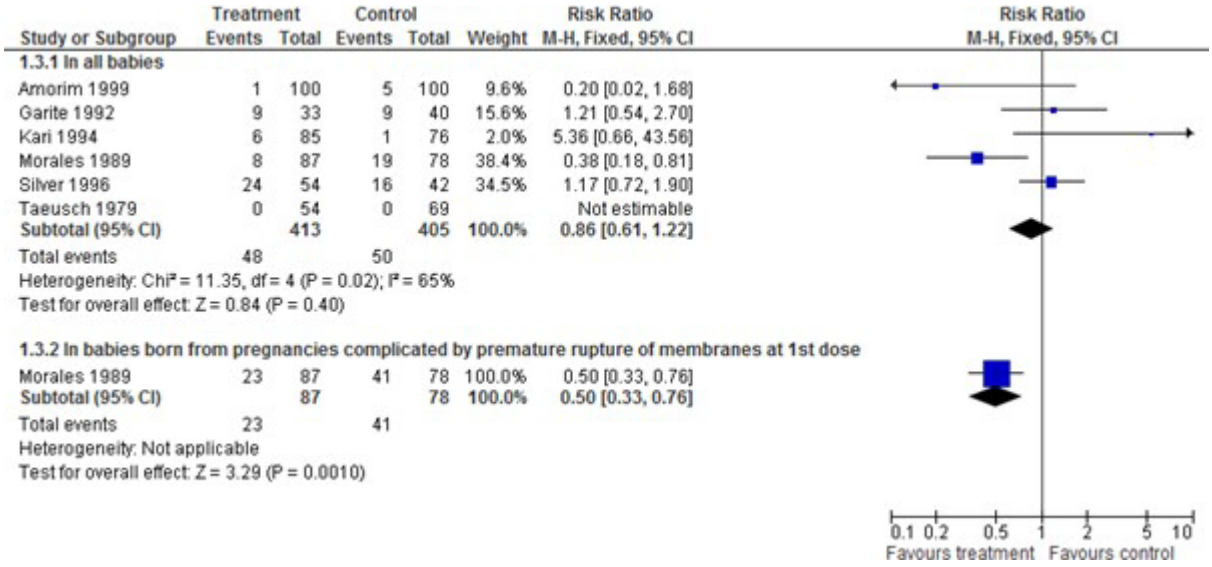


Figure 65: Need for mechanical intervention

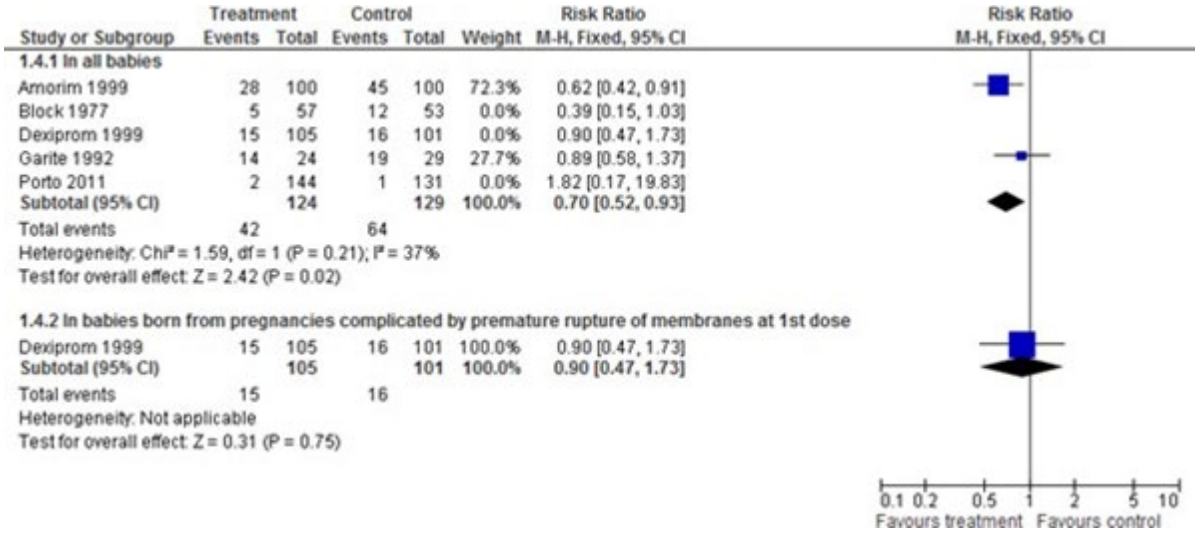


Figure 66: Neonatal sepsis (systemic infection in first 48 hours of life)

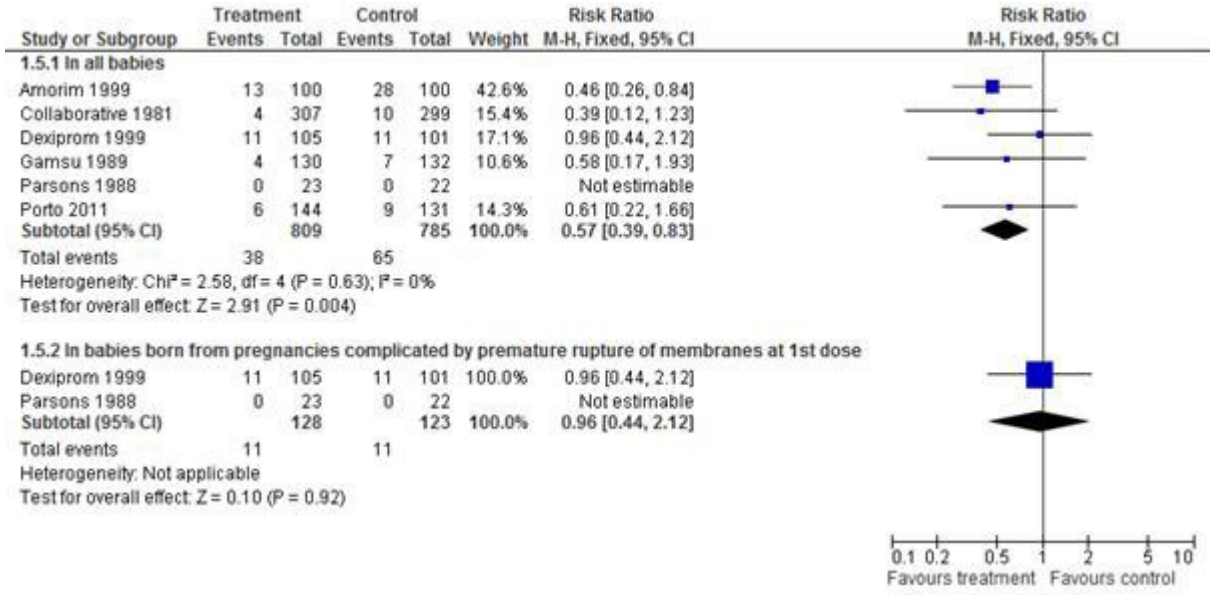


Figure 67: Cerebral palsy in childhood

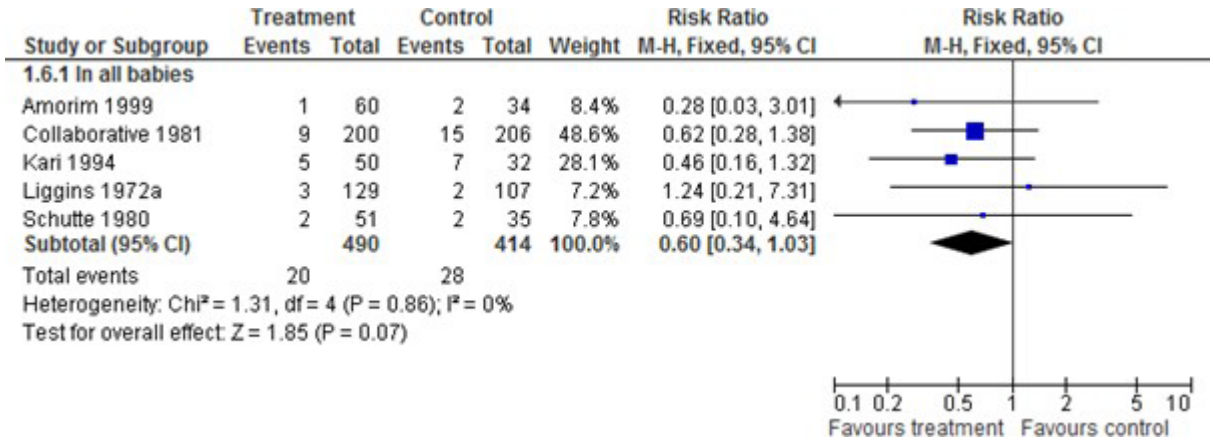


Figure 68: Visual impairment in childhood

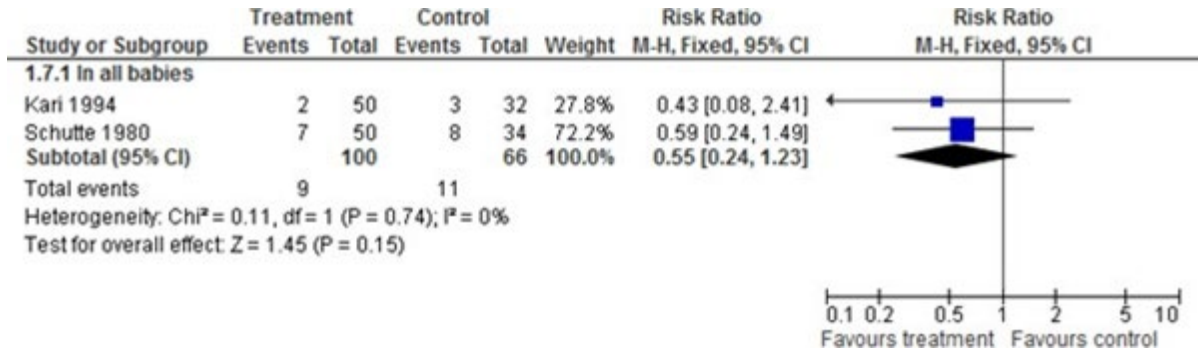


Figure 69: Hearing impairment in childhood

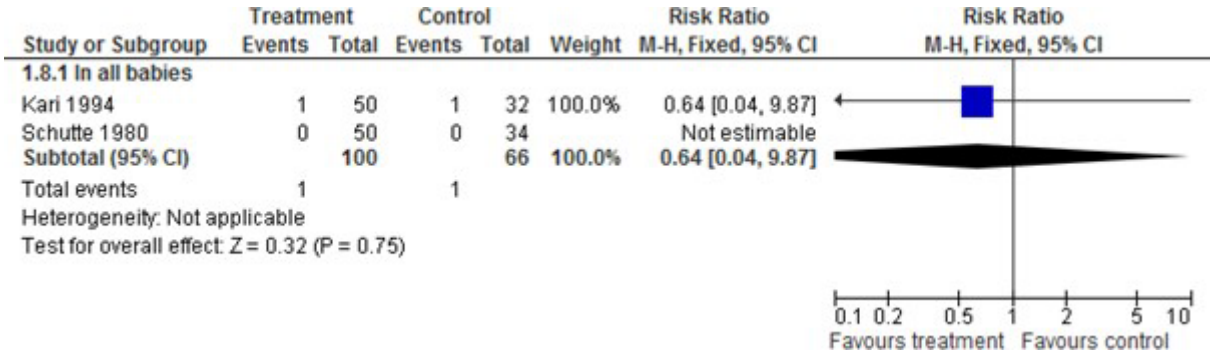


Figure 70: Neurodevelopment delay in childhood

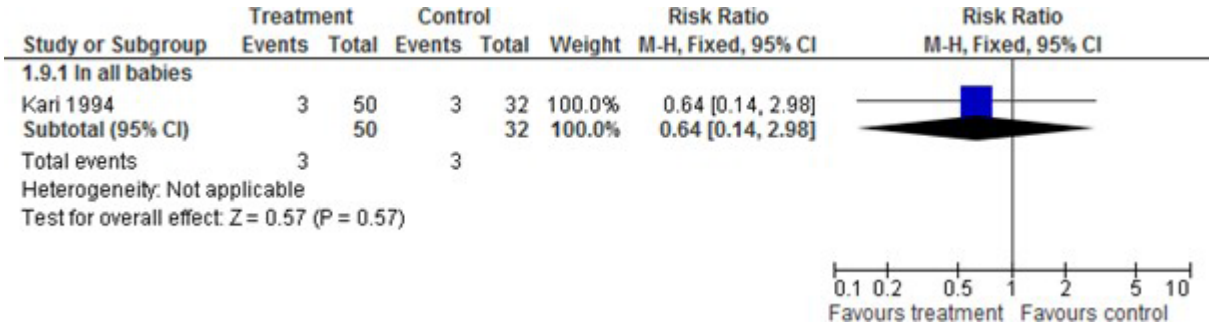


Figure 71: Developmental delay in childhood

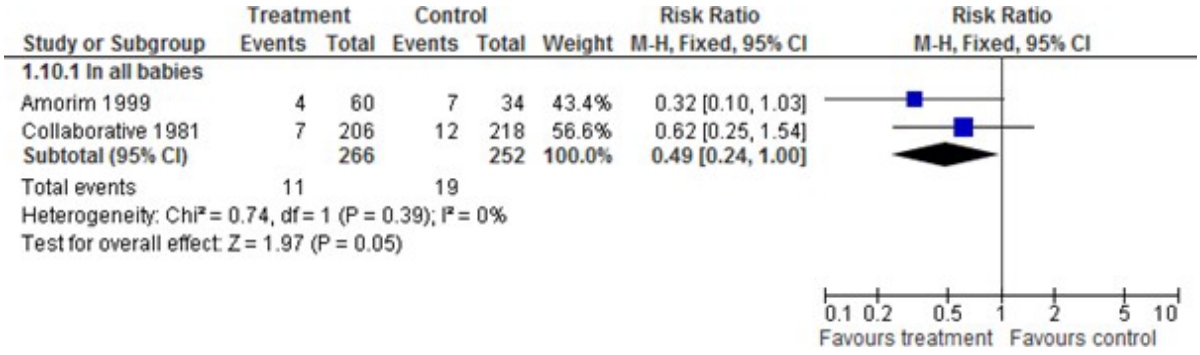


Figure 72: Intellectual impairment in childhood

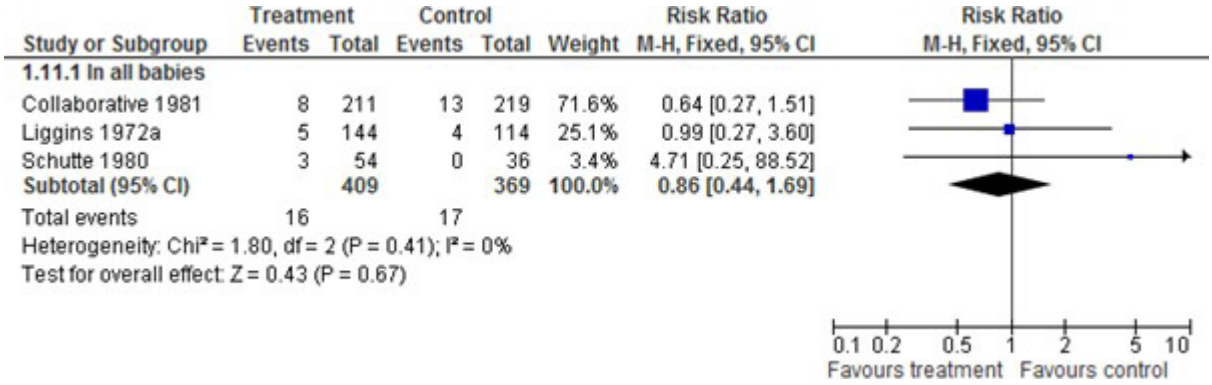


Figure 73: Behavioural/learning difficulties in childhood

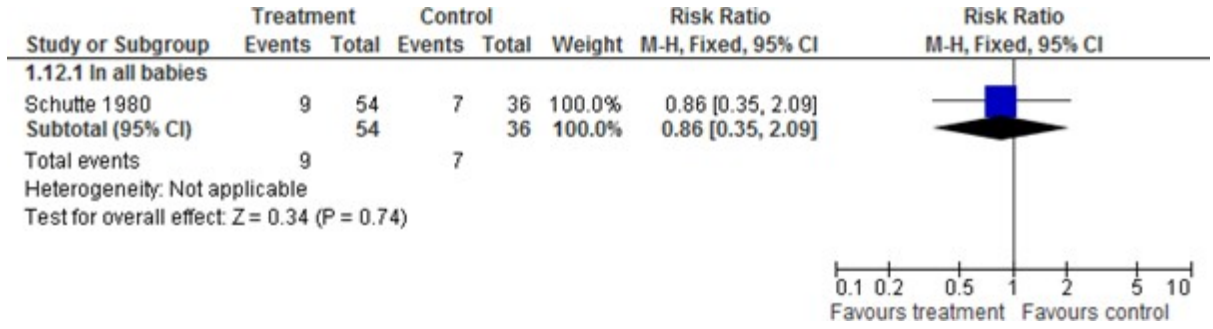


Figure 74: Maternal mortality

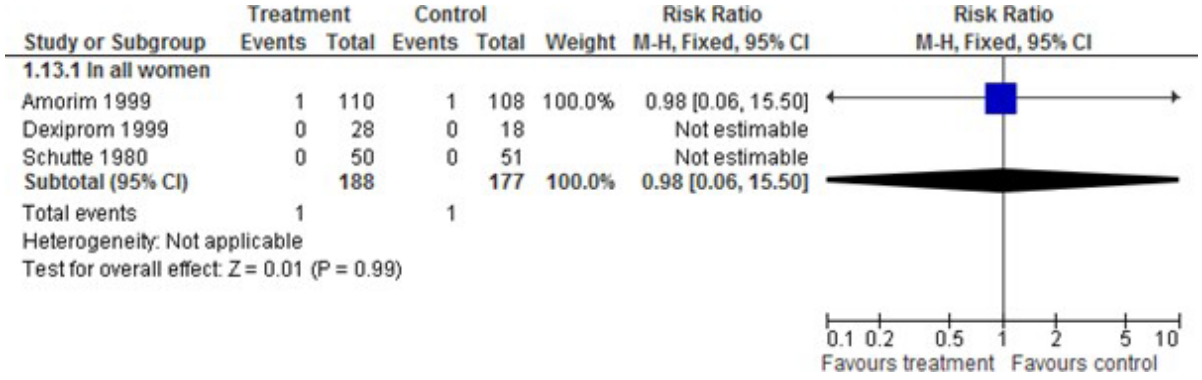
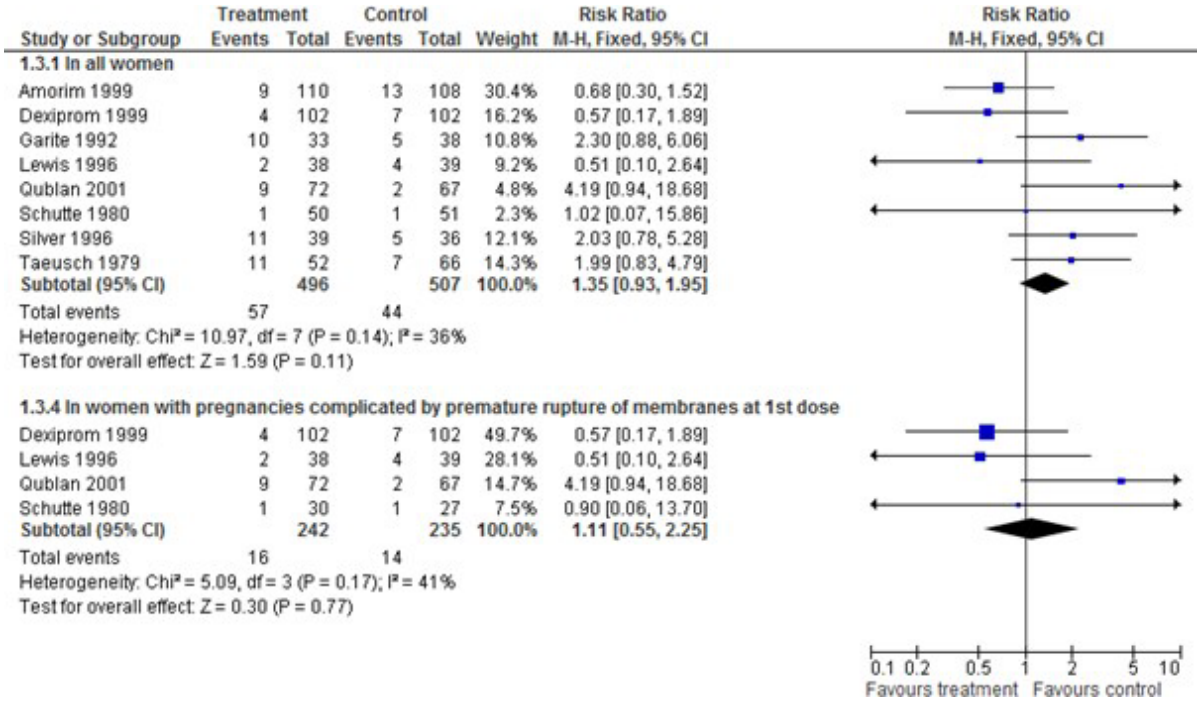


Figure 75: Side-effects of therapy in women

Figure 76: Puerperal sepsis



I.8.2 Repeat courses

This section was updated and replaced in 2022. Please see the NICE website for the updated guideline

I.9 Magnesium sulphate for neuroprotection

Figure 77: Stillbirth

Figure 78: Neonatal mortality: before discharge

Figure 79: Neonatal/paediatric mortality: between discharge and follow-up

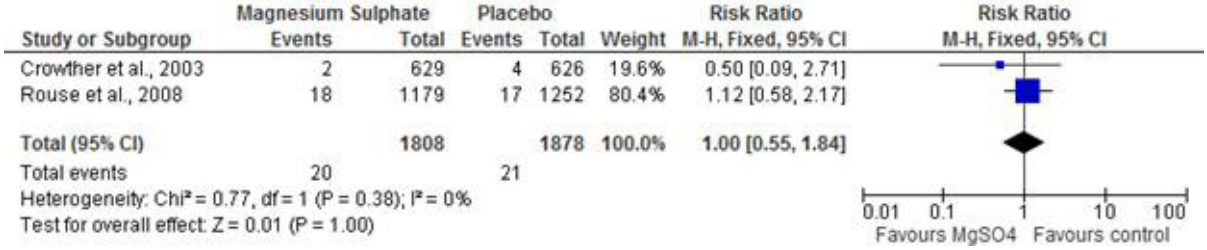


Figure 80: Total perinatal, neonatal and paediatric mortality

Figure 81: Findings on cranial ultrasound: grades III or IV intracranial haemorrhage

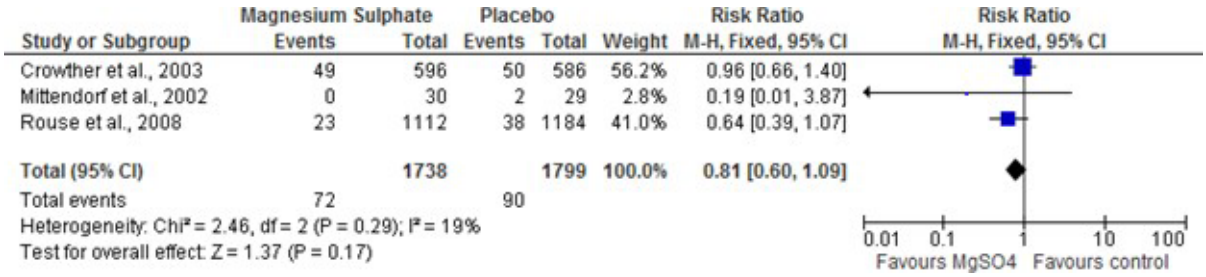


Figure 82: Findings on cranial ultrasound: periventricular leukomalacia

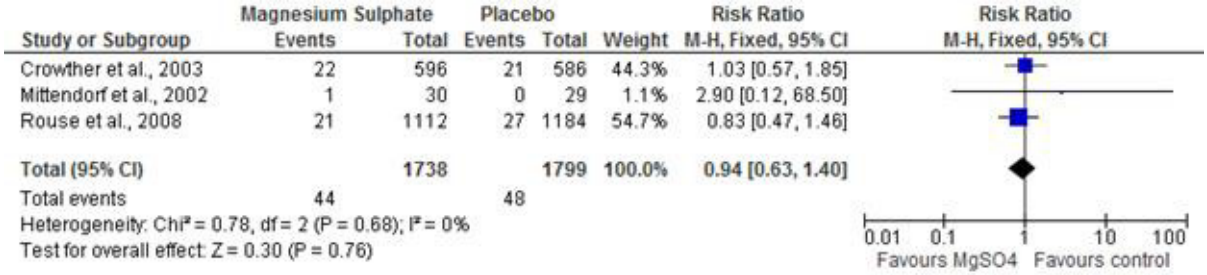


Figure 83: Cerebral palsy: any

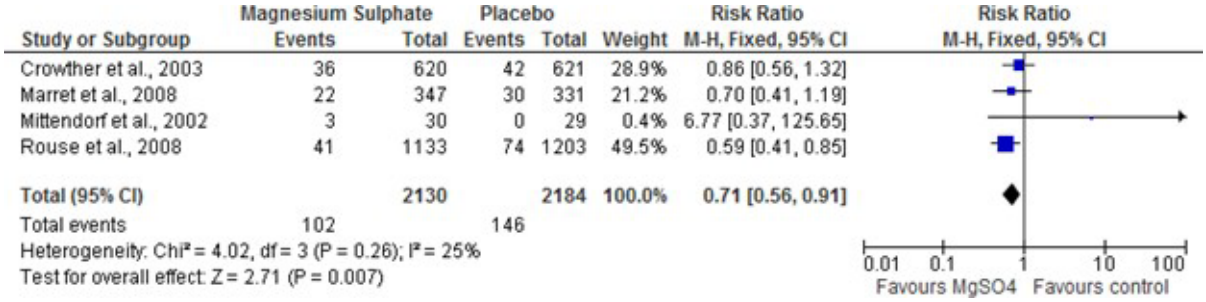


Figure 84: Cerebral palsy: moderate or severe (at 2 years)

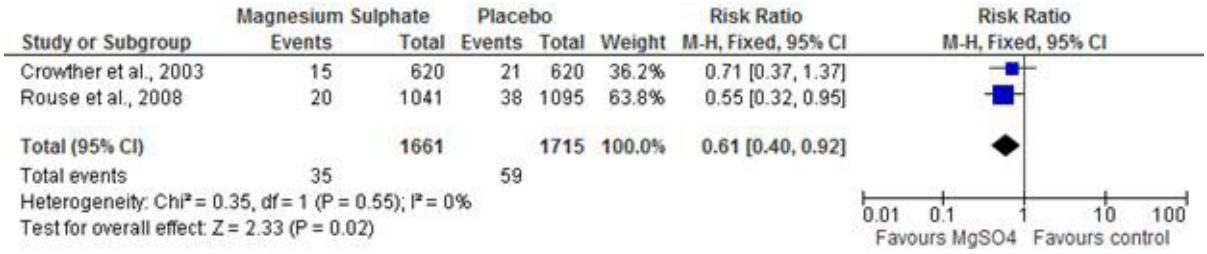


Figure 85: Maternal death

Figure 86: Maternal adverse effects: any

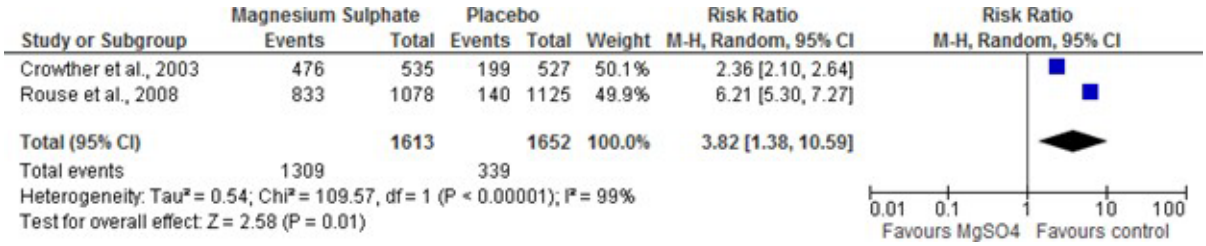


Figure 87: Maternal adverse effects: leading to stopping of infusion

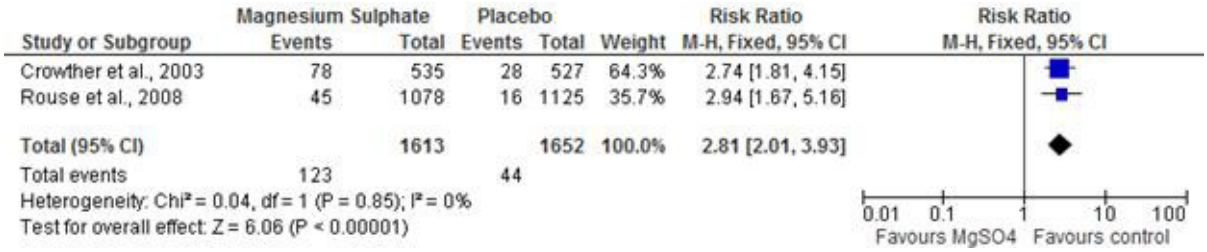
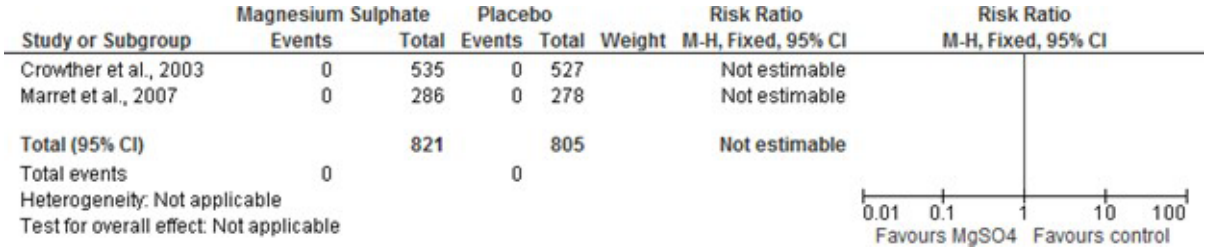


Figure 88: Maternal adverse effects: cardiac or respiratory arrest



I.10 Tocolysis

Figure 89: Neonatal mortality

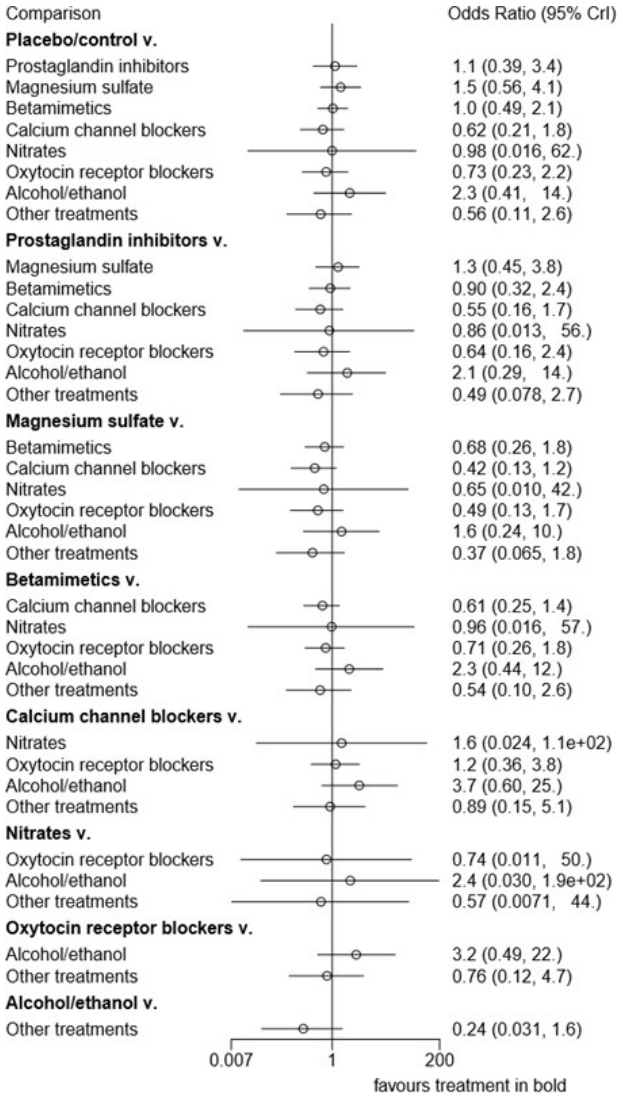


Figure 90: Perinatal mortality

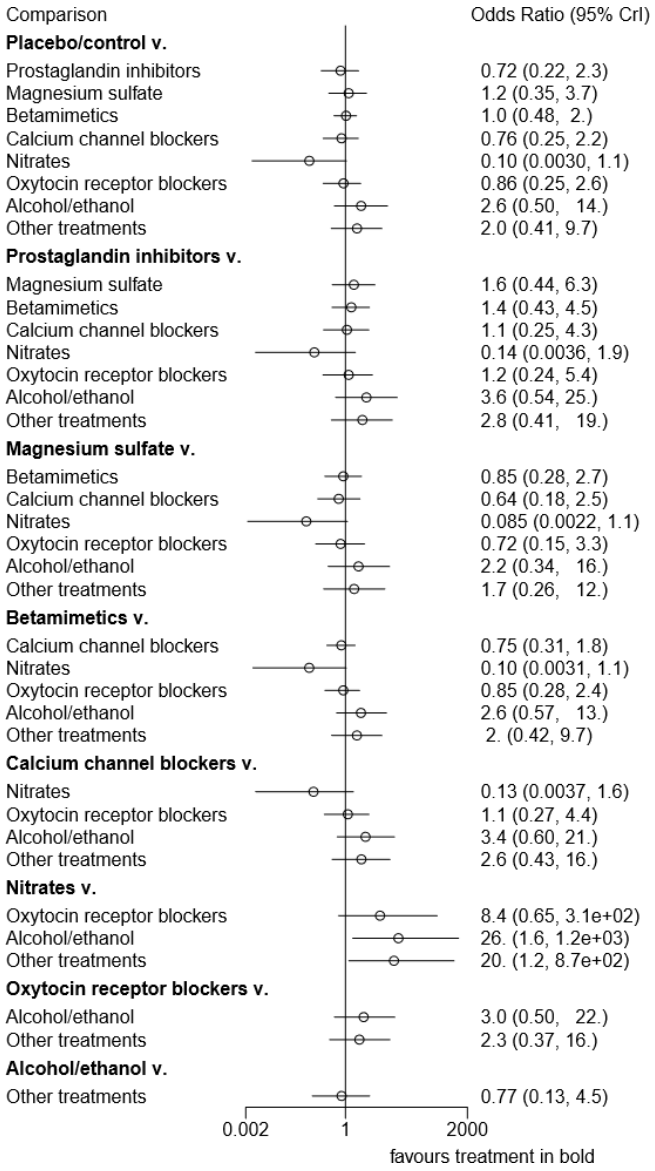


Figure 91: Respiratory distress syndrome

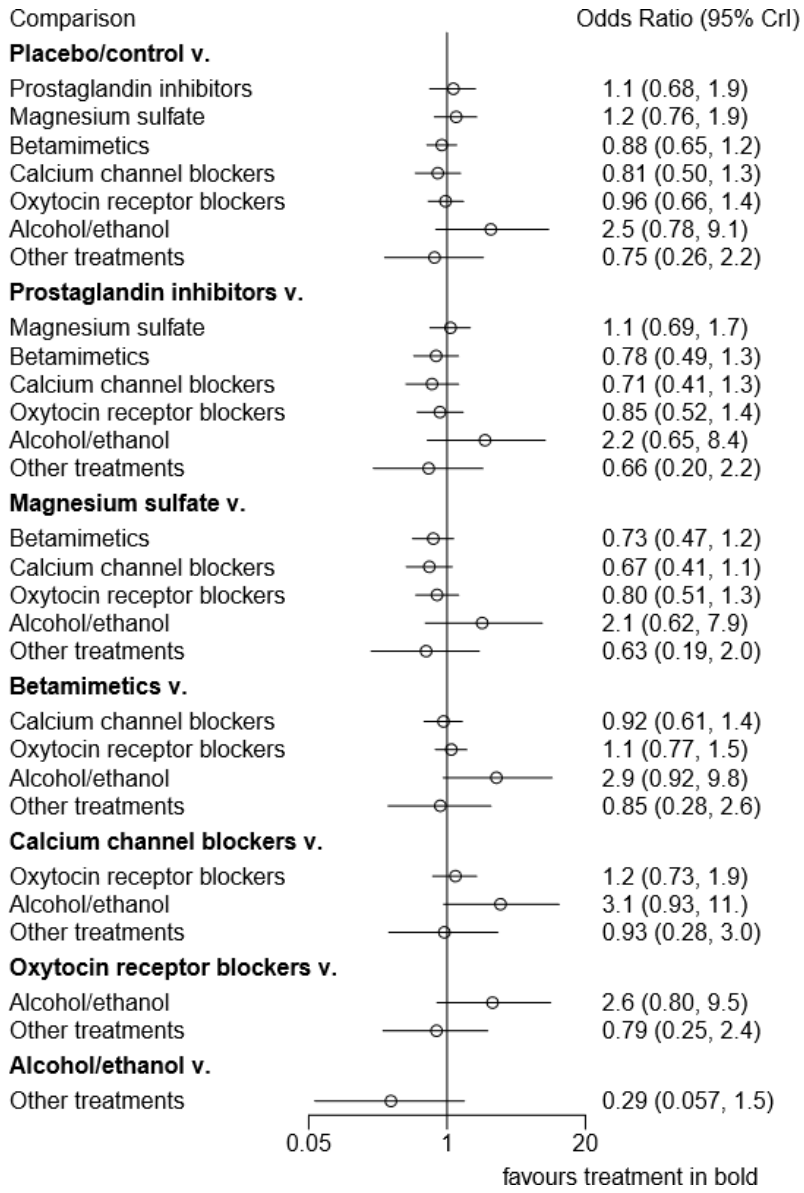


Figure 92: Intraventricular haemorrhage

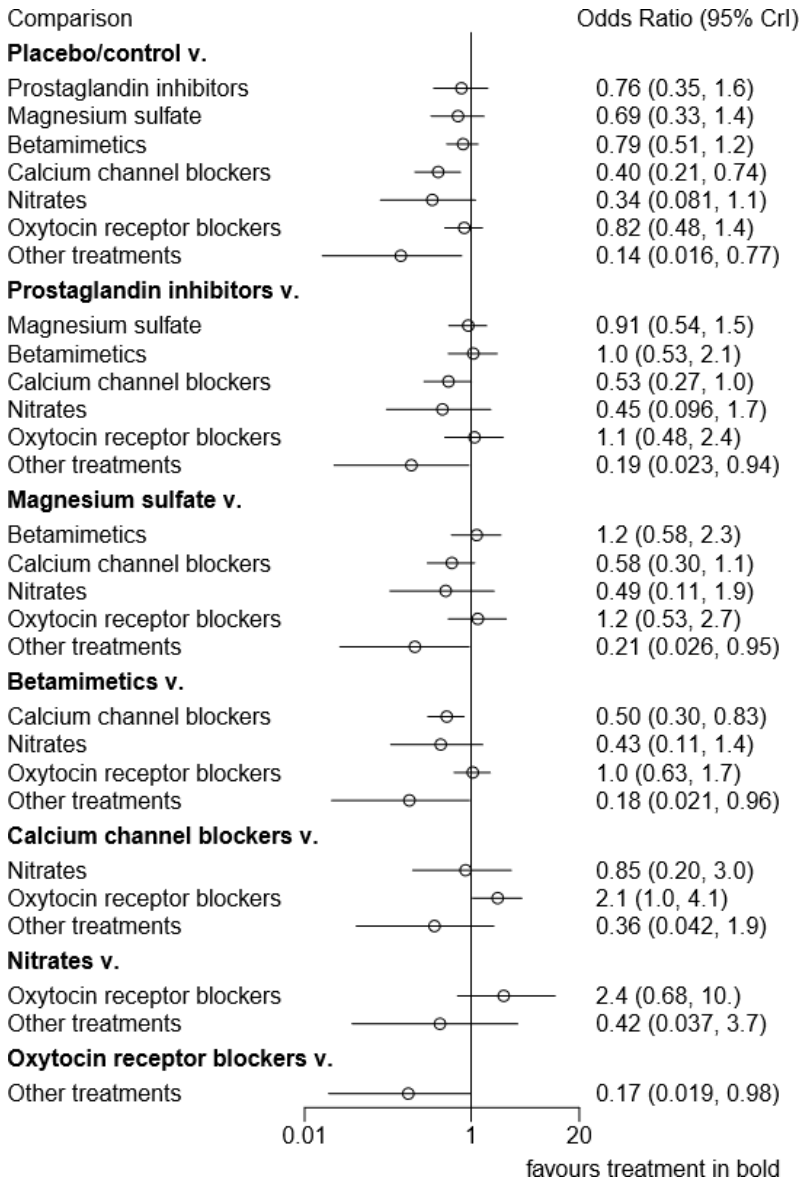


Figure 93: Mothers with adverse events requiring cessation of treatment

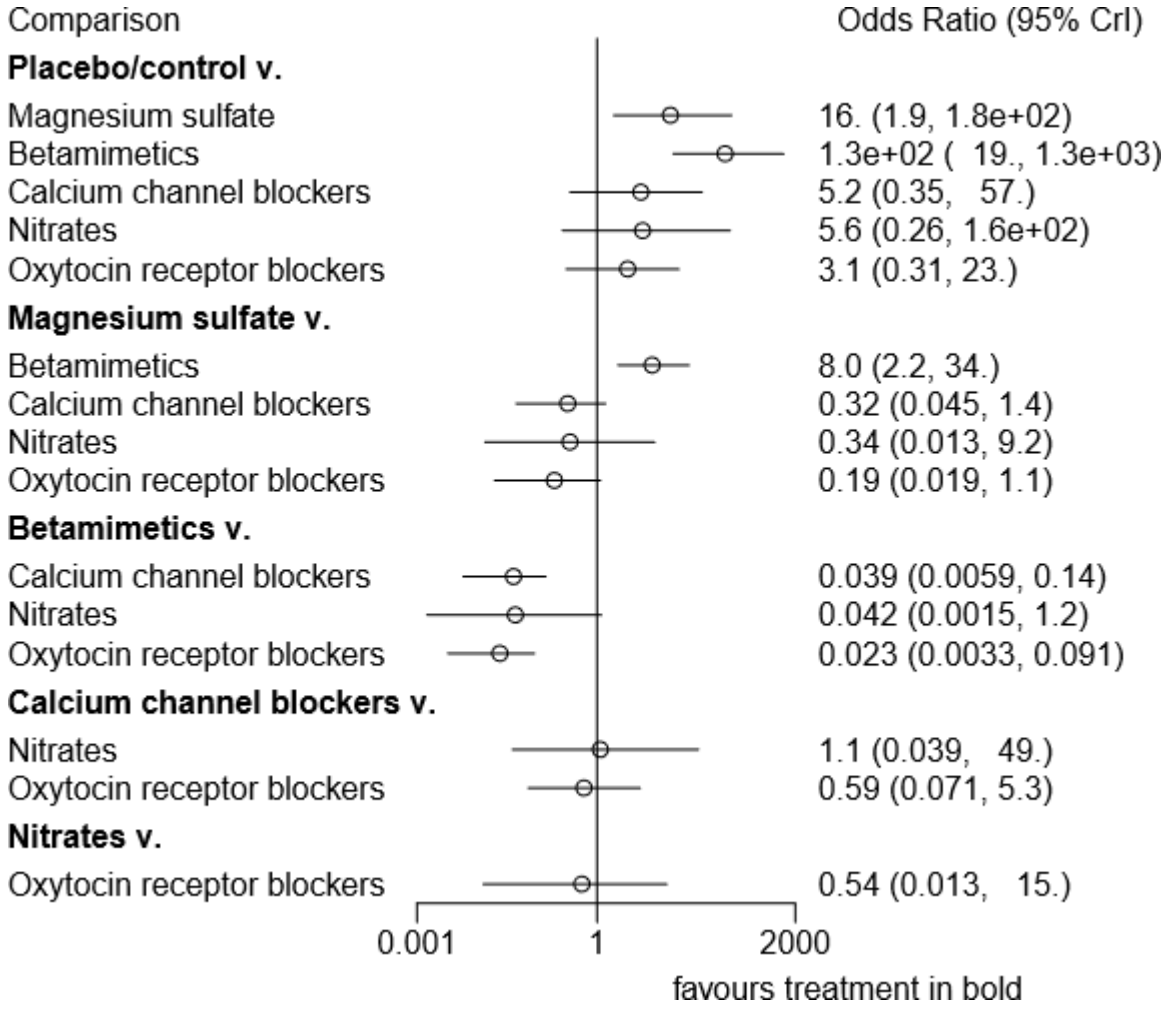


Figure 94: Delay of birth by at least 48 hours

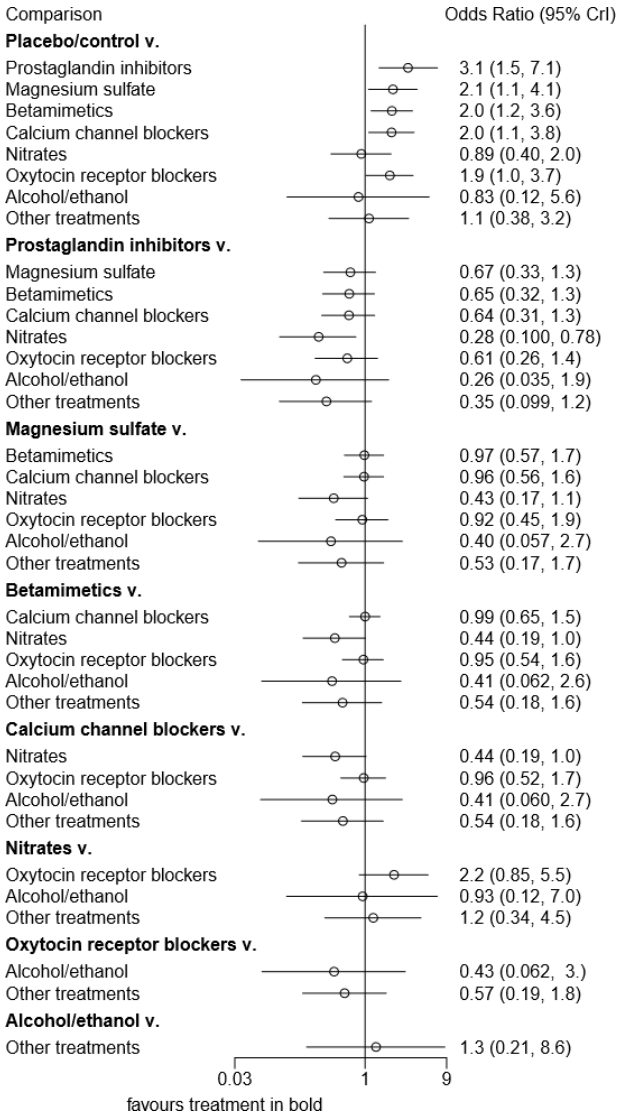


Figure 95: Neonatal sepsis

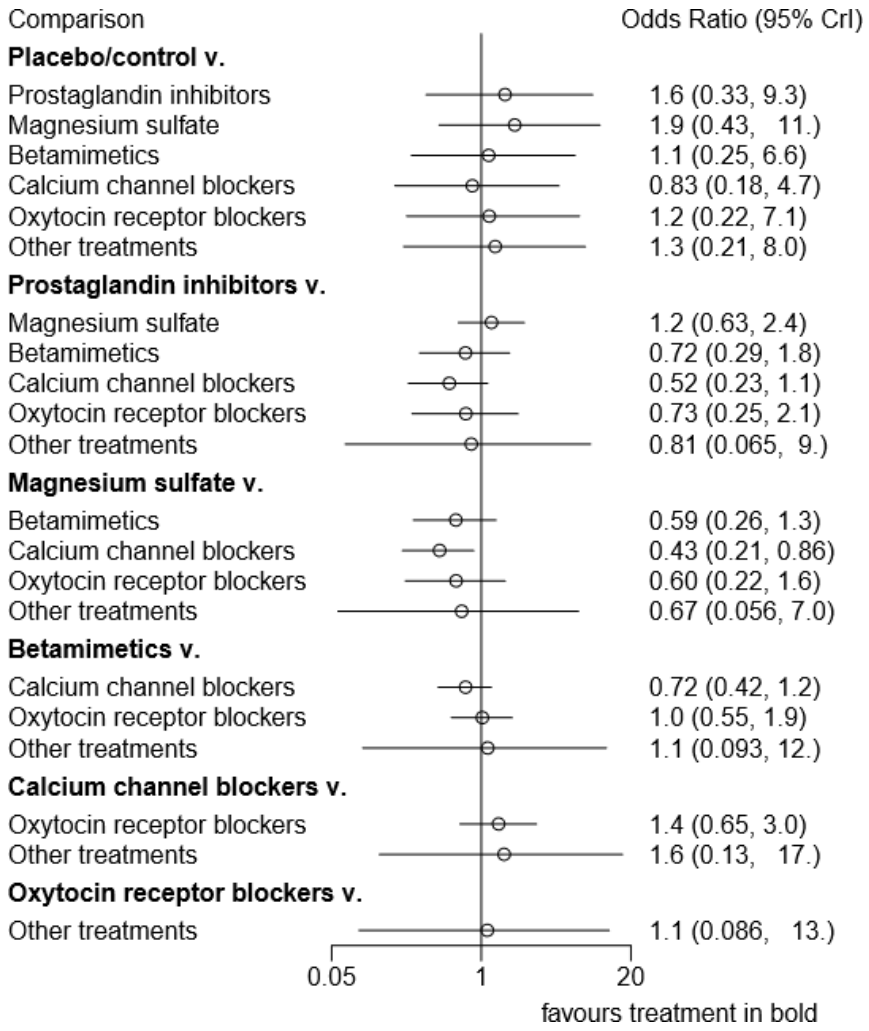
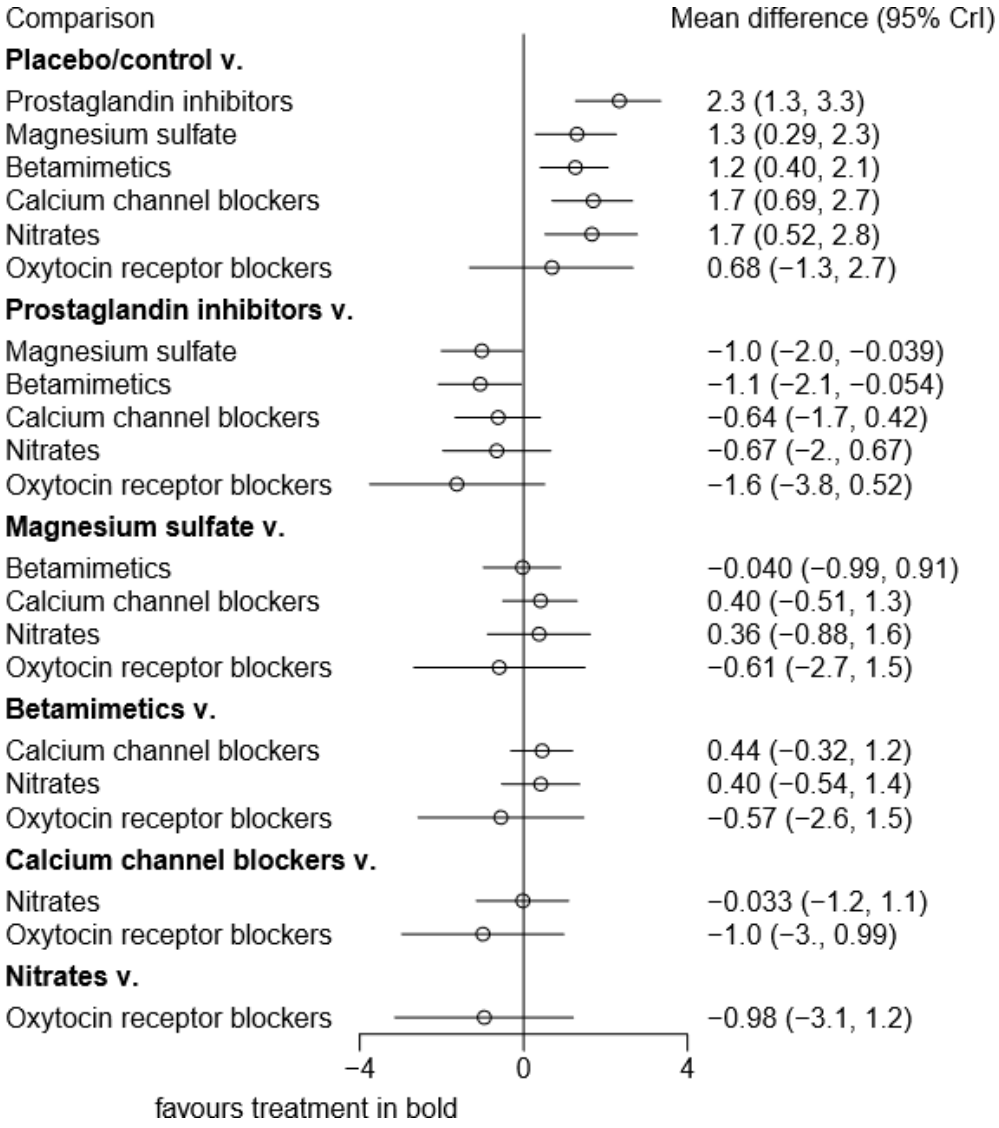


Figure 96: Gestational age at birth



I.11 Fetal monitoring

I.11.1 EFM versus IA

No forest plots were generated for this review question

I.11.2 Use of FSE

No forest plots were generated for this review question

I.11.3 CTG interpretation

No forest plots were generated for this review question

I.11.4 Blood sampling

No forest plots were generated for this review question

I.12 Mode of birth

I.12.1 Planned immediate caesarean section versus planned vaginal delivery in singletons

I.12.1.1 Neonatal outcome

Figure 97: Perinatal death

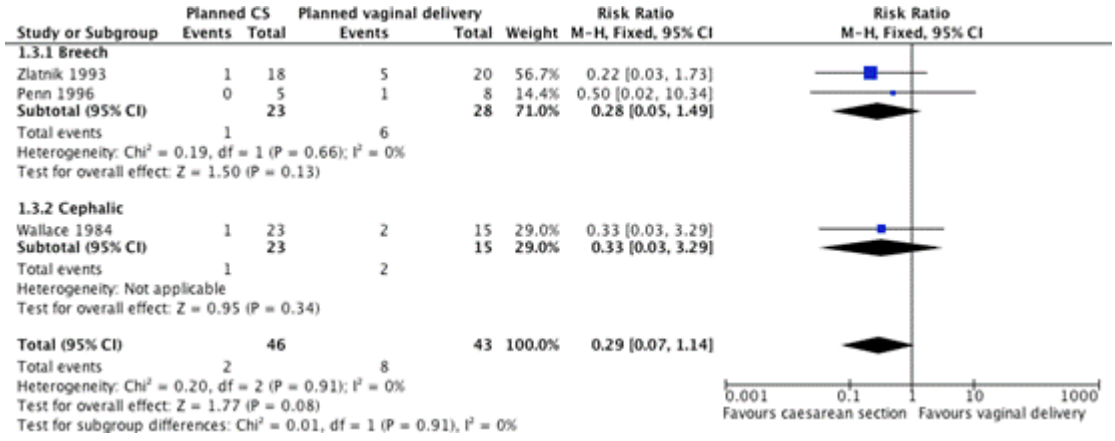


Figure 98: Intracranial pathology (outcome not pre-specified)

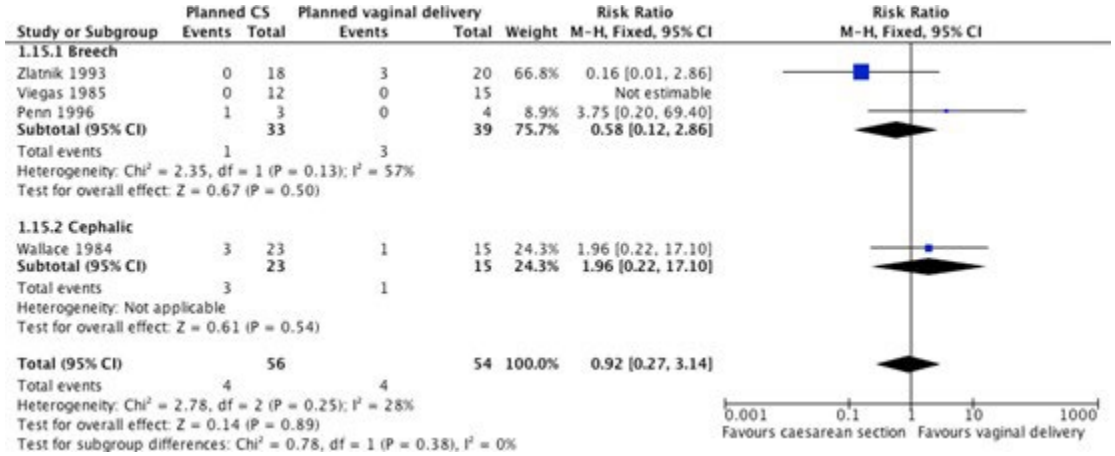


Figure 99: Hypoxic ischemic encephalopathy

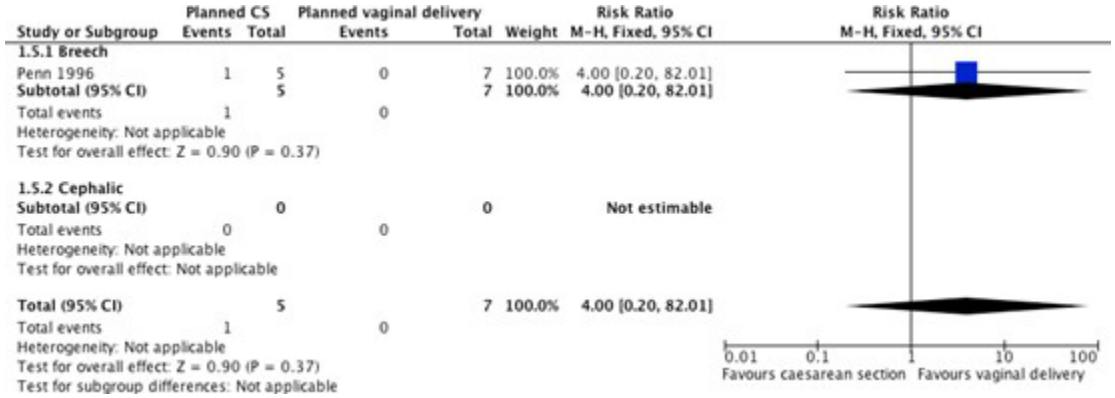
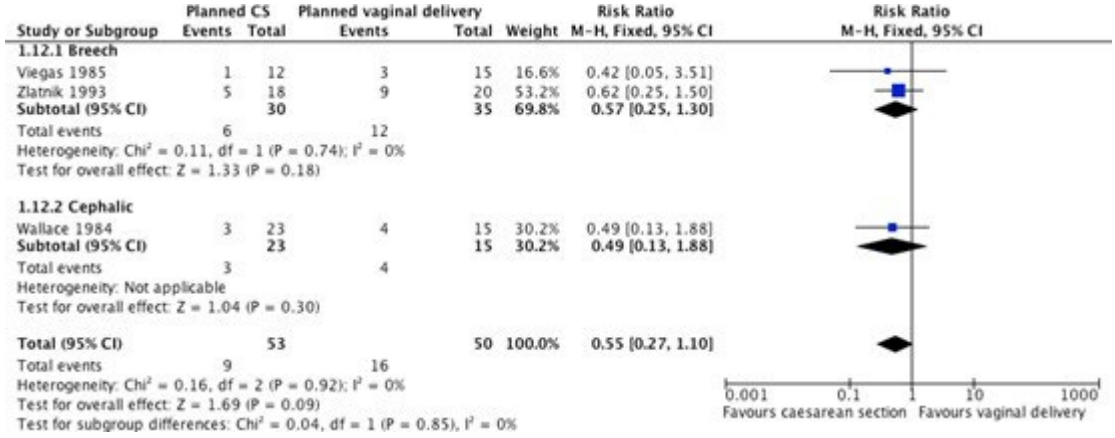


Figure 100: Respiratory distress syndrome



I.12.2 Immediate caesarean section versus planned vaginal delivery in singletons

I.12.2.1 Maternal outcomes

Figure 101: Postpartum haemorrhage

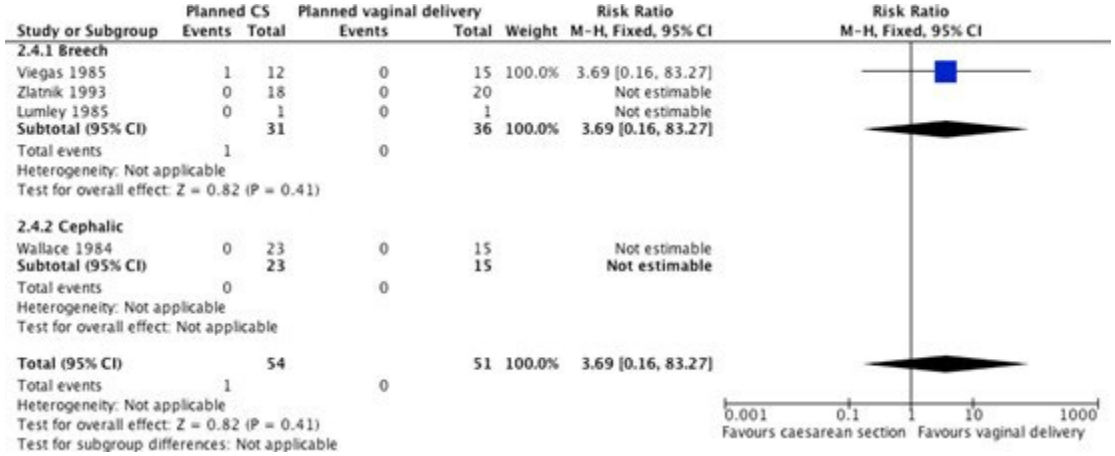


Figure 102: Maternal wound infection

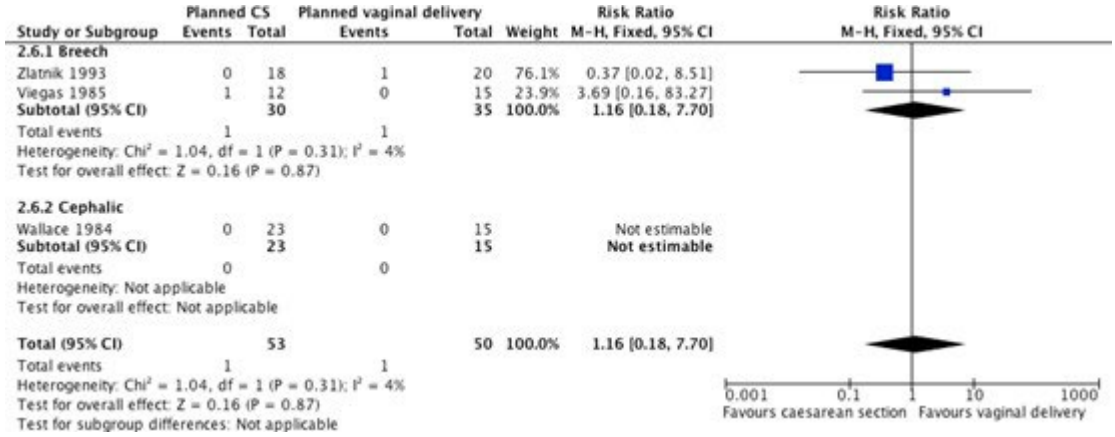
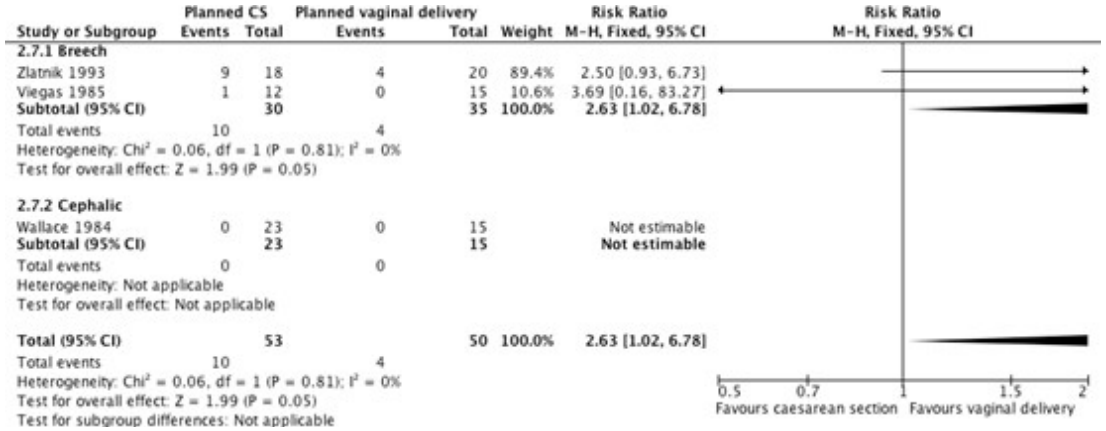


Figure 103: Other maternal infection



I.13 Timing of cord clamping

I.13.1 More placental transfusion (delayed clamping) versus less placental transfusion (early clamping)

Figure 104: Infant death

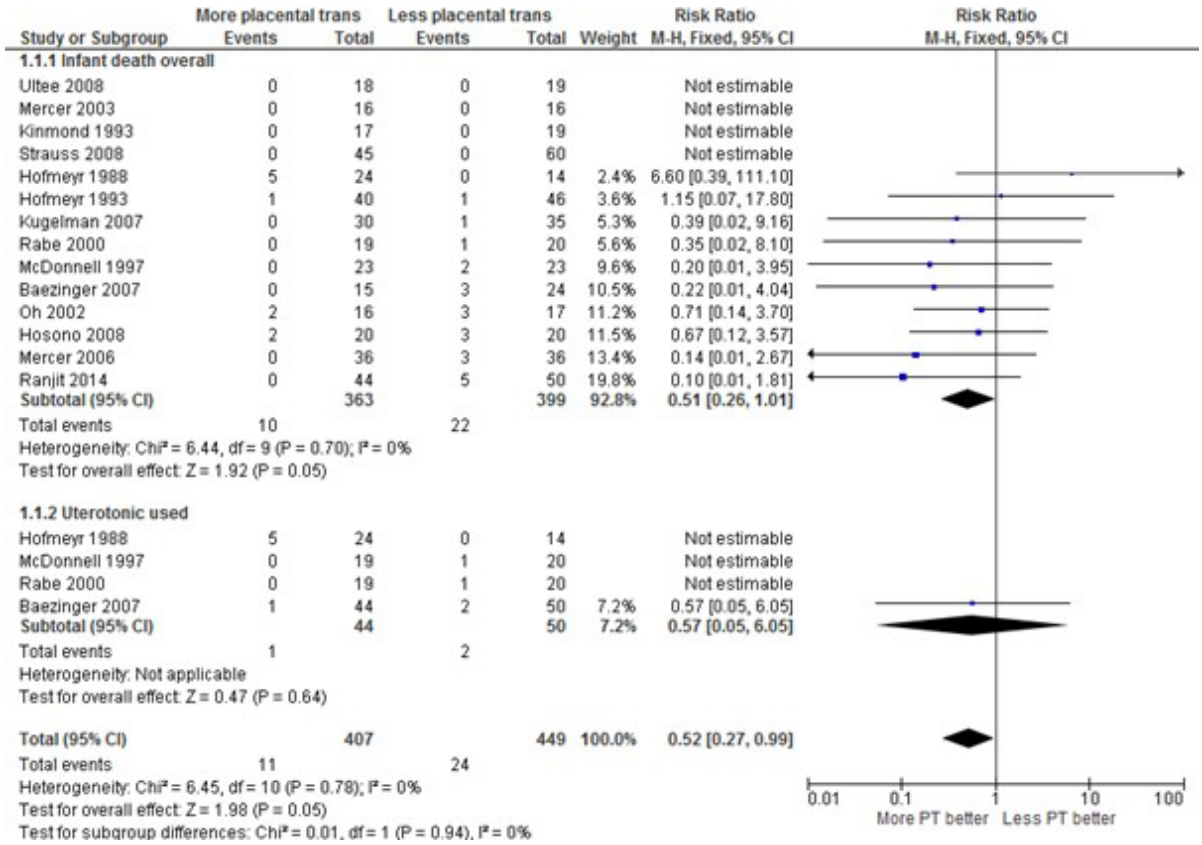


Figure 105: Intraventricular haemorrhage

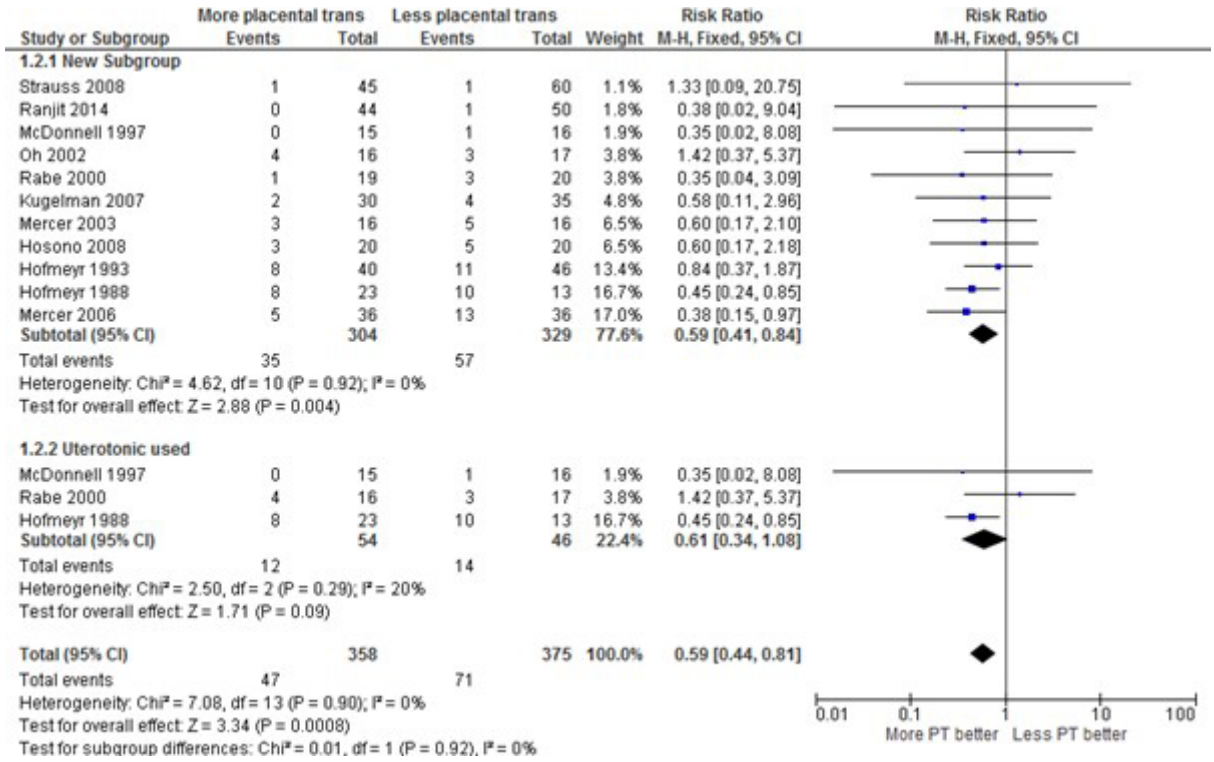


Figure 106: Severe intraventricular haemorrhage

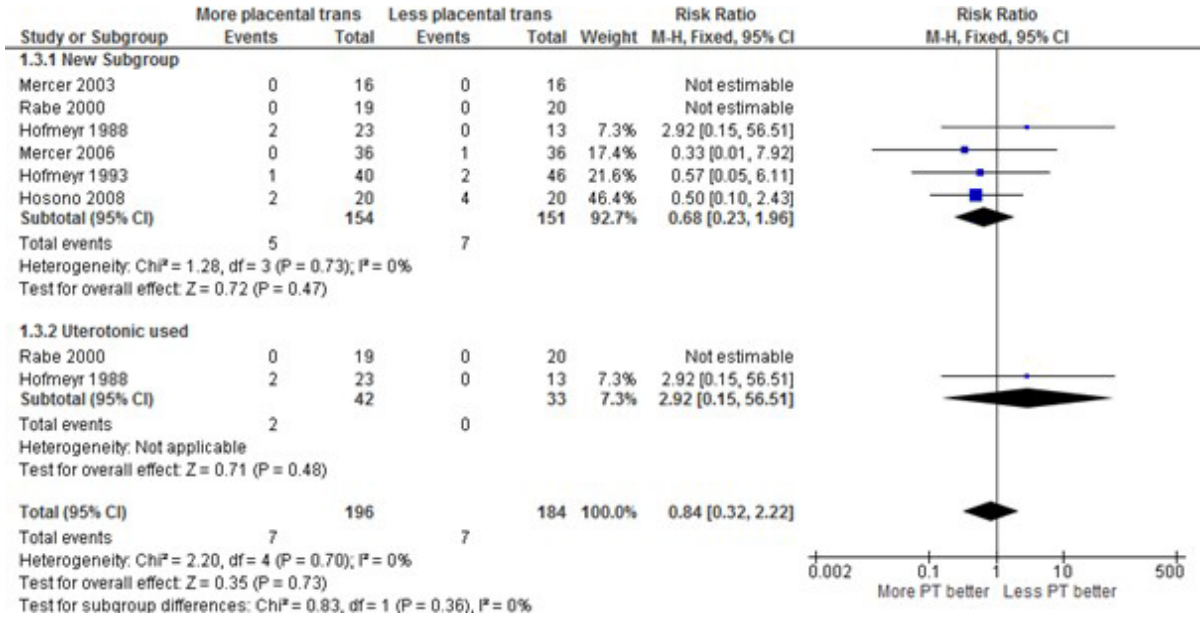


Figure 107: Ventilated for respiratory distress syndrome

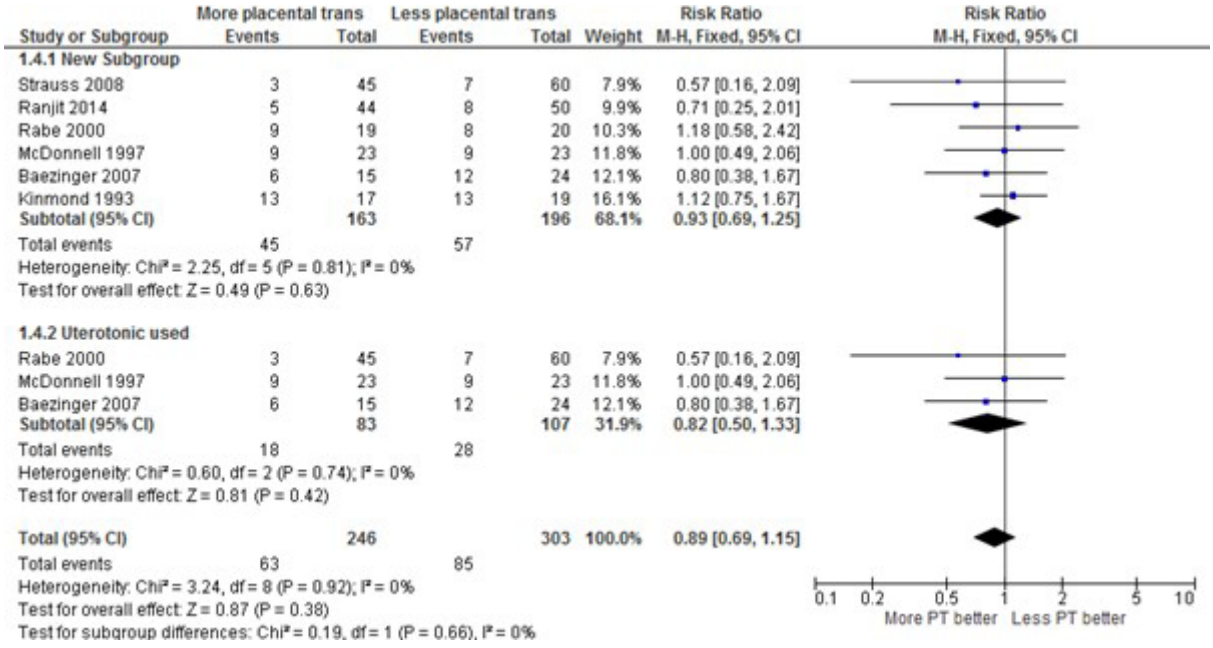


Figure 108: Hyperbilirubinemia

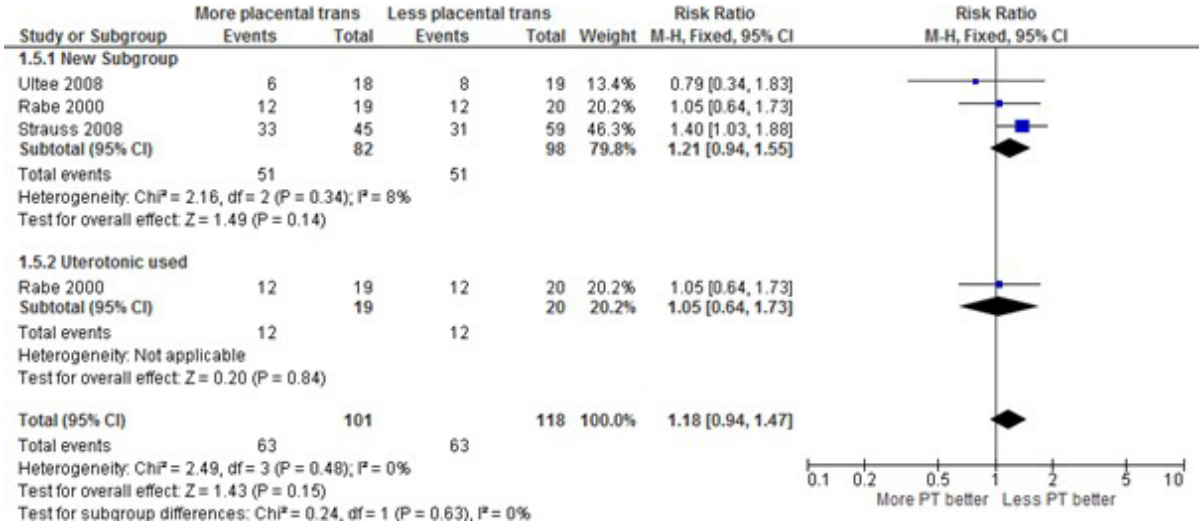


Figure 109: Transfused for anaemia

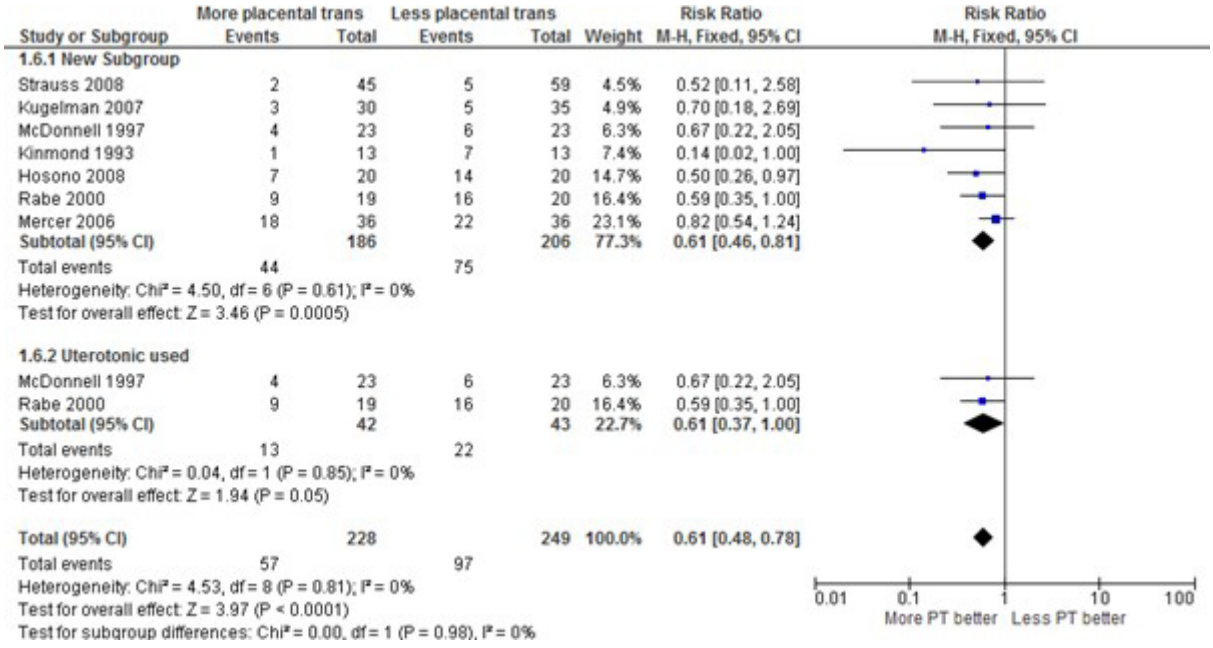


Figure 110: Apgar score at 5th minute < 8

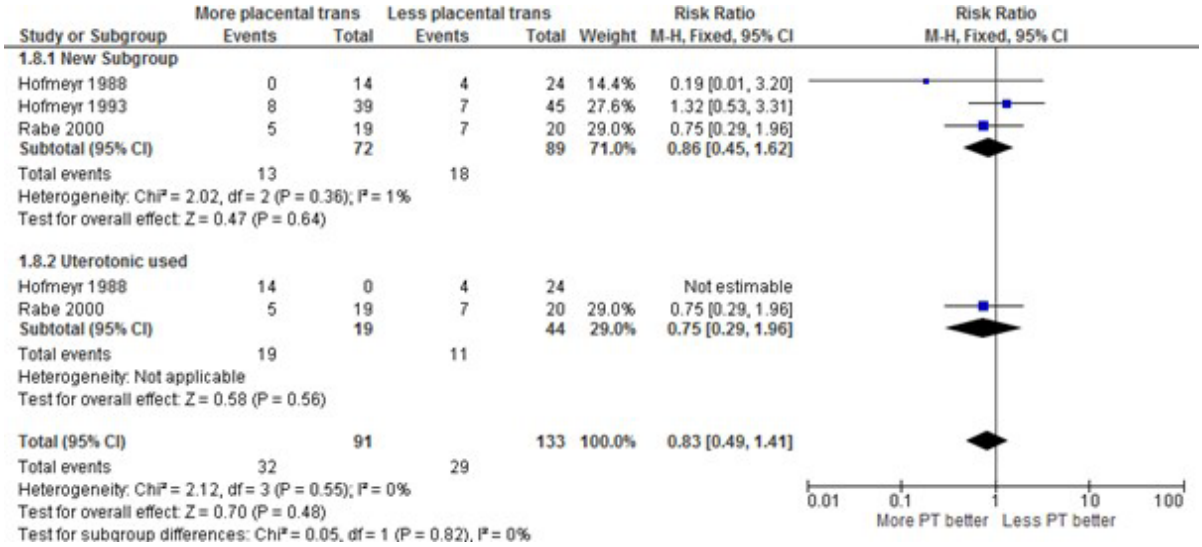


Figure 111: Haematocrit at 4 hours of life (%)

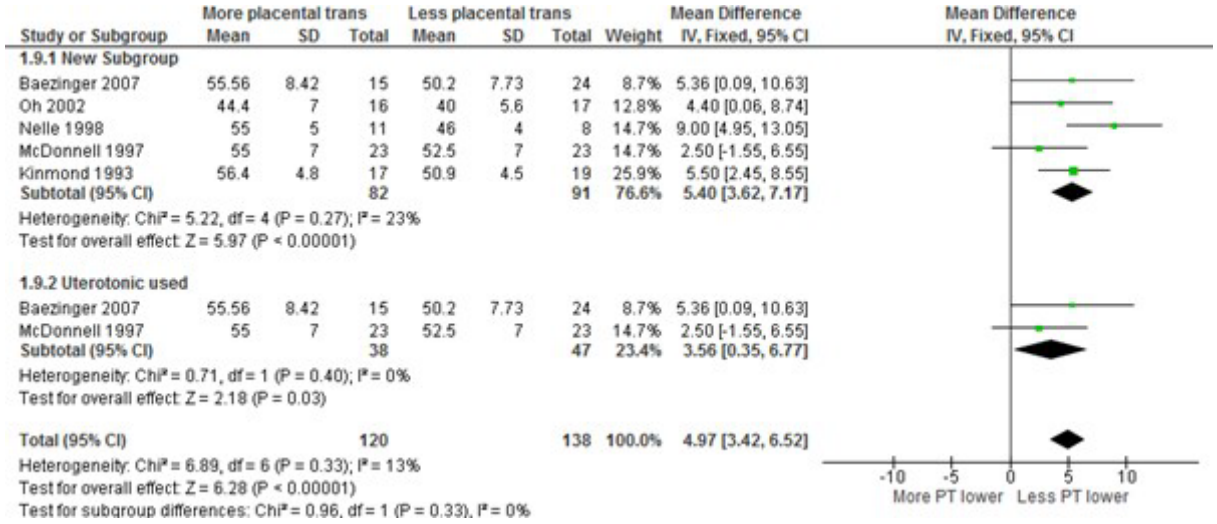
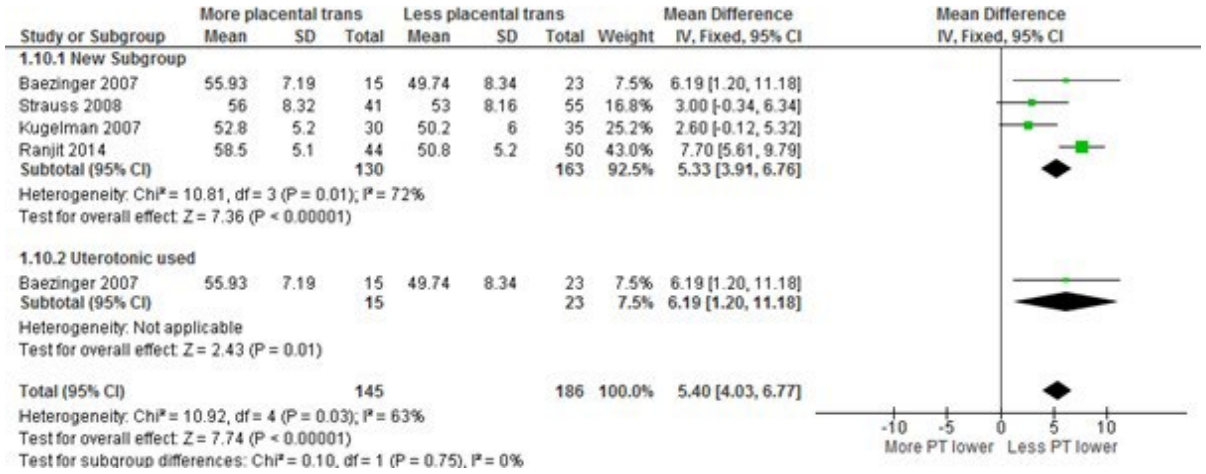


Figure 112: Haematocrit at 24 hours after birth (%)



I.13.2 More placental transfusion versus less placental transfusion: subgroup analysis by strategy for more placental transfusion

Figure 113: Infant death

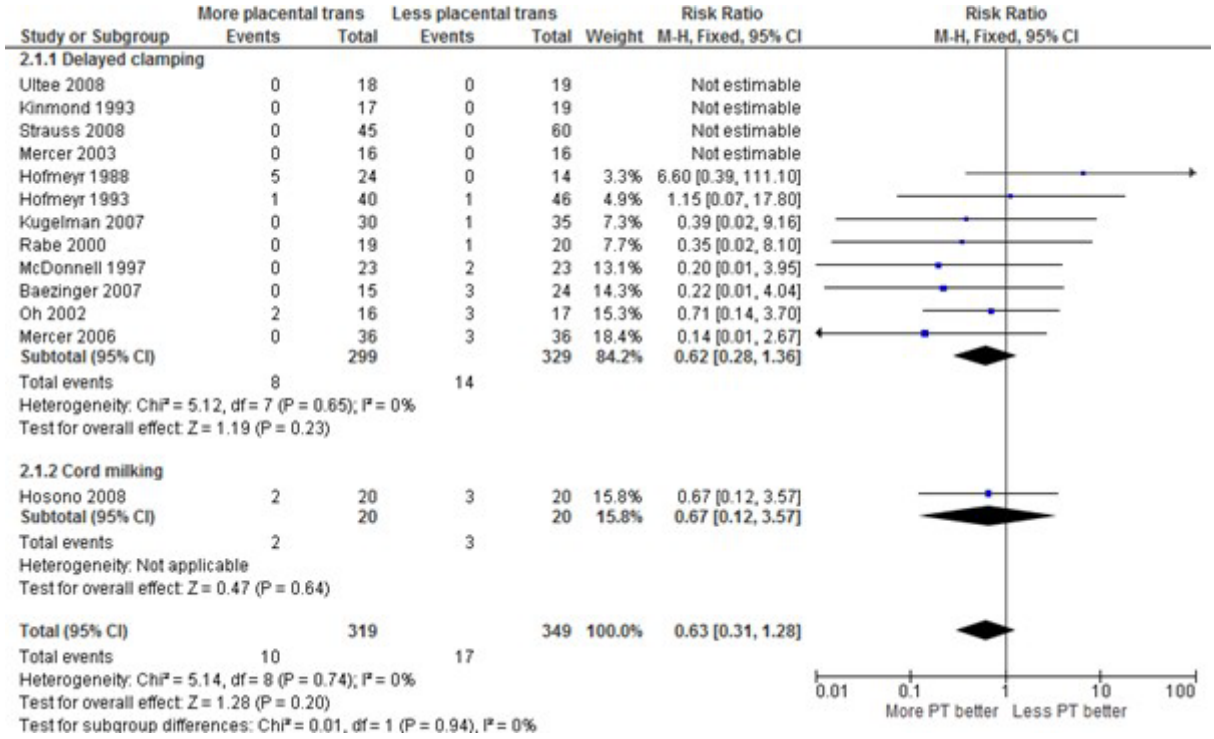


Figure 114: Severe intraventricular haemorrhage

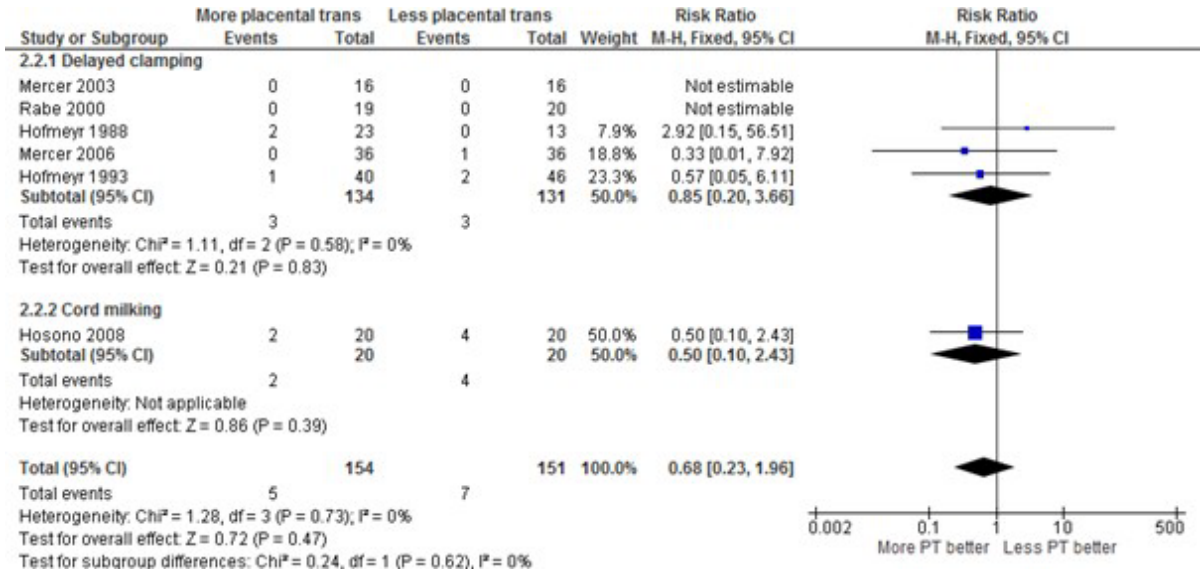
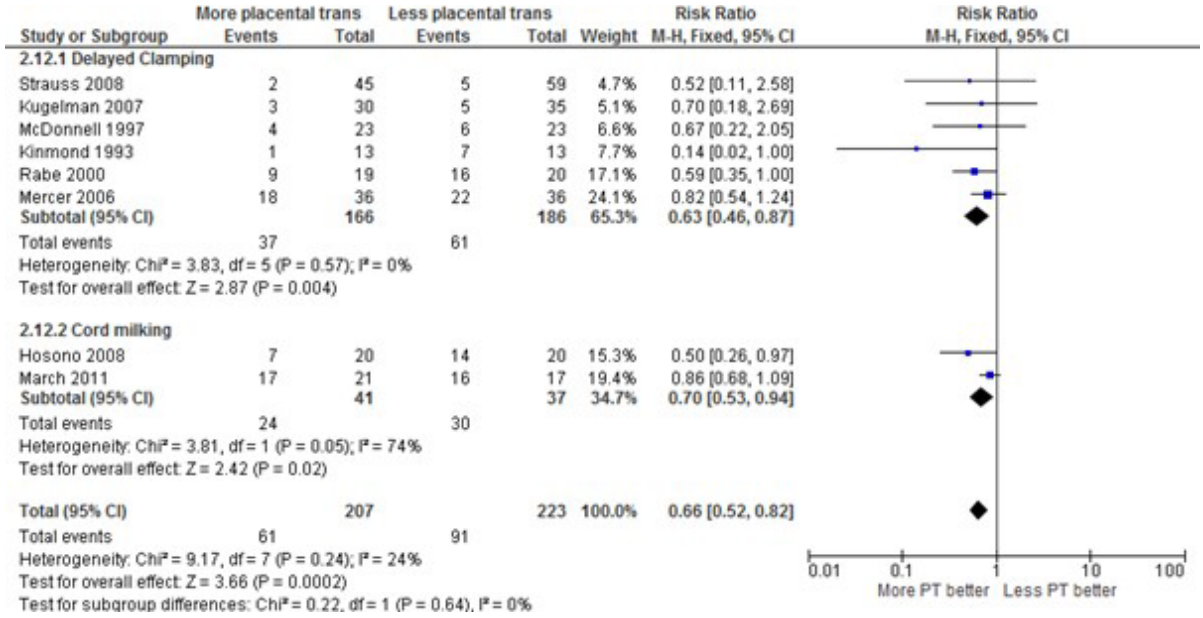


Figure 115: Transfused for anaemia



Appendix J: Network meta-analysis of tocolytics

J.1 Summary

Tocolytics are given to women in preterm labour to delay birth and therefore improve outcomes for the newborn. Whilst the treatment is given to the mother, the aim is to improve outcomes for the infant.

Network meta-analyses (NMA) of outcomes considered important to assess efficacy and safety were conducted. Eight outcomes were suitable for NMA:

1. IVH (infant)
2. RDS (infant)
3. Neonatal mortality (infant)
4. Neonatal sepsis (infant)
5. Perinatal mortality (infant)
6. Delay of birth by at least 48 hours (mother)
7. Termination of treatment due to adverse events (mother)
8. Estimated gestational age (EGA) at delivery (mother)

The first 7 outcomes are reported as the number of observed events out of the total number of infants or mothers, whilst EGA is reported as a continuous outcome (mean EGA) with a standard deviation. 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.

A total of 35 treatments (including Placebo and combinations of treatments) were evaluated in relevant trials. These treatments were classified into 9 classes (Table 1).

A NMA class model (Kew 2014) was used to estimate the relative effects of each treatment class compared to Placebo/control. 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.

A binomial / logit model was used to model outcomes 1 to 7 and a normal model with identity link was used to model EGA (Dias 2011).

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. Treatments were assigned to classes according to Table 2.

J.2 Methods

In order to take all trial information into consideration, without ignoring part of the evidence and without introducing bias by breaking the rules of randomisation (for example, by “naively” combining data across treatment arms from all RCTs), Mixed Treatment Comparison meta-analytic techniques, also termed Network meta-analysis (NMA), were employed. NMA is a generalization 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 (Dias 2001; Lu 2004; Caldwell 2005). 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 a 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. NMA techniques strengthen inference concerning the relative effect of two treatments by including both direct and indirect comparisons between treatments, and, at the same time, allow simultaneous inference on all treatments while respecting randomisation (Lu 2004; Caldwell 2005). Simultaneous inference on the relative effects of all treatments is possible whenever treatments are part of a single "network of evidence", that is, every treatment is linked to at least one of the other treatments under assessment. The correlation between the random effects of multi-arm trials (i.e. those with more than 2 arms) in the network is taken into account in the analysis (Dias 2011).

A Bayesian framework is used to estimate all parameters, using Markov chain Monte Carlo simulation methods implemented in WinBUGS 1.4.3 (Lunn 2000; Lunn 2013). In order to test whether starting values have an impact on the results, three chains with different initial values were run simultaneously. Convergence was assessed by inspection of the Gelman–Rubin diagnostic plots and by examining the history plots. Pre-convergence iterations were discarded, and further iterations on all chains were run on which results are based.

Sample WinBUGS code is provided in Section J.6.

J.2.1 Baseline probability (IVH, RDS and neonatal mortality)

Please see Health Economic Appendix K for details on calculating baseline probabilities for IVH, RDS and neonatal mortality.

J.2.2 Relative effects model

Models allowing for within-class differences in treatment effects were considered with both fixed and random treatment effects. These were compared with models assuming no within-class variability (i.e. all treatments in a class have the same relative effect), allowing for fixed or random treatment effects. Goodness of fit was tested using the posterior mean of the residual deviance, which was compared to the number of data points in the model and by inspecting the fit of each data point. Models were compared using the deviance information criteria (DIC) (Spiegelhalter 2002). The model with the lowest DIC was chosen, with differences of 5 considered meaningful. When models had very similar DIC (differences less than 5), simpler models were preferred, provided the posterior mean of the residual deviance was still close to the number of data points.

J.2.3 NMA model for binary data (outcomes 1 to 7)

A logit model was used to obtain the log-odds ratios of each treatment relative to Placebo. For each arm k of a trial i, the number of events, r_{ik} , have a binomial likelihood

$$r_{ik} \sim \text{Binomial}\left(p_{ik}, n_{ik} \right)$$

where p_{ik} is the probability of an event and n_{ik} the total number of patients in arm k of trial i.

The parameters of interest are the probabilities of an event and these are modelled using a NMA model on the log-odds scale using a logit link such that

$$\text{logit}(p_{ik}) = \mu_i + \delta_{ik}$$

with μ_i being given non-informative normal priors, Normal(0,1000), and $\delta_{i1} = 0$ since there is no relative treatment effect estimated for arm 1 of each trial.

In a random effects (RE) model the trial-specific treatment effects of the treatment in arm k, relative to the treatment in arm 1, are drawn from a common random effects distribution, under the assumption of consistency:

$$\delta_{ik} \sim N(d_{t_{ik}} - d_{t_{i1}}, \tau^2)$$

where $d_{t_{ik}}$ represents the mean effect of the treatment in arm k in trial i, t_{ik}, relative to Placebo, and τ^2 represents the between-trial variability in treatment effects (heterogeneity). The between-trials standard deviation, τ, was given a Uniform(0,5) prior.

In the FE model we replace equation (2) with

$$\text{logit}(p_{ik}) = \mu_i + d_t - d_t$$

J.2.4 NMA model for continuous data (EGA)

For each arm k of a trial i, the observed mean EGA, y_{ik}, has a normal likelihood

$$y_{ik} \sim \text{Normal}(\theta_{ik}, s^2_{ik})$$

where θ_{ik} is the underlying (true) mean EGA and s_{ik} is the standard error of the mean EGA in arm k of trial i.

The mean EGA is modelled using a NMA model such that

$$\theta_{ik} = \mu_i + \delta_{ik}$$

with μ_i being given non-informative normal priors, Normal(0,1000), and $\delta_{i1} = 0$, since there is no relative treatment effect estimated for arm 1 of each trial.

In a random effects (RE) model the trial-specific treatment effects of the treatment in arm k, relative to the treatment in arm 1, are drawn from a common random effects distribution, under the assumption of consistency (equation (3)). The between-trials standard deviation was given a Uniform(0,20) prior.

In the FE model we replace equation (5) with

$$\theta_{ik} = \mu_i + d_t - d_t$$

For studies not reporting the standard error, this was calculated using imputed standard deviations (SD). For each treatment for which a SD was not reported, it was imputed based on the median SD for that treatment reported in other studies. When there were fewer than 2 other studies reporting SD for a given treatment, the SD was imputed based on the median of reported SDs for that class. A sensitivity analysis imputing the upper quartile instead of the median was carried out.

J.2.5 Class model

Due to the sparseness of the network, with most comparisons being informed by only a few trials, a class model was used to borrow strength within treatment classes.

Two models for class were explored: an **exchangeable class effects** model, where the pooled relative treatment effects were assumed exchangeable within class

$$d_{1,k} \sim N(m_{Dk}, \tau^2_D)$$

with D_k indicating the class to which treatment k belongs to; and a **fixed class effects** model,

where the pooled relative treatment effects are assumed equal for all treatments in a class $d_{1,k} = m_{D_k}$. Magnesium sulphate belongs to a class formed only of itself (Class 3), so its relative treatment effect was assumed to be equal to its class effect in both models.

Both class models were considered with fixed or random treatment effects. The within-class mean treatment effects were given vague priors $m_j \sim N(0,100^2)$ and the within-class standard deviations were assumed equal for all classes (due to insufficient data) and given Uniform(0,2) priors.

J.2.6 Consistency

Consistency was assessed by checking the agreement of direct and indirect evidence using a node-split model (Dias 2009) fitted in R (Anonymous 2010) through the GeMTC package (van Valkenhoef 2012). 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.

J.3 Results

J.3.1 Baseline models (IVH, RDS, neonatal mortality)

Convergence was satisfactory by at least 20,000 iterations in all cases. Models were then run for a further 50,000 iterations on three separate chains, and all results are based on this further sample.

Results from these models are used in the relative effects model to generate a baseline $A \sim \text{Normal}(m, sd^2)$ on the log-odds scale on which relative effects were added at each iteration, to deliver the posterior summaries on the absolute probability scale for each treatment (Dias 2011a; Dias 2011b).

The estimated probabilities of events were very imprecise and there was large between-study heterogeneity in the log-odds of an event. This suggests that the included studies are very different in their baseline event rates and that they are perhaps not all representative of the UK population.

J.3.2 Imputing standard deviations (EGA)

51 studies were used in the NMA for EGA. 5 studies (Merkatz 1980, Leveno 1986, Larsen 1986, Rasanen 1995, Holleboom 1996) did not report the standard deviation (SD).

19 treatments were included in the network. No treatments in Class 8 (Alcohol/ethanol) were compared in trials reporting this outcome.

Five studies did not report SD for EGA (Merkatz 1980, Leveno 1986, Larsen 1986, Rasanen 1995, Holleboom 1996). This meant that the SD had to be imputed for 4 treatments: Placebo, Indomethacin, Sulindac and Ritodrine.

Placebo: 11 studies comparing this treatment to other treatments reported the SD, whilst 3 did not. The range of reported SD was 0.5 to 6.6 (Figure 133).

Indomethacin: 10 studies comparing this treatment to other treatments reported the SD, whilst 1 did not. The range of reported SD was 0.7 to 5.6 (Figure 133).

Sulindac: only 1 study comparing this treatment to other treatments reported the SD, whilst one other did not. The reported SD for other treatments of the same class (Class 2) were used as the basis for imputation. The range of reported SD for this class was 0.5 to 5.6 (Figure 133).

Ritodrine: 13 studies comparing this treatment to other treatments reported the SD, whilst 4 did not. The range of reported SD was 1.7 to 4.7 (Figure 133).

Imputed values for the main analysis were based on the median SD (Table 4, Figure 133). A sensitivity analysis using the upper quartile of the reported SD was also carried out (Table 4).

Model comparison using the DIC showed the fixed class with random treatment effects model as the preferred model (**Error! Reference source not found.**). The model with fixed lass and treatment effects was not fitted as it was expected to have a very poor fit, given the results of the exchangeable class, fixed effects model. Node-split models compared direct and indirect evidence on 11 comparisons. Some evidence of inconsistency was found for comparisons of placebo and magnesium sulphate (p=0.01).

J.3.3 Sensitivity to imputed SD

When imputing the upper quartile of the reported SD, the fixed class with fixed treatment effects model was preferred, although there were some poorly fitting data points and there was evidence of inconsistency for comparisons of placebo and prostaglandin inhibitors (p=0.02) and placebo and betamimetics (p=0.49). Apart from increased uncertainty the main results were not affected.

Table 1: Class descriptions

	Classes
1	Placebo/control
2	Prostaglandin inhibitors
3	Magnesium sulfate
4	Betamimetics
5	Calcium channel blockers
6	Nitrates
7	Oxytocin receptor blockers
8	Alcohol/ethanol
9	Other treatments

Table 2: Treatments with class assignments

	Treatment	class
1	Placebo	1
2	No treatment	1
3	Bed rest	1
4	Celecoxib	2
5	Indomethacin	2
6	Ketorolac	2
7	Mefenic Acid	2
8	Nimeluside	2
9	Rofecoxib	2
10	Sulindac	2
11	Magnesium Sulfate	3
12	Beta-Mimetics	4

	Treatment	class
13	Fenoterol	4
14	Hexoprenaline	4
15	Isoxsuprine	4
16	Ritodrine	4
17	Salbutamol	4
18	Terbutaline	4
19	Nylidrin	4
20	Calcium-Channel Blocker	5
21	Nicardipine	5
22	Nifedipine	5
23	Nitric Oxide	6
24	Nitroglycerin	6
25	Atosiban	7
26	Barisiban 1.0	7
27	Barusiban 0.3	7
28	Barusiban 10	7
29	Barusiban 3.0	7
30	Alcohol	8
31	Ethanol	8
32	Beta-Mimetics + Mag	9
33	Alcohol + Indomethacin	9
34	Other Tocolytic(s)	9
35	Tocolysis	9

Treatment classes are defined in Table 1

Table 3: Posterior mean of the residual deviance ($\bar{D}_{res}$) DIC for all models

Outcome (number of data points)	Measures of model fit	Exchangeable class effects		Fixed class effects	
		RE	FE	RE	FE
IVH (61)	$\bar{D}_{res}$	65.7	68.6	66.1	69.2
	DIC	285.1	284.2	284.0	282.9
	between-study standard deviation	0.27 (0.01, 0.83)	-	0.27 (0.01, 0.81)	-
	within-class standard deviation	0.44 (0.02, 1.78)	0.43 (0.02, 1.77)	-	-
RDS (102)	$\bar{D}_{res}$	110.0	114.3	112.3	121.3
	DIC	506.5	505.8	506.9	507.6
	between-study standard deviation	0.20 (0.01, 0.50)	-	0.25 (0.02, 0.54)	-
	within-class standard deviation	0.30 (0.02, 0.87)	0.36 (0.04, 0.92)	-	-
Neonatal mortality (102)	$\bar{D}_{res}$	111.6	132.5	112.2	144.0
	DIC	429.1	437.4	429.2	443.3
	between-study standard deviation	0.79 (0.24, 1.42)	-	0.86 (0.39, 1.47)	-
	within-class standard deviation	0.79 (0.04, 1.90)	1.16 (0.14, 7.95)	-	-
Neonatal sepsis (39)	$\bar{D}_{res}$	42.8	45.4	44.0	47.0
	DIC	181.2	180.1	181.0	179.8
	between-study standard deviation	0.44 (0.02, 1.49)	-	0.41 (0.02, 1.41)	-
	within-class standard deviation	0.65 (0.03, 1.87)	0.60 (0.03, 1.84)	-	-
Perinatal mortality (88)	$\bar{D}_{res}$	*	*	95.6	115.1
	DIC	*	*	365.1	371.8
	between-study	*	*	0.79 (0.19, 1.47)	-

Outcome (number of data points)	Measures of model fit	Exchangeable class effects		Fixed class effects	
		RE	FE	RE	FE
	standard deviation				
	within-class standard deviation	*	*	-	-
Delay by 48hrs (132)	$\bar{D}_{res}$	130.7	301.0	130.7	NA
	DIC	727.9	862.6	727.2	NA
	between-study standard deviation	0.89 (0.68, 1.16)	-	0.89 (0.68, 1.14)	-
	within-class standard deviation	0.14 (0.01, 0.55)	0.29 (0.05, 0.61)	-	-
Termination due to AE (75)	$\bar{D}_{res}$	80.1	103.2	82.0	102.5
	DIC	297.7	308.7	298.5	306.7
	between-study standard deviation	1.34 (0.26, 2.68)	-	1.17 (0.18, 2.74)	-
	within-class standard deviation	0.36 (0.02, 1.60)	0.18 (0.01, 0.97)	-	-
EGA (101)	$\bar{D}_{res}$	100.3	352.7	100.0	NA
	DIC	191.0	418.4	190.4	NA
	between-study standard deviation	1.25 (0.96, 1.64)	-	1.25 (0.98, 1.62)	-
	within-class standard deviation	0.25 (0.01, 0.98)	1.53 (0.96, 2.67)	-	-

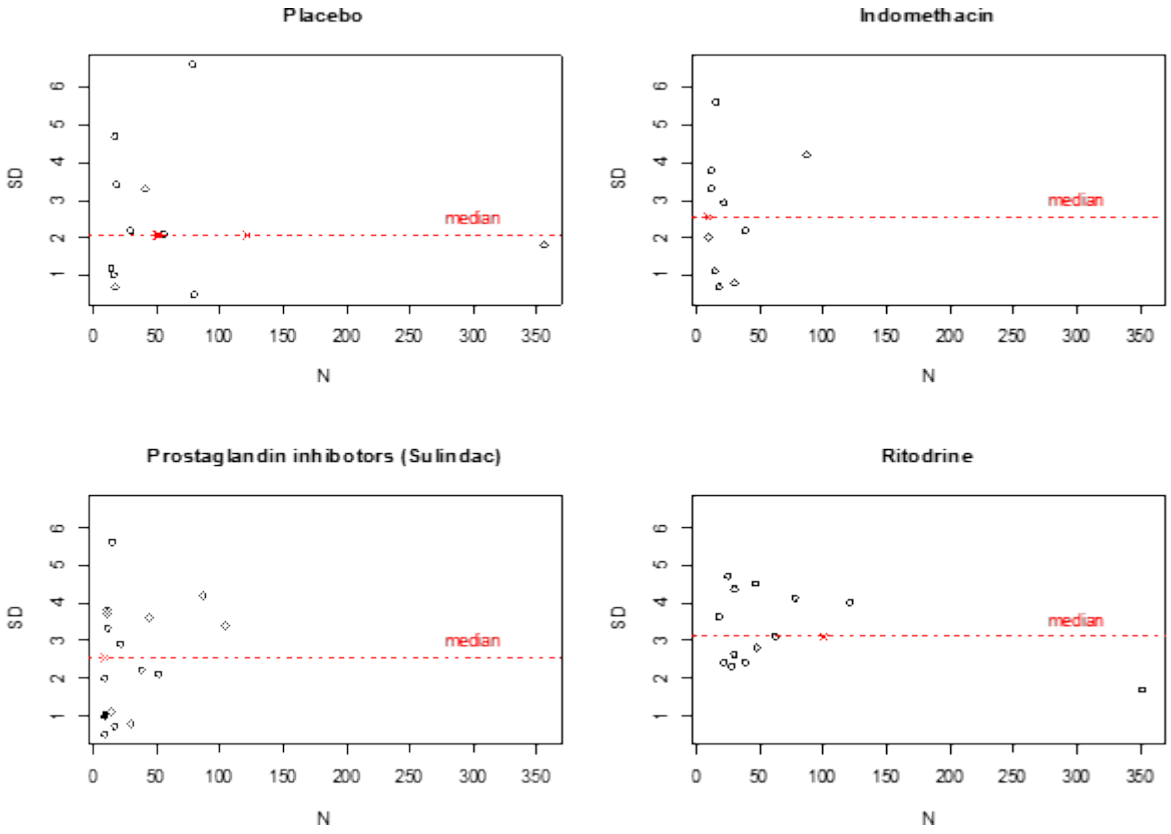
'NA' indicates the model was not fitted as it was expected to be a poor fit, and '' indicated that the model was not fitted because there was not enough evidence to estimate all the parameters. Shaded cells indicate the preferred model. The median and 95% Credible Intervals of the between-study deviation (heterogeneity) and within-class standard deviation are also presented, A '–' indicates that this value was fixed at zero in the model.*

Table 4: Vales used for the imputation of SD with these were not reported

Treatment	Median	Upper quartile
Placebo	2.1	3.35
Indomethacin	2.555	3.675
Sulindac	2.555	3.625
Ritodrine	3.1	4.1

J.4 Figures

Figure 133: Reported standard deviations (SD) in trials comparing the difference treatments, or treatments of the same class (open circles); SD in the only sulindac trial to report it (filled circle); imputed values (red crosses) and median SD, plotted against sample size



J.5 References

Kew 2014

Kew KM, Dias S, Cates CJ. Long-acting inhaled therapy (beta-agonists, anticholinergics and steroids) for COPD: a network meta-analysis. Cochrane Database of Systematic Reviews. 2014(3):Art No.: CD010844.

Dias 2011a

Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. 2011.

Lu 2004

Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 2004;23:3105-3124.

Caldwell 2005

Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005 2005;331:897-900.

Lunn 2000a

Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000;10:325-337.

Lunn 2000b

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS book. Boca Raton, FL: CRC Press; 2013.

Dias 2011b

Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011.

Spiegelhalter 2002

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society (B). 2002;64(4):583-616.

Dias 2010

Dias S, Welton NJ, Caldwell DM, Ades AE. Checking Consistency in Mixed Treatment Comparison Meta-analysis. Stat Med. 2010;29:932-944.

Anonymous 2010

R: A Language and Environment for Statistical Computing [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2010.

van Valkenhoef 2012

van Valkenhoef G, Lu G, De Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. Research Synthesis Methods. 2012;3:285-299.

J.6 Sample WINGBUGS code for binary outcome analyses

FIXED CLASS, FIXED TREATMENT EFFECTS

Tocolytics: outcome is IVH
Class model - treatments exchangeable within class, within-class variance is zero (fixed class effects)

==========

21 May 2014

Treatments (code, Class, Treat)

1	1	Placebo	
2	2	Indomethacin	
3	2	Ketorolac	
4	2	Rofecoxib	
5	3	Magnesium Sulfate	
6	4	Beta-Mimetics	
7	4	Ritodrine	
8	4	Salbutamol	
9	4	Terbutaline	
10	4	Nylidrin	(NOT TO BE USED FOR RANKING)
11	5	Nifedipine	
12	6	Nitric Oxide	
13	7	Atosiban	
14	8	Other Tocolytic(s)	(NOT TO BE USED FOR RANKING)

Class "Alcohol/ethanol" not compared
Class 8 not to be used for ranking

==========

```
# Binomial likelihood, logit link
# Fixed effects model
# class effects - zero within-class variance
model{                              # *** PROGRAM STARTS
for(i in 1:ns){                     # LOOP THROUGH STUDIES
    mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
    for (k in 1:na[i])  {           # LOOP THROUGH ARMS
        r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
        logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
# expected value of the numerators
        rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
        dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
            +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-
rhat[i,k])))
      }
# summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
     }
totresdev <- sum(resdev[])      # Total Residual Deviance
d[1]<-0    # treatment effect is zero for reference treatment
# treatment effects from Class - fixed class effects
for (k in 2:nt){  d[k] <- m[D[k]]   }
```

```
m[1] <- 0
for (k in 2:nc){ m[k] ~ dnorm(0, .0001) } # priors for mean class effect
# all pairwise ORs
for (c in 1:(nt-1)) {
    for (k in (c+1):nt)  {
        lor[c,k]<- d[k]-d[c]
        OR[c,k] <- exp(lor[c,k])
      }
  }
# select treatments to be used for ranking and economic analysis
for(k in 1:9){ dR[k] <- d[k] }
# not treatment 10
for(k in 11:13){ dR[k-1] <- d[k] }
# not treatment 14
# ranking on relative scale
for (k in 1:ntR) {
#    rk[k]<- (ntR+1)-rank(dR[],k)      # events are "good"
    rk[k]<- rank(dR[],k)              # events are "bad"
    best[k] <- equals(rk[k],1)        # rank=1 is best
#calculate probability that treat k is h-th best
    for (h in 1:nt) { prob[h,k] <- equals(rk[k],h) }
  }
# Provide estimates of treatment effects T[k] on the natural scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:ntR) { logit(T[k]) <- A + dR[k]  }
# all pairwise ORs for classes
for (c in 1:(nc-1)){
    for (k in (c+1):nc){
        lorClass[c,k] <- m[k] - m[c]
        ORClass[c,k] <- exp(m[k] - m[c])
    }
 }
# rank all classes except last
for (k in 1:nc-1)  {
#    rkClass[k]  <- (nc+1)-rank(m[],k)      # events are "good"
    rkClass[k]  <- rank(m[1:(nc-1)],k)            # events are "bad"
    bestClass[k]  <- equals(rkClass[k],1)  # rank=1 is best
# prob class k is h-th best, prob[1,k]=best[k]
    for (h in 1:nc-1) { probClass[h,k] <- equals(rkClass[k],h) }
    }
}                                  # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments; nc=number of classes; D=index of classes
# ntR = number of treat for ranking
list(ns=29, nt=14, nc=8, meanA=-2.814, precA=0.9861, ntR=12,
D=c(1, 2, 2, 2, 3, 4, 4, 4, 4, 4, 5, 6, 7, 8))
```

na[]	t[,1]	t[,2]	t[,3]	r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]	#Study	Year
3	1	5	9	3	1	2	19	16	19	#Cotton	1984
3	2	5	11	14	11	10	103	95	119	#Klauser	2012
3	7	13	13	5	7	4	56	61	58	#Goodwin	1996
2	1	2	NA	0	1	NA	20	19	NA	#Panter	1999
2	1	5	NA	4	4	NA	89	78	NA	#Cox	1990
2	1	7	NA	4	2	NA	55	56	NA	#Leveno	1986
2	1	7	NA	31	21	NA	391	380	NA	#CPLIG	1992
2	1	12	NA	1	2	NA	79	74	NA	#Smith	2007
2	1	13	NA	19	16	NA	246	243	NA	#Romero	2000
2	2	5	NA	4	4	NA	49	52	NA	#Morales	1993
2	2	5	NA	4	6	NA	14	18	NA	#Parilla	1997
2	2	7	NA	1	4	NA	47	50	NA	#Morales	1989

2	2	7	NA	3	2	NA	25	20	NA	#Besinger	1991
2	2	10	NA	2	0	NA	30	30	NA	#Kurki	1991
2	3	5	NA	1	0	NA	45	43	NA	#Schorr	1998
2	4	5	NA	6	7	NA	92	102	NA	#McWhorter	2004
2	5	11	NA	3	2	NA	106	110	NA	#Lyell	2007
2	5	14	NA	8	2	NA	55	51	NA	#Mittendorf MAGnet	2002
2	6	12	NA	8	2	NA	116	120	NA	#Bisits	2004
2	7	7	NA	15	4	NA	111	111	NA	#Holleboom	1996
2	7	11	NA	1	1	NA	35	35	NA	#Maitra	2007
2	7	11	NA	7	4	NA	43	48	NA	#Van de Water	2008
2	7	11	NA	28	17	NA	90	95	NA	#Papatsonis (1997/2000)	
2	7	13	NA	1	3	NA	63	63	NA	#Shim	2006
2	7	13	NA	5	3	NA	107	107	NA	#Moutquin	2000
2	8	13	NA	2	4	NA	99	109	NA	#French/Australian	2001
2	9	11	NA	3	0	NA	16	20	NA	#Laohapojanart	2007
2	9	13	NA	4	3	NA	105	101	NA	#European	2001
2	11	11	NA	0	4	NA	48	52	NA	#Nassar	2009

END

FIXED CLASS, RANDOM TREATMENT EFFECTS

Tocolytics: outcome is RDS
Class model - treatments exchangeable within class, within-class variance is zero (fixed class effects)

==========

6 August 2014

Treatments (code, Class, Treat)

1	1	Placebo	
2	2	Celecoxib	
3	2	Indomethacin	
4	2	Ketorolac	
5	2	Rofecoxib	
6	2	Sulindac	
7	3	Magnesium Sulfate	(TREATMENT IS ITS OWN CLASS)
8	4	Fenoterol	
9	4	Hexoprenaline	
10	4	Ritodrine	
11	4	Salbutamol	
12	4	Terbutaline	
13	4	Nylidrin	(NOT TO BE USED FOR RANKING)
14	5	Nicardipine	
15	5	Nifedipine	
16	6	Atosiban	
17	6	Barisiban 1.0	(NOT TO BE USED FOR RANKING)
18	6	Barusiban 0.3	(NOT TO BE USED FOR RANKING)
19	6	Barusiban 10	(NOT TO BE USED FOR RANKING)
20	6	Barusiban 3.0	(NOT TO BE USED FOR RANKING)
21	7	Ethanol (NOT TO BE USED FOR RANKING)	
22	8	Tocolysis	(NOT TO BE USED FOR RANKING)

Class "Nitrates" not compared
Classes 7 and 8 not to be used for ranking

==========

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials
# class effects - zero within-class variance
model{                                  # *** PROGRAM STARTS
for(i in 1:ns){                         # LOOP THROUGH STUDIES
    w[i,1] <- 0    # adjustment for multi-arm trials is zero for control
arm
    delta[i,1] <- 0             # treatment effect is zero for control arm
    mu[i] ~ dnorm(0,.0001)           # vague priors for all trial baselines
    for (k in 1:na[i]) {             # LOOP THROUGH ARMS
        r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
        logit(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
        rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
        dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
            +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-
rhat[i,k])))          }
#  summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
    for (k in 2:na[i]) {             # LOOP THROUGH ARMS
# trial-specific LOR distributions
        delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
```

```
        md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
        taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
        w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
        sw[i,k] <- sum(w[i,1:k-1])/(k-1)
      }
  }
totresdev <- sum(resdev[])           # Total Residual Deviance
d[1]<-0       # treatment effect is zero for reference treatment
# treatment effects from Class - fixed class effects
for (k in 2:nt){  d[k] <- m[D[k]]   }
sd ~ dunif(0,5)      # vague prior for between-trial SD
tau <- pow(sd,-2)    # between-trial precision = (1/between-trial variance)
m[1] <- 0
for (k in 2:nc){ m[k] ~ dnorm(0, .0001) } # priors for mean class effect
# all pairwise ORs
for (c in 1:(nt-1)) {
    for (k in (c+1):nt)  {
        lor[c,k]<- d[k]-d[c]
        OR[c,k] <- exp(lor[c,k])
      }
  }
# select treatments to be used for ranking and economic analysis
for(k in 1:12){ dR[k] <- d[k] }
# not treatment 13
for(k in 14:16){ dR[k-1] <- d[k] }
# not treatments 17-22
# ranking on relative scale
for (k in 1:ntR) {
#    rk[k]<- (ntR+1)-rank(dR[],k)      # events are "good"
    rk[k]<- rank(dR[],k)              # events are "bad"
    best[k] <- equals(rk[k],1)        # rank=1 is best
#calculate probability that treat k is h-th best
    for (h in 1:nt) { prob[h,k] <- equals(rk[k],h) }
  }
# Provide estimates of treatment effects T[k] on the natural scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:ntR) { logit(T[k]) <- A + dR[k]  }
# all pairwise ORs for classes
for (c in 1:(nc-1)){
    for (k in (c+1):nc){
        lorClass[c,k] <- m[k] - m[c]
        ORClass[c,k] <- exp(m[k] - m[c])
    }
 }
# rank all classes except last two
for (k in 1:nc-2)  {
    rkClass[k]  <- rank(m[1:(nc-2)],k)     # events are "bad"
    bestClass[k]  <- equals(rkClass[k],1)  # rank=1 is best
# prob class k is h-th best, prob[1,k]=best[k]
    for (h in 1:nc-2) { probClass[h,k] <- equals(rkClass[k],h) }
    }
}                                    # *** PROGRAM ENDS
```

```
# ns= number of studies; nt=number of treatments; nc=number of classes; D=index of classes
# ntR = number of treat for ranking
list(ns=47, nt=22, nc=8, meanA=-1.75, precA=0.555, ntR=15,
D=c(1, 2, 2, 2, 2, 2, 3, 4, 4, 4, 4, 4, 4, 5, 5, 6, 6, 6, 6, 6, 7, 8))
```

na[]	t[,1]	t[,2]	t[,3]	t[,4]	t[,5]	r[,1]	r[,2]	r[,3]	r[,4]	r[,5]	n[,1]	n[,2]
	n[,3]	n[,4]	n[,5]	#	Study							
5	1	17	18	19	20	1	2	0	7	2	32	31
	32	36	32	#	Thornton	2009						
4	1	10	10	10	NA	1	4	5	2	NA	45	44
	41	46	NA	#	Larsen	1980						
3	1	7	12	NA	NA	6	6	4	NA	NA	19	16
	19	NA	NA	#	Cotton	1984						
3	3	7	18	NA	NA	41	39	34	NA	NA	103	95
	119	NA	NA	#	Klauser	2012						
3	10	16	16	NA	NA	5	8	7	3	2	56	61
	58	62	57	#	Goodwin	1996						
2	1	3	NA	NA	NA	2	3	NA	NA	NA	15	16
	NA	NA	NA	#	Niebyl	1980						
2	1	3	NA	NA	NA	2	4	NA	NA	NA	20	19
	NA	NA	NA	#	Panter	1999						
2	1	3	NA	NA	NA	4	1	NA	NA	NA	18	18
	NA	NA	NA	#	Zuckerman		1984					
2	1	7	NA	NA	NA	15	15	NA	NA	NA	89	78
	NA	NA	NA	#	Cox	1990						
2	1	10	NA	NA	NA	3	0	NA	NA	NA	15	14
	NA	NA	NA	#	Spellacy	1979						
2	1	10	NA	NA	NA	6	3	NA	NA	NA	50	49
	NA	NA	NA	#	Larsen	1986						
2	1	10	NA	NA	NA	24	20	NA	NA	NA	122	187
	NA	NA	NA	#	Merkatz	1980						
2	1	10	NA	NA	NA	24	25	NA	NA	NA	55	56
	NA	NA	NA	#	Leveno	1986						
2	1	10	NA	NA	NA	90	69	NA	NA	NA	391	380
	NA	NA	NA	#	CPLIG	1992						
2	1	16	NA	NA	NA	0	3	NA	NA	NA	57	57
	NA	NA	NA	#	Goodwin	1994						
2	1	16	NA	NA	NA	54	64	NA	NA	NA	292	283
	NA	NA	NA	#	Romero	2000						
2	1	22	NA	NA	NA	22	15	NA	NA	NA	42	33
	NA	NA	NA	#	Weiner	1988						
2	2	3	NA	NA	NA	1	1	NA	NA	NA	12	12
	NA	NA	NA	#	Stika	2002						
2	3	6	NA	NA	NA	1	0	NA	NA	NA	10	10
	NA	NA	NA	#	Rasanen	1995						
2	3	7	NA	NA	NA	5	5	NA	NA	NA	49	52
	NA	NA	NA	#	Morales	1993						
2	3	7	NA	NA	NA	5	5	NA	NA	NA	14	18
	NA	NA	NA	#	Parilla	1997						
2	3	10	NA	NA	NA	8	12	NA	NA	NA	47	50
	NA	NA	NA	#	Morales	1989						
2	3	13	NA	NA	NA	3	2	NA	NA	NA	30	30
	NA	NA	NA	#	Kurki	1991						
2	4	7	NA	NA	NA	2	4	NA	NA	NA	45	43
	NA	NA	NA	#	Schorr	1998						
2	5	7	NA	NA	NA	18	19	NA	NA	NA	92	102
	NA	NA	NA	#	McWhorter		2004					
2	7	12	NA	NA	NA	3	2	NA	NA	NA	15	16
	NA	NA	NA	#	Miller	1982						
2	7	15	NA	NA	NA	4	5	NA	NA	NA	40	50
	NA	NA	NA	#	Floyd	1995						
2	7	15	NA	NA	NA	24	21	NA	NA	NA	106	110
	NA	NA	NA	#	Lyell	2007						
2	8	10	NA	NA	NA	4	2	NA	NA	NA	48	48
	NA	NA	NA	#	Essed	1978						
2	9	11	NA	NA	NA	7	4	NA	NA	NA	70	70
	NA	NA	NA	#	Gummerus		1983					
2	10	10	NA	NA	NA	17	12	NA	NA	NA	111	111
	NA	NA	NA	#	Holleboom		1996					
2	10	12	NA	NA	NA	2	5	NA	NA	NA	31	26
	NA	NA	NA	#	Caritis	1984						
2	10	15	NA	NA	NA	1	0	NA	NA	NA	35	35
	NA	NA	NA	#	Maitra	2007						
2	10	15	NA	NA	NA	3	2	NA	NA	NA	39	39
	NA	NA	NA	#	Cararach	2006						

```
2   10  15  NA  NA  NA  3   3   NA  NA  NA  43  48
    NA  NA  NA  #   Van de Water 2008
2   10  15  NA  NA  NA  4   4   NA  NA  NA  28  30
    NA  NA  NA  #   Al-Qattan 2000
2   10  15  NA  NA  NA  31  23  NA  NA  NA  90  95
    NA  NA  NA  #   Papatsonis (1997/2000)   1997
2   10  16  NA  NA  NA  0   3   NA  NA  NA  63  63
    NA  NA  NA  #   Shim 2006
2   10  16  NA  NA  NA  1   0   NA  NA  NA  22  23
    NA  NA  NA  #   Lin 2009
2   10  16  NA  NA  NA  14  15  NA  NA  NA  107 107
    NA  NA  NA  #   Moutquin 2000
2   10  21  NA  NA  NA  6   15  NA  NA  NA  73  76
    NA  NA  NA  #   Lauersen 1977
2   11  14  NA  NA  NA  3   5   NA  NA  NA  21  24
    NA  NA  NA  #   Trabelsi 2008
2   11  16  NA  NA  NA  10  14  NA  NA  NA  99  109
    NA  NA  NA  #   French/Australian 2001
2   12  15  NA  NA  NA  2   2   NA  NA  NA  16  20
    NA  NA  NA  #   Laohapojanart 2007
2   12  16  NA  NA  NA  28  17  NA  NA  NA  105 101
    NA  NA  NA  #   European 2001
2   15  15  NA  NA  NA  6   10  NA  NA  NA  48  52
    NA  NA  NA  #   Nassar 2009
2   15  16  NA  NA  NA  10  5   NA  NA  NA  23  25
    NA  NA  NA  #   Al-Omari 2006
END
```

SAMPLE WINBUGS CODE FOR EGA

FIXED CLASS, RANDOM TREATMENT EFFECTS

Tocolytics: outcome is EGA at delivery
Class model - treatments exchangeable within class, within-class variance is zero (fixed class effects)

==========

1 August 2014

Treatments (code, Class, Treat)

1	1	Placebo	
2	2	Celecoxib	
3	2	Indomethacin	
4	2	Ketorolac	
5	2	Nimeluside	
6	2	Rofecoxib	
7	2	Sulindac	
8	3	Magnesium Sulfate	(TREATMENT IS ITS OWN CLASS)
9	4	Fenoterol	
10	4	Isoxsuprine	
11	4	Ritodrine	
12	4	Salbutamol	
13	4	Terbutaline	
14	4	Nylidrin	(NOT TO BE USED FOR RANKING)
15	5	Nicardipine	
16	5	Nifedipine	
17	6	Nitric Oxide	
18	7	Atosiban	
19	8	Tocolysis	(NOT TO BE USED FOR RANKING)

Class "Alcohol/ethanol" not compared
Class 8 not to be used for ranking

==========

```
# Normal likelihood, identity link
# Random effects model for multi-arm trials
# class effects - zero within-class variance
model{                                  # *** PROGRAM STARTS
for(i in 1:ns){                         #   LOOP THROUGH STUDIES
    w[i,1] <- 0    # adjustment for multi-arm trials is zero for control
arm
    delta[i,1] <- 0             # treatment effect is zero for control arm
    mu[i] ~ dnorm(0,.0001)           # vague priors for all trial baselines
    for (k in 1:na[i]) {             #  LOOP THROUGH ARMS
        var[i,k] <- pow(se[i,k],2)   # calculate variances
        prec[i,k] <- 1/var[i,k]      # set precisions
        y[i,k] ~ dnorm(theta[i,k],prec[i,k]) # binomial likelihood
        theta[i,k] <- mu[i] + delta[i,k]  # model for linear predictor
#Deviance contribution
        dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]
      }
#  summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
    for (k in 2:na[i]) {             # LOOP THROUGH ARMS
# trial-specific LOR distributions
        delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions, with multi-arm trial correction
```

```
        md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
        taud[i,k] <- tau *2*(k-1)/k
# adjustment, multi-arm RCTs
        w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
        sw[i,k] <- sum(w[i,1:k-1])/(k-1)
      }
  }
totresdev <- sum(resdev[])            #Total Residual Deviance
d[1]<-0       # treatment effect is zero for control arm
# treatment effects from Class - fixed class effects
for (k in 2:nt){  d[k] <- m[D[k]]   }
sd ~ dunif(0,20)     # vague prior for between-trial SD
tau <- pow(sd,-2)    # between-trial precision = (1/between-trial variance)
m[1] <- 0
for (k in 2:nc){ m[k] ~ dnorm(0, .0001) } # priors for mean class effect
# all pairwise differencess
for (c in 1:(nt-1)) {
    for (k in (c+1):nt)  { diff[c,k]<- d[k]-d[c] }
  }
# select treatments to be used for ranking
for(k in 1:13){ dR[k] <- d[k] }
# not treatment 14
for(k in 15:18){ dR[k-1] <- d[k] }
# not treatment 19
# ranking on relative scale
for (k in 1:ntR) {
    rk[k]<- (ntR+1)-rank(dR[],k)     # larger values are "good"
#    rk[k]<- rank(dR[],k)             # larger values are "bad"
    best[k] <- equals(rk[k],1)       # rank=1 is best
#calculate probability that treat k is h-th best
    for (h in 1:nt) { prob[h,k] <- equals(rk[k],h) }
  }
# all pairwise differences for classes
for (c in 1:(nc-1)){
    for (k in (c+1):nc){ diffClass[c,k] <- m[k] - m[c] }
 }
# rank all classes except 8
for (k in 1:nc-1)  {
    rkClass[k]  <- nc-rank(m[1:(nc-1)],k)  # larger values are "good"
    bestClass[k]  <- equals(rkClass[k],1)  # rank=1 is best
# prob class k is h-th best, prob[1,k]=best[k]
    for (h in 1:nc-1) { probClass[h,k] <- equals(rkClass[k],h) }
    }
}                                      # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments; nc=number of classes; D=index of classes
# ntR = number of treat for ranking
list(ns=49, nt=19, nc=8, ntR=17,
D=c(1, 2, 2, 2, 2, 2, 2, 3, 4, 4, 4, 4, 4, 4, 5, 5, 6, 7, 8))
```

na[]	t[,1]	t[,2]	t[,3]	y[,1]	y[,2]	y[,3]	se[,1]	se[,2]	se[,3]	#	Study	Year
3	1	8	13	32	31	33.1	0.780013495		0.475	0.757071922	#	
	Cotton	1984										
3	3	5	7	37.2	38.4	38.1	0.632455532		0.158113883	0.316227766		
	#	Sawdy	2003									
3	3	8	16	31.8	31.2	31.8	0.450287265		0.423014393	0.441261304		
	#	Klauser	2012									

2	1	3	NA	31.2	36.4	NA	0.164991582	0.164991582	NA	#
	Zuckerman		1984							
2	1	3	NA	33	35.2	NA	0.309838668	0.284018779	NA	#
	Niebyl	1980								
2	1	3	NA	29.1	29.1	NA	1.107800624	1.4	NA	#
	Panter	1999								
2	1	8	NA	33	33.8	NA	0.055901699	0.057353933	NA	#
	Cox	1990								
2	1	8	NA	36.5	35.7	NA	0.401663209	0.367423461	NA	#
	How	2006								
2	1	10	NA	32.9	38.7	NA	0.242535625	0.114707867	NA	#
	Casapo	1977								
2	1	11	NA	33.4	34	NA	0.095399809	0.090610304	NA	#
	CPLIG	1992								
2	1	11	NA	32.5	34.6	NA	0.190125067	0.226694451	NA	#
	Merkatz	1980								
2	1	11	NA	32.6	32.8	NA	0.291217603	0.421856567	NA	#
	Leveno	1986								
2	1	11	NA	36.3	37.2	NA	0.296984848	0.442857143	NA	#
	Larsen	1986								
2	1	17	NA	34.1	35.2	NA	0.742558015	0.56961343	NA	#
	Smith	2007								
2	1	18	NA	38.3	37.8	NA	0.280624304	0.467707173	NA	#
	Goodwin	1994								
2	1	19	NA	30.1	31	NA	0.509201055	0.504825202	NA	#
	Weiner	1988								
2	2	3	NA	35.7	35.7	NA	1.068097998	0.952627944	NA	#
	Stika	2002								
2	2	8	NA	35.5	35.7	NA	0.291217603	0.402157642	NA	#
	Borna	2007								
2	3	7	NA	39	39	NA	0.807961942	0.807961942	NA	#
	Rasanen	1995								
2	3	8	NA	30.8	31.1	NA	1.096965511	1.241303079	NA	#
	Parilla	1997								
2	3	11	NA	35.5	33.8	NA	0.620414085	0.853242183	NA	#
	Besinger	1991								
2	3	14	NA	36.7	35.2	NA	0.146059349	0.146059349	NA	#
	Kurki	1991								
2	3	16	NA	35.2	34.1	NA	0.352281938	0.43204938	NA	#
	Kashanian		2011							
2	4	8	NA	34.9	34.8	NA	0.536656315	0.655743852	NA	#
	Schorr	1998								
2	6	8	NA	35.3	34.7	NA	0.331806025	0.40228704	NA	#
	McWhorter		2004							
2	8	13	NA	36.21	36.01	NA	0.46	0.474976691	NA	#
	Surichamorn		2001							
2	8	15	NA	35.5	35.6	NA	0.396911151	0.490076972	NA	#
	Larmon	1999								
2	8	16	NA	34.1	34.3	NA	0.191502	0.176162803	NA	#
	Taherian	2007								
2	8	16	NA	35.2	34.5	NA	0.484138662	0.448358831	NA	#
	Glock	1993								
2	8	16	NA	35.8	36	NA	0.354474504	0.31	NA	#
	2007									
2	9	11	NA	37.4	36.9	NA	0.346410162	0.404145188	NA	#
	Essed	1978								
2	10	11	NA	35	35.6	NA	0.547722558	0.481995851	NA	#
	Sirohiwal	2001								
2	10	16	NA	33.46	34.98	NA	0.394360241	0.4118897	NA	#
	Rayamajhi		2003							
2	11	11	NA	35.7	35.4	NA	0.308461529	0.31	NA	#
	Holleboom		1996							
2	11	16	NA	29.5	30.2	NA	0.434659144	0.474692883	NA	#
	Al-Qattan	2000								
2	11	16	NA	36.1	36.2	NA	0.384307569	0.384307569	NA	#
	Cararach	2006								
2	11	16	NA	32.1	33.4	NA	0.464233584	0.461690258	NA	#
	Papatsonis (1997/2000)			1997						
2	11	16	NA	34.07	34.71	NA	0.794197708	0.495710634	NA	#
	Fan	2003								
2	11	16	NA	31.8	33.3	NA	0.656392462	0.593295879	NA	#
	Koks	1998								
2	11	18	NA	37.4	37.1	NA	0.511681719	0.521286035	NA	#
	Lin	2009								