National Institute for Health and Care Excellence

Final

Postnatal care

[L2] Scoring systems for illness in babies

NICE guideline NG194

Evidence review underpinning recommendations 1.3.11 to 1.3.12 and 1.4.2

April 2021

Final

This evidence review was developed by the National Guideline Alliance, part of the Royal College of Obstetricians and Gynaecologists

Disclaimer

Copyright

ISBN: 978-1-4731-4078-3

Contents

Scoring systems for illness in babies

Review question

Which scoring systems are accurate in identifying or predicting illness severity in babies?

Introduction

Following the neonatal period, the highest incidence of illness and death occurs in the first six months of life compared to the rest of childhood. Although many babies showing signs and symptoms will have a self-limiting illness, a minority will have a serious or even life-threatening illness. Early recognition of signs and symptoms in babies, and early treatment, is therefore important to help reduce the severity of illness and prevent deaths. The aim of this review is to find out which scoring systems are accurate in identifying or predicting illness severity in babies.

Summary of the protocol

Please see Table 1 for a summary of the population, index tests/clinical prediction models, and outcome characteristics of this review.

Table 1: Summary of the protocol

Population	Babies born at term, between 37 and 42 weeks of pregnancy
Index tests/clinical prediction models	A validated scoring system based on a combination of symptoms and/or signs for babies within the first 8 weeks after birth used by healthcare professionals or parents, either face-to-face or remotely.
Outcomes	• Well/mildly unwell for example defined as no clinical intervention needed • Moderately unwell for example defined as requiring clinical attention • Seriously unwell for example defined as admission to hospital or treatment in hospital Severity of illness (or absence of) defined by a qualified assessor through a comprehensive assessment

For further details see the review protocol in appendix A.

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual 2014. Methods specific to this review question are described in the review protocol in appendix A.

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy until March 2018. From April 2018 until June 2019, declarations of interest were recorded according to NICE's 2018 conflicts of interest policy. From July 2019 onwards, the declarations of interest were recorded according to NICE's 2019

conflicts of interest policy. Those interests declared before July 2019 were reclassified according to NICE's 2019 conflicts of interest policy (see Register of Interests).

Clinical evidence

Included studies

Five publications from 4 prospective cohort studies were included in this review (Chandran 1998, Chen 1997, Cole 1991, Morley 1991, Thornton 1991). Two publications (Cole 1991; Morley 1991) were based on the same data collection, however one paper focused on more detailed methods of the scoring system development (Cole 1991) and the other the accuracy of the scoring system (Morley 1991).

All studies reported on the Baby Check scoring system and included babies up to 6 months of age.

Studies were conducted in Australia (Cole 1991, Morley 1991); Oman (Chandran 1998); Taiwan (Chen 1997); and the UK (Cole 1991, Morley 1991, Thornton 1991).

Two studies took place in hospital (Chen 1997, Thornton 1991), 1 study was based in a polyclinic (Chandran 1998), and 1 study from 2 publications collected data from both a hospital and community cohort (Cole 1991, Morley 1991).

One study reported in 2 publications (Cole 1991, Morley 1991) extrapolated data from a hospital and community cohort to model a theoretical community cohort of 10,000 infants.

The included studies are summarised in Table 2.

See the literature search strategy in appendix B and also the study selection flow chart in appendix C.

Excluded studies

Studies not included in this review with reasons for their exclusions are provided in appendix K.

Summary of studies included in the evidence review

A summary of the study included in this review is presented in Table 2.

Table 2: Summary of included study.

Study	Population	Scoring system	Reference standard	Outcome
Cole 1991; Morley 1991 Prospective cohort study UK, Australia	N=1007 infants younger than 6 months of age n=298 infants assessed at home n=709 infants assessed in hospital	Baby check scoring system conducted by 2 independent assessors (healthcare professionals) in hospital setting. Unclear who conducted the Baby check	Babies assessed in hospital Paediatrician subjectively graded each baby's illness into categories of well, mildly ill, moderately ill, and seriously ill based on a 7-point scale. Other criteria, such as investigation results and a review of the notes by 3 independent	• Infants needing to be admitted to hospital through to infants that need urgent hospital attention for a life threatening condition (score of 13 or more)** • Infants who were moderately ill (score 8-12) • Infants who were well or mildly ill (score 0-7)

Study	Population	Scoring system	Reference standard	Outcome
	Prevalence of infants requiring hospital admission for observation or treatment (score 13 or more): 0% (home cohort); 23% (hospital cohort)	scoring system in the community setting.	paediatricians were used, where possible. <u>Babies assessed in the community</u> A research nurse on 2 weekdays per week subjectively graded the babies based on a 7-point scale.	
Chandran 1998 Prospective cohort study Oman	N= 90 infants younger than 6 months of age All infants were assessed in a polyclinic, which had a paediatrician with specialist training available for consultation and a facility for observation for a limited period. Prevalence of infants requiring further consultation and/or referral to tertiary care (score 13 or more): 52%	Baby check scoring system conducted by a junior doctor	Junior doctor graded infants as well or mildly ill, moderately ill, and seriously ill. The majority of cases were reviewed by the same physician or the specialist (paediatrician with specialist training).	• Infants needing immediate referral to tertiary care (score of 20 or more)** • Infants needing observation and consultation (score 13-19)** • Infants needing minor medication and/or reassurance (score of less than 13)
Chen 1997 Prospective cohort study Taiwan	N=134 infants younger than 6 months of age All infants were assessed in a hospital emergency department Prevalence of infants requiring hospital admission for observation or treatment (score 13 or more): 31% (paediatrician 'B')	Baby check scoring system translated into Chinese conducted by an inter*n	Two senior paediatricians* 'A' and 'B' (third-year paediatric residents) reviewed the medical records after the babies were discharged, and graded the severity of the illness as well, mildly ill, moderately ill, and seriously ill.	• Infants who required hospital treatment (score 20 or more)** • Infants who required hospital admission for observation when there was uncertainty about the severity of illness (score 13-19)** • Infants who required careful observation and treatment, but could be managed at home by a capable person (score 8-12) • Infants who could be managed at home (score 0-7)
Thornton 1991 Prospective cohort study	N=193 infants younger than 6 months of age All infants were assessed in hospital	Baby check scoring system conducted by a paediatric house officer	Two consultant paediatricians 'A' and 'B' reviewed each baby's notes after discharge, and graded the severity of illness into 1 of 4	• Infants who required hospital treatment (score 20 or more)** • Infants who required hospital admission for observation when there was uncertainty about the

Study	Population	Scoring system	Reference standard	Outcome
UK	(majority in a casualty department) Prevalence of infants requiring hospital admission for observation or treatment (score 13 or more): 73% (consultant 'A'); 65% (consultant 'B')		categories (see outcomes).	severity of illness (score 13-19)** • Infants who required careful observation and treatment, but could be managed at home by a capable person (score 8-12) • Infants who could be managed at home (score 0-7)

*Two senior paediatricians 'A' and 'B' conducted the assessments in Chen 1997, however there was only sufficient data for paediatrician 'B' reported to calculate diagnostic accuracy outcomes
**Babies who required hospital admission for observation and treatment were reported separately in Chandran 1998, Chen 1997, and Thornton 1991, where as in Cole 1991 babies who required hospital admission for observation and treatment were reported together. We combined data for babies who required hospital admission for observation and treatment in Chandran 1998, Chen 1997, and Thornton 1991 from the 2x2 tables in the primary studies to give the same definition across all studies and therefore appropriately pool the data for this outcome.

See the full evidence tables in appendix D and the forest plots in appendix E.

Quality assessment of studies included in the evidence review

See the evidence profile in appendix F.

Economic evidence

Included studies

A single economic search was undertaken for all topics included in the scope of this guideline but no economic studies were identified which were applicable to this review question. See the literature search strategy in appendix B and economic study selection flow chart in appendix G.

Excluded studies

No economic studies were reviewed at full text and excluded from this review.

Economic model

No economic modelling was conducted for this review question because the committee agreed that other topics were higher priorities for economic evaluation.

Evidence statements

Clinical evidence statements

Babies assessed in secondary care

Babies seriously unwell defined as requiring admission to hospital for observation or treatment

- Very low to moderate quality evidence from 3 prospective cohort studies (N=417) showed mixed results, with sensitivity ranging from 0.46 (95% CI 0.35 to 0.56) to 1.00 (95% CI 0.54 to 1.00), and specificity ranging from 0.81 (95% CI 0.71 to 0.89) to 0.94 (95% CI 0.88 to 0.97) for a Baby Check score of 20 or more to identify babies with serious illness requiring hospital treatment. The evidence for a Baby Check score of 20 or more was of a wide range, therefore it is not possible to ascertain how useful it is in identifying babies with serious illness requiring hospital treatment.

- Very low quality evidence from 4 meta-analysed prospective cohort studies (N=1,126) showed a sensitivity of 0.75 (95% CI 0.57 to 0.87) and a specificity of 0.79 (95% CI 0.72 to 0.85) for a Baby Check score of 13 or more to identify babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness or for serious illness requiring hospital treatment. The evidence suggests that a Baby Check score of 13 or more is moderately useful for identifying babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness and for identifying serious illness requiring hospital treatment.

- Very low to moderate quality evidence from 3 prospective cohort studies (N=417) showed mixed results, with sensitivity ranging from 0.23 (95% CI 0.11 to 0.38) to 0.37 (95% CI 0.22 to 0.53), and specificity ranging from 0.80 (95% CI 0.73 to 0.86) to 0.94 (95% CI 0.83 to 0.99) for a Baby Check score of 13 to 19 to identify babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness. The evidence suggests that a Baby Check score of 13 to 19 is not useful in identifying babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness.

Babies moderately unwell defined as requiring clinical attention

- Very low to moderate quality evidence from 3 prospective cohort studies (N=1,036) showed mixed results, with sensitivity ranging from 0.19 (95% CI 0.08 to 0.33) to 0.33 (95% CI 0.22 to 0.45), and specificity ranging from 0.83 (95% CI 0.79 to 0.86) to 0.86 (95% CI 0.79 to 0.91) for a Baby Check score of 8 to 12 to identify babies that need careful observation and treatment ("could be managed at home by a capable mother"). The evidence suggests that a Baby Check score of 8 to 12 is not useful in identifying babies that need careful observation and treatment.

Babies well or mildly unwell defined as no clinical intervention needed

- Very low to moderate quality evidence from 3 prospective cohort studies (N=1,036) showed mixed results, with sensitivity ranging from 0.62 (95% CI 0.56 to 0.67) to 0.92 (95% CI 0.62 to 1.00), and specificity ranging from 0.64 (95% CI 0.55 to 0.73) to 0.86 (95% CI 0.83 to 0.90) for a Baby Check score of 0 to 7 to identify babies that are well and could be managed at home by any mother. The evidence for a Baby Check Score of 0 to 7 was of a wide range, therefore it is not possible to ascertain how useful it is in identifying babies that are well and "could be managed at home by any mother".

Babies assessed in the community

Babies seriously unwell defined as requiring admission to hospital for observation or treatment

- Low quality evidence from 1 prospective cohort study (N=298) showed a specificity of 0.99 (0.97 to 1.00) for a Baby Check score of 13 or more to identify babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness or for serious illness requiring hospital treatment. The sensitivity and usefulness of the Baby Check score of 13 or more for identifying babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness or for serious illness requiring hospital treatment could not be ascertained. This is because no events were recorded in the cohort.

Babies moderately unwell defined as requiring clinical attention

- Very low quality evidence from 1 prospective cohort study (N=298) showed a sensitivity of 0.38 (95% CI 0.085 to 0.76) and a specificity of 0.99 (95% CI 0.97 to 1.00) for a Baby Check score of 8 to 12 to identify babies that need careful observation and treatment ("could be managed at home by a capable mother"). The evidence suggests that a Baby Check score of 8 to 12 is not useful for identifying babies that need careful observation and treatment.

Babies well or mildly unwell defined as no clinical intervention needed

- Low quality evidence from 1 prospective cohort study (N=298) showed a sensitivity of 0.99 (95% CI 0.97 to 1.00) and a specificity of 0.63 (95% CI 0.25 to 0.92) for a Baby Check score of 0 to 7 to identify babies that are well and could be managed at home by any mother. The evidence suggests that a Baby Check score of 0 to 7 is very useful in identifying babies that are well and "could be managed at home by any mother".

Babies in a theoretical community cohort

Babies seriously unwell defined as requiring admission to hospital for observation or treatment

- Very low quality evidence from an analysis of 1 theoretical cohort showed a sensitivity of 0.93 (95% CI 0.82 to 0.99) and a specificity of 0.99 (95% CI 0.98 to 1.00) for a Baby Check score of 13 or more for identifying babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness or for serious illness requiring hospital treatment. The evidence suggests that a Baby Check score of 13 or more is very useful for identifying babies that need to be admitted to hospital for observation due to uncertainty about the severity of illness and for identifying serious illness.

Babies moderately unwell defined as requiring clinical attention

- Very low quality evidence from an analysis of 1 theoretical cohort showed a sensitivity of 0.36 (95% CI 0.31 to 0.41) and a specificity of 0.99 (95% CI 0.985 to 0.99) for a Baby Check score of 8 to 12 to identify babies that need careful observation and treatment ("could be managed at home by a capable mother"). The evidence suggests that a Baby Check score of 8 to 12 is not useful in identifying babies that need careful observation and treatment.

Babies well or mildly unwell defined as no clinical intervention needed

- Very low quality evidence from an analysis of 1 theoretical cohort showed a sensitivity of 0.98 (95% CI 0.98 to 0.99) and a specificity of 0.70 (95% CI 0.65 to 0.74) to identify babies that are well and could be managed at home by any

mother. The evidence suggests that a Baby Check score of 0 to 7 is very useful in identifying babies that are well and "could be managed at home by any mother".

Economic evidence statements

No economic evidence was identified which was applicable to this review question.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee prioritised sensitivity as a critical outcome for this review. Severity of illness scoring systems aim to identify babies that are well or unwell and either offers reassurance that the baby is healthy or ensures that the unwell baby is identified so that he or she can receive appropriate monitoring and/or management to reduce the risk of complications. Therefore, the priority is to ensure that the scoring system identifies the baby's true health state, for example seriously ill. While false positives may mean that babies undergo unnecessary follow up, this is less of a concern than failing to identify babies who are seriously unwell and need intensive monitoring or intervention.

Calibration and discrimination were also identified as critical outcomes in this review, however no clinical prediction model studies were identified so these outcomes were not reported.

The quality of the evidence

The committee were aware that evidence from a clinical prediction model study was the most appropriate study design to answer the review question, allowing calibration and discrimination of the data. In view of the absence of a clinical prediction model study, diagnostic accuracy studies were included and the limitations of the data were discussed with the committee when assessing the evidence.

The evidence was assessed using a modified GRADE for diagnostic test accuracy. The overall confidence in the review findings ranged from very low to moderate.

<u>Babies assessed in secondary care</u>

The quality of the evidence ranged from very low to moderate. There was no serious risk of bias across any of the included studies: often not all babies enrolled in the study were included in the analysis, but reasons for exclusion were well documented and valid (not all babies who were scored using Baby Check were assessed by the consultant; or babies who were seen by the consultant were not previously scored using Baby Check); also in 1 study (Chandran 1998) there was ambiguity around whether every baby was assessed by an experienced paediatrician, however as the assessment was comprehensive and there is no "gold standard" for assessment the study was not downgraded.

Only grade 1 and 2 illness (score of 13 or more) had sufficient data for meta-analysis. One study (Cole 1991) reported babies admitted to hospital for observation and treatment together, whereas 3 studies (Chandran 1998, Chen 1997, Thornton 1991) reported babies admitted to hospital for observation and for treatment separately. However, we were able to combine data for babies admitted to hospital for observation and treatment in the 3 studies (Chandran 1998, Chen 1997, Thornton

1991) from the 2x2 tables in the primary studies to give the same definition across all studies and therefore appropriately meta-analyse. The results for the remaining health states reported were from individual studies. In the meta-analysis for grade 1 and 2 illness, the evidence was downgraded due to very high and high heterogeneity (assessed using the I^2 statistic). However, it was noted that heterogeneity is often high with diagnostic accuracy studies, and therefore this downgrading of the evidence should be interpreted with caution. However, the driving factor influencing the decision making was the relatively low sensitivity of the Baby Check scoring system to identify serious illness (grade 1 and 2) in babies, and therefore heterogeneity, while acknowledged, was not a determining factor in their decision making.

The quality of the evidence was downgraded for indirectness as the population included were infants under 6 months old, whereas the population of interest for this review was infants 8 weeks old or less.

Individual studies were downgraded due to serious or very serious imprecision of the effect estimate, that is, the confidence interval crossed the upper threshold of 0.9 and/or the lower threshold of 0.75.

Babies assessed in the community

The quality of the evidence ranged from very low to moderate. There was serious bias in the included study as a research nurse's grading of health state served as the reference standard compared to the usual experienced paediatrician or at minimum a physician. This is not to say that a research nurse's diagnosis of serious illness is necessarily inferior to a physician's, but given the lack of information on the research nurse's expertise and experience and the differing qualifications compared to the reference standard in the other included studies, it was deemed appropriate to downgrade the quality of evidence.

The quality of the evidence was downgraded for serious indirectness as the population included were infants under 6 months old, whereas the population of interest for this review were infants 8 weeks old or less.

Individual studies were downgraded due to serious or very serious imprecision of the effect estimate, that is, the confidence interval crossed the upper threshold of 0.9 and/or the lower threshold of 0.75.

Babies in a theoretical community cohort

The quality of evidence was very low. There was serious bias in the included study as a research nurse's grading of health state served as the reference standard compared to the usual experienced paediatrician or at minimum a physician.

The quality of the evidence was downgraded for very serious indirectness as the population included were infants under 6 months old whereas the population of interest for this review was infants 8 weeks old or less. Additionally, the results were based on a theoretical cohort extrapolated from 2 cohorts in the study, which were based on assumptions.

The committee agreed that the evidence in the theoretical community cohort was not considered particularly helpful as the data were based on assumptions from a proportion of babies in the study. The committee were therefore not sufficiently confident in the accuracy of the data to base recommendations on this data, not least given the consequences of failing to identify a seriously ill baby.

Benefits and harms

The committee noted that the only available evidence located was on the Baby Check scoring system. The committee agreed that the main priority for a scoring system is to identify well babies from those that need further assessment or treatment. The evidence showed that the sensitivity of the Baby Check scoring system to identify well or mildly unwell babies in the community was high. However, because of the uncertainties and concerns around the evidence this should not be taken to be a definitive indication that no further assessment or care is needed. However, the committee agreed that the Baby Check scoring system may be a useful tool to be used in the community when parents/caregivers are unsure whether their baby is unwell to help them decide whether to seek help from a healthcare professional.

The committee discussed the potential harms of the poor specificity of the Baby Check scoring system to identify when the baby is unwell. The committee agreed that this could cause undue anxiety for parents/caregiver and burden on healthcare services if the baby is well. Nonetheless, the committee agreed that the benefits of not missing a seriously unwell baby outweigh the harms potential harms discussed. The committee also recognised that the evidence review does not tell whether the use of the Baby Check scoring system will increase requests for medical advice nor whether it will ultimately improve health outcomes for the baby.

The committee discussed the differing accuracy of the Baby Check scoring system in the community compared to secondary care. The committee agreed that the differences might be due to the fact that it's easier to identify babies that are well in the community as the majority of babies are fit and healthy. Whereas, babies presenting to secondary care are assessed as there is some concern over their health, therefore making it more difficult to identify babies that are well from those that are unwell. The difference could also potentially be explained by the different type of assessors used in the settings.

The committee discussed the 2 groups of users of a scoring system to identify serious illness: parents/caregivers and healthcare professionals (HCPs). The committee agreed that recommendations for parents/caregiver and HCPs using the scoring system should be separate as different considerations would need to be made for both groups of users to ensure that babies who are seriously unwell are admitted to hospital for observation and/or treatment and not missed.

The committee highlighted that in the included studies, healthcare professionals completed the scoring system to identify serious illness in babies. In view of this, it is difficult to ascertain whether the same diagnostic accuracy would have been achieved if parents/caregivers had used the scoring systems with their babies. The committee agreed that given the serious consequences of failing to identify a seriously ill baby, the diagnostic accuracy of the evidence from healthcare professionals should not be extrapolated to parents/caregivers. Therefore, if a parent/caregiver thinks that their baby is ill or unwell, they should seek advice from a healthcare professional. The committee emphasised that if parents or caregivers think their baby is seriously ill, they should contact 999 immediately without delay and not rely on a scoring system to confirm that their baby is seriously ill.

The committee discussed the potential harms of the Baby Check scoring system, where temperature was measured rectally, which is not in line with current practice. In the NICE guideline on fever in under 5s (CG160), axillary temperature measurement with an electronic thermometer for babies less than 4 weeks of age, and an axillary temperature measurement with an electronic or chemical dot

thermometer, or ear temperature measurement with an infrared tympanic thermometer for babies 4 weeks of age or older is recommended. Axillary and rectal temperature measurement are not equivalent with rectal temperature being slightly higher than axillary temperature (around 0.5^0C). The implications of this may be that a baby with an axillary temperature of for example 37.9^0C would score lower on the Baby Check scoring system than if they had their rectal temperature taken which would be roughly equivalent to 38.4^0C or more. In some situations this would give false reassurance that the baby is well.

The committee also discussed the transferability of the Baby Check scoring system which was developed in babies up to 6 months of age and very young babies for example a 1 week old. The committee agreed that babies at both ends of the age range are quite different and that some of the Baby Check scoring system domains are difficult to assess in a very young baby for example “Is the baby concentrating on you less than you would expect?”. Despite the limitations of the Baby Check scoring system, the committee agreed there were benefits in using it to prompt parents/caregivers to seek advice from a healthcare professional when they are unsure if their baby is unwell.

The committee agreed that parents/caregivers could be provided with information about the Baby Check scoring system to allow them to familiarise themselves with the scoring system domains, helping them identify when their baby is unwell and giving a reference point (or baseline) with which to compare. The committee discussed the importance of the timing of information provision on the Baby Check scoring system to parents/caregivers, with the emphasis that parents/caregivers should be introduced to the Baby Check scoring system with other postnatal information. This was seen as more beneficial than waiting until there is a suspicion that the baby is unwell before introducing the Baby Check scoring system.

The sensitivity of the Baby Check scoring system to identify seriously ill babies requiring hospital admission for observation or treatment was moderate in secondary care. The committee emphasised that the accuracy of the Baby Check scoring system to identify serious illness is insufficient to recommend its use in isolation at the risk of missing seriously ill babies. Nonetheless, the committee agreed that the Baby Check scoring system may be a useful tool to use alongside the clinical assessment of the baby. In particular, the Baby Check scoring system may be useful when the healthcare professional can’t assess the baby physically, for example during a remote appointment via video call or phone, thus giving a more comprehensive assessment. Furthermore, it may aid communication around the baby’s condition when using a pre-defined checklist that both healthcare professionals and parents/caregivers can work from.

The committee were aware that a modified version of the Baby Check scoring system is available through the Lullaby Trust. The Lullaby Trust’s ‘Baby Check’ derives from the original Baby Check scoring system and can be accessed via a mobile app (the ‘Lullaby Trust Baby Check’ app) for those with access to a smartphone or a booklet version available as a pdf on their website that can be printed out, where the same questions and scores are given. Advantages of the Lullaby Trust’s ‘Baby Check’ is that it guides parents/caregivers through the meaning of each item and how to check each of the observations involved.

The Lullaby Trust’s ‘Baby Check’ mobile app is currently the only mobile app that is based on the evidence underpinning the Baby Check scoring system. Although the committee were aware of other mobile apps and online checklists assessing illness in babies, they know these are not based on the evidence located by this review.

The committee highlighted that the original Baby Check scoring system and the Lullaby Trust's 'Baby Check' (the app and the pdf booklet) are very similar, however it is important to note that there are differences between the systems. As mentioned previously, the original Baby Check scoring system measured temperature in babies rectally, which is not in line with clinical practice. The Lullaby Trust's 'Baby Check' has taken this into account and advises temperature measurement via the axilla or ear in line with current practice, nonetheless the cut-off remains in line with the original Baby Check scoring system thus there is still potential for error and giving false reassurance that their baby is well.

A further discrepancy in temperature measurement between the Lullaby Trust's 'Baby Check' mobile app and pdf booklet version was discussed by the committee. In the pdf booklet version different temperature cut-offs were used for babies under and over 3 months of age which is in line with NICE guideline on fever in under-5s (CG160) whereas the mobile app has one cut-off for all babies. The committee agreed that fever in younger babies is cause for concern and less common than in older babies, thus the lower cut-off for younger babies is appropriate, with the risk of one temperature cut-off for all babies being that younger babies with a low-grade fever are missed and deemed well when they may be unwell and need monitoring or treatment.

The committee also discussed the equal scoring of the temperature cut-offs for younger and older babies in the Baby Check scoring system. In current practice, which is aligned with the NICE guideline on fever in under 5s (CG160) traffic light system for identifying risk of serious illness, a lower fever in a younger baby would be more of a concern than an older baby with a higher fever. Potential implications of this would be that younger babies with a low-grade fever score lower than their actual risk might be and therefore deemed well when they may be unwell and need monitoring or treatment.

The committee also discussed the changes in wording between the original Baby Check scoring system and the Lullaby Trust's Baby Check mobile app and pdf booklet version, for example:

- "frank blood mixed with the baby's stools" in the Baby Check scoring system was replaced with "large amounts of obvious blood in your baby's nappy (not just on the stool"
- "Is the baby's muscle tone reduced?" in the Baby Check scoring system was replaced with "Is your baby more floppy than usual?"
- "Is the baby concentrating on you less that you would expect?" in the Baby Check scoring system was replaced with "Is your baby watching you less than usual?"

The committee acknowledged the differences but agreed that these represent small changes which make the scoring more user friendly and understandable for 'lay people', without altering the original, intended meaning. This did however provoke discussion about potential future revisions to the app or the booklet, which might create further differences between them and the original Baby Check scoring system. The committee agreed any such changes would need to be carefully considered.

The committee also discussed the poor sensitivity of the Baby Check scoring system in identifying babies that are moderately unwell. They agreed that it is very difficult to diagnose a moderately unwell baby and they were therefore unsurprised that there was very poor correlation between the scoring system and diagnosis. The committee emphasised that the baby check scoring system was most useful in identifying well babies, thus wrote recommendations based on the evidence around this health state.

The evidence from the theoretical community cohort (Cole 1991; Morley 1991) was not considered particularly helpful as the data was based on assumptions from a proportion of babies in the study. The committee were therefore not sufficiently confident in the accuracy of the data to base recommendations on this data, not least given the consequences of failing to identify a seriously ill baby.

Cost effectiveness and resource use

No economic evidence is available for this review question. The committee agreed that providing parents with information on the Baby Check scoring system is likely to have small resource implications relating to the health professional's time. The scoring system is freely available. Its use may lead to benefits for the babies and their parents and cost-savings to the health service, if illness is identified and treated earlier, resulting in need for less intensive intervention and lower mortality and morbidity for the baby. Therefore, the committee agreed that the recommendations ensure efficient use of healthcare resources.

References

Chandran 1998

Chandran, S., Sunita, K., Nair, A. K., et al. A trial of baby check scoring system to identify high-risk infants in a polyclinic in Oman. Journal of Tropical Pediatrics 1998; 44: 218-221

Chen 1997

Chen, C. K., Chen, S. J., Hwang, B. Evaluation of the severity of illness in infants by the Baby Check Score. Chinese medical journal; Free China ed 1997; 59:.15-20

Cole 1991 and Morley 1991

Cole TJ, Morley CJ, Thornton AJ, et al. A scoring system to quantify illness in babies under 6 months of age. Royal Statistical Society 1991; 2: 287-304

Morley CJ, Thornton AJ, Cole TJ, et al. Baby check: a scoring system to grade the severity of acute systemic illness in babies under 6 months old. Arch of Dis in Child 1991; 66: 100-106

Thornton 1991

Thornton AJ, Morley CJ, Cole TJ, et al. Field trials of the Baby Check score card in hospital. Arch of Dis in Child1991; 66: 115-120

Appendices

Appendix A – Review protocol

Review protocol for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

Table 3: Review protocol

Field (based on PRISMA-P)	Content
Review question	Which scoring systems are accurate in identifying or predicting illness severity in babies?
Type of review question	Clinical prediction model review
Objective of the review	To determine if a scoring system can accurately assess illness severity in babies.
Eligibility criteria – population	Exclude studies with a specific population of babies who were born pre-term. This means babies born before 37 weeks since 'term' is considered to be between 37 and 42 weeks of pregnancy. For studies with a mixed population, they will be included if at least 66% of babies are born at term. Exclude studies specifically focused on babies in which fever was an entry criterion. Exclude babies in neonatal units when signs and symptoms occur. Exclude studies focused on babies with a major underlying morbidity (e.g. congenital heart disease).
Eligibility criteria – index tests /clinical prediction model	A validated scoring system based on a combination of symptoms and/or signs for babies within the first 8 weeks after birth used by healthcare professionals or parents, either face-to-face or remotely.
Eligibility criteria – outcome to be modelled	• Well/mildly unwell for example defined as no clinical intervention needed • Moderately unwell for example defined as requiring clinical attention

	• Seriously unwell for example defined as admission to hospital or treatment in hospital Severity of illness (or absence of) will be defined by a qualified assessor through a comprehensive assessment Exclude studies specifically focused on infection in babies with onset in the first 72 hours after birth. Exclude studies focused on a specific disorder already covered by separate NICE guidelines (sepsis, bacterial meningitis and meningococcal septicaemia, early onset neonatal infection, urinary tract infection, gastro-oesophageal reflux disease).
Confounding factors for prognostic estimates	Analysis should adjust for important confounding factors. Multivariate analysis should be used for clinical prediction models
Outcomes and prioritisation	<u>Model performance:</u> **Critical outcomes:** • Calibration • Discrimination (AUC/C-statistic) <u>Accuracy of prediction:</u> **Critical outcomes:** • Sensitivity **Important outcomes:** • Specificity • Positive likelihood ratio • Negative likelihood ratio
Eligibility criteria – study design	Include published full text papers: • systematic reviews

	• cohort studies (prospective cohort studies will be prioritised over retrospective cohort studies. If insufficient data for decision making is available from prospective cohort studies, then retrospective cohort studies will be considered). • cross-sectional studies Exclude: • conference abstracts
Other inclusion exclusion criteria	**Inclusion** • English-language Studies from low- and middle-income countries, as defined by the World Bank, will be excluded, as the configuration of antenatal and postnatal services in these countries might not be representative of that in the UK. • Studies published from 1990
Proposed sensitivity/sub-group analysis, or meta-regression	Scoring system used in different settings, by different assessors (for example healthcare professionals or parents), or different versions of the scoring system will be analysed separately.
Selection process – duplicate screening/selection/analysis	Sifting, data extraction and appraisal of methodological quality will be performed by the systematic reviewer. Any disputes will be resolved in discussion with the senior systematic reviewer and the Topic Advisor. Quality control will be performed by the senior systematic reviewer. This review question was not prioritised for health economic analysis and so no formal dual weeding, study selection (inclusion/exclusion) or data extraction into evidence tables will be undertaken. (However, internal (NGA) quality assurance processes will include consideration of the outcomes of weeding, study selection and data extraction and the committee will review the results of study selection and data extraction).
Data management (software)	NGA STAR software will be used for study sifting, data extraction, recording quality assessment using checklists and generating bibliographies/citations.

	For the diagnostic component of the review, a modified 'GRADE' method will be used to assess the quality of evidence for each index test. RevMan v.5, STATA and WinBUGS software will be used for data analysis, as appropriate. This will be described in the separate methods chapter for the guideline.
Information sources – databases and dates	The following databases will be searched: • CDSR • DARE • Embase • EMCare • HTA Database • MEDLINE and MEDLINE IN-PROCESS Searches will be restricted by: • English language • human studies • observational studies • systematic reviews. Other searches: • inclusion lists of systematic reviews.
Identify if an update	This is an update. However, the review and drafting of recommendations are being completed afresh. The 2006 version of the postnatal care guideline included these recommendations: 1.4.1 Healthy babies should have normal colour for their ethnicity, maintain a stable body temperature, and pass urine and stools at regular intervals. They initiate feeds, suck well on the breast (or bottle) and settle between feeds. They are not excessively irritable, tense, sleepy or floppy. The vital signs of a healthy baby should fall within the following ranges: • respiratory rate normally 30−60 breaths per minute

	• heart rate normally between 100 and 160 beats per minute in a newborn • temperature in a normal room environment of around 37°C (if measured). [2006] 1.4.2 At each postnatal contact, parents should be offered information and advice to enable them to: • assess their baby's general condition • identify signs and symptoms of common health problems seen in babies • contact a healthcare professional or emergency service if required. [2006]
Author contacts	National Guideline Alliance https://www.nice.org.uk/guidance/indevelopment/gid-ng10070
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual 2014
Search strategy – for one database	For details please see appendix B of the full guideline
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix D (clinical evidence tables) or H (economic evidence tables) of the full guideline.
Data items – define all variables to be collected	For details please see evidence tables in appendix D (clinical evidence tables) or H (economic evidence tables) of the full guideline.
Methods for assessing bias at outcome/study level	Standard study checklists were used to critically appraise individual studies. For details please see section 6.2 of Developing NICE guidelines: the manual 2014 For the diagnostic component of the review the risk of bias across all available evidence was evaluated for each outcome 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/
Criteria for quantitative synthesis (where suitable)	For details please see section 6.4 of Developing NICE guidelines: the manual 2014 Minimum important differences Default values will be used of: Sensitivity and specificity high when ≥ 90% Sensitivity and specificity moderate when between 75 and 89%

	Good model performance will be defined as AUC > 0.75 and O:E ratio between 0.8 and 1.2 (as suggested by Debray 2017), unless more appropriate values are identified by the guideline committee or in the literature.
Methods for analysis – combining studies and exploring (in)consistency	For a full description of methods see Supplement 1.
Meta-bias assessment – publication bias, selective reporting bias	For details please see section 6.2 of Developing NICE guidelines: the manual 2014
Assessment of confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual 2014
Rationale/context – Current management	For details please see the introduction to the evidence review
Describe contributions of authors and guarantor	A multidisciplinary committee developed the guideline. The committee was convened by The National Guideline Alliance and chaired by Dr David Jewell in line with section 3 of Developing NICE guidelines: the manual 2014 Staff from The National Guideline Alliance undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the guideline in collaboration with the committee. For a full description of methods see Supplement 1.
Sources of funding/support	The National Guideline Alliance is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists
Name of sponsor	The National Guideline Alliance is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists
Roles of sponsor	NICE funds The National Guideline Alliance to develop guidelines for those working in the NHS, public health, and social care in England
PROSPERO registration number	This protocol has not been registered in PROSPERO

AUC: Area Under Curve; C-statistic: concordance statistic; GRADE: Grading of Recommendations Assessment, Development and Evaluation; NGA: National Guideline Alliance; NICE: National Institute for Health and Care Excellence; O:E: Observed to Expected ratio; PROSPERO: Prospective Register for Systematic Reviews;: Preferred Reporting Items for Systematic and Meta-analysis Protocols; QUADAS: quality assessment of diagnostic accuracy studies;

Appendix B – Literature search strategies

Literature search strategies for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

Clinical search

The search for this topic was last run on 26th May 2020.

Database: Emcare, Embase, Medline, Medline Ahead of Print and In-Process & Other Non-Indexed Citations – OVID [Multifile]

#	Search
1	"area under the curve"/ or instrument validation/ or performance/ or predictive validity/ or predictive value/ or receiver operating characteristic/ or reliability/ or reproducibility/ or "sensitivity and specificity"/ or test retest reliability/ or validity/
2	1 use emez, emcr
3	"area under curve"/ or "predictive value of tests"/ or "reproducibility of results"/ or roc curve/ or "sensitivity and specificity"/ or validation studies/
4	3 use ppez
5	(accurac* or accurat* or area under curve or auc value* or (likelihood adj3 ratio*) or (diagnostic adj2 odds ratio*) or ((pretest or pre test or posttest or post test) adj2 probabilit*) or (predict* adj3 value*) or receiver operating characteristic or (roc adj2 curv*) or reliabil* or sensititiv* or specificit* or valid*).tw.
6	(calibration or discrimination).ti,ab.
7	(or/2,4-6) or diagnostic value.sh.
8	disease severity/ use emez, emcr or "severity of illness index"/ use ppez or (((assess* or illness* or sickness*) adj5 sever*) or ((grad* or scor* or quantif*) adj3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) adj3 illness*) or sign* or symptom* or complain* or (clinical adj3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting adj3 (feature* or finding* or factor*)) or presentation* or (physical adj3 (manifestation* or characteristic* or feature* or finding*)))))).ti,ab.
9	(index or scale* or score* or scoring* or test* or tool*).ti,ab,hw.
10	baby/ use emez, emcr or newborn/ use emez, emcr or exp infant, newborn/ use ppez or (babies or baby or infant* or neonat* or newborn* or new born*).ti,ab.
11	7 and 8 and 9 and 10
12	(baby check*).ti,ab.
13	((index or scale* or score* or scoring or test* or tool*) adj5 (assess* or grad* or score* or scoring or quantif*) adj3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) adj3 illness*) or sign* or symptom* or complain* or (clinical adj3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting adj3 (feature* or finding* or factor*)) or presentation* or (physical adj3 (manifestation* or characteristic* or feature* or finding*))))).ti,ab. and 7 and 10
14	((index or scale* or score* or scoring or test* or tool*) adj5 (assess* or grad* or score* or scoring or quantif*) adj3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) adj3 illness*) or sign* or symptom* or complain* or (clinical adj3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting adj3 (feature* or finding* or factor*)) or presentation* or (physical adj3 (manifestation* or characteristic* or feature* or finding*))))).ti,ab. and (babies or baby or infant* or neonat* or newborn* or new born*).ti.
15	(((scor* adj (card* or system*)) and (babies or baby or infant* or neonat* or newborn* or new born*) and (((assess* or illness* or sickness*) adj5 sever*) or ((grad* or scor* or

#	Search
	quantif*) adj3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) adj3 illness*) or sign* or symptom* or complain* or (clinical adj3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting adj3 (feature* or finding* or factor*)) or presentation* or (physical adj3 (manifestation* or characteristic* or feature* or finding*))))))) or (scor* adj (card* or system*) adj5 (babies or baby or infant* or neonat* or newborn* or new born*) adj5 (accurac* or accurat* or area under curve or auc value* or (likelihood adj3 ratio*) or (diagnostic adj2 odds ratio*) or ((pretest or pre test or posttest or post test) adj2 probabilit*) or (predict* adj3 value*) or receiver operating characteristic or (roc adj2 curv*) or reliabil* or sensititiv* or specificit* or valid*))).ti,ab.
16	((scor* adj (card* or system*)) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or new born*) and (((assess* or illness* or sickness*) adj5 sever*) or ((grad* or scor* or quantif*) adj3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or sign* or symptom* or complain* or (clinical adj3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting adj3 (feature* or finding* or factor*)) or presentation* or (physical adj3 (manifestation* or characteristic* or feature* or finding*))))))).ti,ab.
17	((scor* adj (card* or system*)) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or new born*) and (accurac* or accurat* or area under curve or auc value* or (likelihood adj3 ratio*) or (diagnostic adj2 odds ratio*) or ((pretest or pre test or posttest or post test) adj2 probabilit*) or (predict* adj3 value*) or receiver operating characteristic or (roc adj2 curv*) or reliabil* or sensititiv* or specificit* or valid*)).ti,ab.
18	((index or scale* or test* or tool*) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or new born*) and (accurac* or accurat* or area under curve or auc value* or (likelihood adj3 ratio*) or (diagnostic adj2 odds ratio*) or ((pretest or pre test or posttest or post test) adj2 probabilit*) or (predict* adj3 value*) or receiver operating characteristic or (roc adj2 curv*) or reliabil* or sensititiv* or specificit* or valid*) and (((assess* or illness* or sickness*) adj5 sever*) or ((grad* or scor* or quantif*) adj3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or sign* or symptom* or complain* or (clinical adj3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting adj3 (feature* or finding* or factor*)) or presentation* or (physical adj3 (manifestation* or characteristic* or feature* or finding*))))))).ti,ab.
19	or/11-18
20	(((letter/ or editorial/ or news/ or exp historical article/ or anecdotes as topic/ or comment/ or case report/ or (letter or comment*).ti.) not (case control* or cohort* or cross sectional* or follow* up* or longitudinal* or metaanal* or meta anal* or observational* or prospective* or random* or retrospective* or systematic review*).sh,pt,ti,ab.) or (animals not humans).sh. or exp animals, laboratory/ or exp animal experimentation/ or exp models, animal/ or exp rodentia/ or (rat or rats or mouse or mice).ti.) use ppez
21	20 use ppez
22	(((letter.pt. or letter/ or note.pt. or editorial.pt. or case report/ or case study/ or (letter or comment*).ti.) not (case control* or cohort* or cross sectional* or follow* up* or longitudinal* or metaanal* or meta anal* or observational* or prospective* or random* or retrospective* or systematic review*).sh,pt,ti,ab.) or ((animal/ not human/) or nonhuman/ or exp animal experiment/ or exp experimental animal/ or animal model/ or exp rodent/ or (rat or rats or mouse or mice).ti.)) use emez, emcr
23	22 use emez, emcr
24	or/21,23
25	19 not 24
26	limit 25 to (conference abstract or conference paper or conference review or conference proceeding)
27	26 use emez, emcr
28	25 not 27

#	Search
29	28
30	limit 29 to english language
31	limit 30 to yr="1990 -current"

Database: CDSR (global) [Wiley]

#	Search
#1	MeSH descriptor: [Area Under Curve] this term only
#2	MeSH descriptor: [Predictive Value of Tests] this term only
#3	MeSH descriptor: [Reproducibility of Results] this term only
#4	MeSH descriptor: [Sensitivity and Specificity] this term only
#5	MeSH descriptor: [Validation Study] this term only
#6	MeSH descriptor: [ROC Curve] this term only
#7	((calibration or discrimination)):ti,ab,kw
#8	((accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near/3 ratio*) or (diagnostic near/2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near/2 probabilit*) or (predict* near/3 value*) or “receiver operating characteristic” or (roc near/2 curv*) or reliabil* or sensititiv* or specificit* or valid*)):ti,ab,kw
#9	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10	MeSH descriptor: [Severity of Illness Index] this term only
#11	((((assess* or illness* or sickness*) near/5 sever*) or ((grad* or scor* or quantif*) near/3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near/3 illness*) or sign* or symptom* or complain* or (clinical near/3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near/3 (feature* or finding* or factor*)) or presentation* or (physical near/3 (manifestation* or characteristic* or feature* or finding*))))):ti,ab,kw
#12	#10 or #11
#13	((index or scale* or score* or scoring* or test* or tool*)):ti,ab,kw
#14	MeSH descriptor: [Infant, Newborn] explode all trees
#15	((babies or baby or infant* or neonat* or newborn* or “new born*”)):ti,ab,kw
#16	#14 or #15
#17	#9 and #12 and #13 and #16
#18	(“baby check*”):ti,ab,kw
#19	(((index or scale* or score* or scoring or test* or tool*) near/5 (assess* or grad* or score* or scoring or quantif*) near/3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near/3 illness*) or sign* or symptom* or complain* or (clinical near/3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near/3 (feature* or finding* or factor*)) or presentation* or (physical near/3 (manifestation* or characteristic* or feature* or finding*))))):ti,ab,kw
#20	#19 and #9 and #16
#21	(((index or scale* or score* or scoring or test* or tool*) near/5 (assess* or grad* or score* or scoring or quantif*) near/3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near/3 illness*) or sign* or symptom* or complain* or (clinical near/3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near/3 (feature* or finding* or factor*)) or presentation* or (physical near/3 (manifestation* or characteristic* or feature* or finding*)))):ti,ab,kw and (babies or baby or infant* or neonat* or newborn* or “new born*”)):ti
#22	((((scor* near/1 (card* or system*)) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (((assess* or illness* or sickness*) near/5 sever*) or ((grad* or scor* or quantif*) near/3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near/3 illness*) or sign* or symptom* or complain* or (clinical near/3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting

#	Search
	near/3 (feature* or finding* or factor*)) or presentation* or (physical near/3 (manifestation* or characteristic* or feature* or finding*)))))) or (scor* near/1 (card* or system*) near/5 (babies or baby or infant* or neonat* or newborn* or “new born*”) near/5 (accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near/3 ratio*) or (diagnostic near/2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near/2 probabilit*) or (predict* near/3 value*) or “receiver operating characteristic” or (roc near/2 curv*) or reliabil* or sensititiv* or specificit* or valid*)))):ti,ab,kw
#23	(((scor* near/1 (card* or system*)) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (((assess* or illness* or sickness*) near/5 sever*) or ((grad* or scor* or quantif*) near/3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or sign* or symptom* or complain* or (clinical near/3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near/3 (feature* or finding* or factor*)) or presentation* or (physical near/3 (manifestation* or characteristic* or feature* or finding*))))))):ti,ab,kw
#24	(((scor* near/1 (card* or system*)) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near/3 ratio*) or (diagnostic near/2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near/2 probabilit*) or (predict* near/3 value*) or “receiver operating characteristic” or (roc near/2 curv*) or reliabil* or sensititiv* or specificit* or valid*))):ti,ab,kw
#25	(((index or scale* or test* or tool*) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near/3 ratio*) or (diagnostic near/2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near/2 probabilit*) or (predict* near/3 value*) or “receiver operating characteristic“ or (roc near/2 curv*) or reliabil* or sensititiv* or specificit* or valid*) and (((assess* or illness* or sickness*) near/5 sever*) or ((grad* or scor* or quantif*) near/3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or sign* or symptom* or complain* or (clinical near/3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near/3 (feature* or finding* or factor*)) or presentation* or (physical near/3 (manifestation* or characteristic* or feature* or finding*))))))):ti,ab,kw
#26	#17 or #18 or #20 or #21 or #22 or #23 or #24 or #25

Database: DARE, HTA (global) [CRD Web]

#	Search
#1	MeSH descriptor Area Under Curve in dare,hta
#2	MeSH descriptor Predictive Value of Tests in dare,hta
#3	MeSH descriptor Reproducibility of Results in dare,hta
#4	MeSH descriptor Sensitivity and Specificity in dare,hta
#5	MeSH descriptor Validation Study in dare,hta
#6	MeSH descriptor ROC Curve in dare,hta
#7	((calibration or discrimination)) in dare, hta
#8	((accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near3 ratio*) or (diagnostic near2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near2 probabilit*) or (predict* near3 value*) or “receiver operating characteristic” or (roc near2 curv*) or reliabil* or sensititiv* or specificit* or valid*)) in dare, hta
#9	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10	MeSH descriptor Severity of Illness Index in dare,hta
#11	((((assess* or illness* or sickness*) near5 sever*) or ((grad* or scor* or quantif*) near3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near3 illness*) or sign* or symptom* or

#	Search
	complain* or (clinical near3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near3 (feature* or finding* or factor*)) or presentation* or (physical near3 (manifestation* or characteristic* or feature* or finding*))))))) in dare, hta
#12	#10 or #11
#13	((index or scale* or score* or scoring* or test* or tool*)) in dare, hta
#14	MeSH descriptor Infant, Newborn explode all trees in dare,hta
#15	((babies or baby or infant* or neonat* or newborn* or “new born*”)) in dare, hta
#16	#14 or #15
#17	#9 and #12 and #13 and #16
#18	(“baby check*”) in dare, hta
#19	(((index or scale* or score* or scoring or test* or tool*) near5 (assess* or grad* or score* or scoring or quantif*) near3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near3 illness*) or sign* or symptom* or complain* or (clinical near3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near3 (feature* or finding* or factor*)) or presentation* or (physical near3 (manifestation* or characteristic* or feature* or finding*))))) in dare, hta
#20	#19 and #9 and #16
#21	(((index or scale* or score* or scoring or test* or tool*) near5 (assess* or grad* or score* or scoring or quantif*) near3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near3 illness*) or sign* or symptom* or complain* or (clinical near3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near3 (feature* or finding* or factor*)) or presentation* or (physical near3 (manifestation* or characteristic* or feature* or finding*)))) and (babies or baby or infant* or neonat* or newborn* or “new born*”))
#22	((((scor* near (card* or system*)) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (((assess* or illness* or sickness*) near5 sever*) or ((grad* or scor* or quantif*) near3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or ((grad* or scor* or quantif*) near3 illness*) or sign* or symptom* or complain* or (clinical near3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near3 (feature* or finding* or factor*)) or presentation* or (physical near3 (manifestation* or characteristic* or feature* or finding*)))))) or (scor* near (card* or system*) near5 (babies or baby or infant* or neonat* or newborn* or “new born*”) near5 (accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near3 ratio*) or (diagnostic near2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near2 probabilit*) or (predict* near3 value*) or “receiver operating characteristic” or (roc near2 curv*) or reliabil* or sensititiv* or specificit* or valid*)))) in dare, hta
#23	(((scor* near1 (card* or system*)) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (((assess* or illness* or sickness*) near5 sever*) or ((grad* or scor* or quantif*) near3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or sign* or symptom* or complain* or (clinical near3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near3 (feature* or finding* or factor*)) or presentation* or (physical near3 (manifestation* or characteristic* or feature* or finding*))))))) in dare, hta
#24	(((scor* near (card* or system*)) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near3 ratio*) or (diagnostic near2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near2 probabilit*) or (predict* near3 value*) or “receiver operating characteristic” or (roc near2 curv*) or reliabil* or sensititiv* or specificit* or valid*))) in dare, hta

#	Search
#25	(((index or scale* or test* or tool*) and (identif* or predict*) and (babies or baby or infant* or neonat* or newborn* or “new born*”) and (accurac* or accurat* or “area under curve” or “auc value*” or (likelihood near3 ratio*) or (diagnostic near2 odds ratio*) or ((pretest or “pre test” or posttest or “post test”) near2 probabilit*) or (predict* near3 value*) or “receiver operating characteristic“ or (roc near2 curv*) or reliabil* or sensititiv* or specificit* or valid*) and (((assess* or illness* or sickness*) near5 sever*) or ((grad* or scor* or quantif*) near3 (disease* or disorder* or infection* or ill* or morbidit* or mortalit* or sick* or unwell* or sign* or symptom* or complain* or (clinical near3 (manifestation* or feature* or finding* or aspect* or marker*)) or (presenting near3 (feature* or finding* or factor*)) or presentation* or (physical near3 (manifestation* or characteristic* or feature* or finding*))))))) in dare, hta
#26	#17 or #18 or #20 or #21 or #22 or #23 or #24 or #25

Health economic search

The search for this topic was last run on 5th December 2019.

Database: Emcare, Embase, Medline, Medline Ahead of Print and In-Process & Other Non-Indexed Citations (global) – OVID [Multifile]

#	Search
1	puerperium/ or perinatal period/ or postnatal care/
2	1 use emczd, emcr
3	postpartum period/ or peripartum period/ or postnatal care/
4	3 use ppez
5	(nullipara* or peri natal* or perinatal* or postbirth or post birth or postdelivery or post delivery or postnatal* or post natal* or postpartum* or post partum* or primipara* or puerpera* or puerperium* or ((after or follow*) adj2 birth*)).ti,ab.
6	or/2,4-5
7	breast feeding/ or breast feeding education/ or lactation/
8	7 use emczd, emcr
9	exp breast feeding/ or lactation/
10	9 use ppez
11	(breastfeed* or breast feed* or breastfed* or breastfeed* or breast fed or breastmilk or breast milk or expressed milk* or lactat* or (nursing adj (baby or infant* or mother* or neonate* or newborn*))).ti,ab.
12	or/8,10-11
13	artificial food/ or bottle feeding/ or infant feeding/
14	13 use emczd, emcr
15	bottle feeding/ or infant formula/
16	15 use ppez
17	(((bottle or formula or synthetic) adj2 (artificial or fed or feed* or infant* or milk*)) or (artificial adj (formula or milk)) or bottlefed or bottlefeed or cup feeding or (milk adj2 (substitut* or supplement*)) or ((infant or milk or water or glucose or dextrose or formula) adj supplement) or formula supplement* or supplement feed or milk feed or ((baby or babies or infant* or neonate* or newborn*) adj (formula* or milk)) or formulafeed or formulated or (milk adj2 powder*) or hydrolyzed formula* or (((feeding or baby or infant) adj bottle*) or infant feeding or bottle nipple* or milk pump*)).ti,ab.
18	or/14,16-17

#	Search
19	or/6,12,18
20	budget/ or exp economic evaluation/ or exp fee/ or funding/ or exp health care cost/ or health economics/
21	20 use emczd, emcr
22	exp budgets/ or exp "costs and cost analysis"/ or economics/ or exp economics, hospital/ or exp economics, medical/ or economics, nursing/ or economics, pharmaceutical/ or exp "fees and charges"/ or value of life/
23	22 use ppez
24	budget*.ti,ab. or cost*.ti. or (economic* or pharmaco?economic*).ti. or (price* or pricing*).ti,ab. or (cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab. or (financ* or fee or fees).ti,ab. or (value adj2 (money or monetary)).ti,ab.
25	or/21,23-24
26	economic model/ or quality adjusted life year/ or "quality of life index"/
27	(cost-benefit analysis.sh. and (cost-effectiveness ratio* and (perspective* or life expectanc*)).tw.)
28	((quality of life or qol).tw. and cost benefit analysis.sh.)
29	or/26-28 use emczd, emcr
30	models, economic/ or quality-adjusted life years/
31	(cost-benefit analysis.sh. and (cost-effectiveness ratio* and (perspective* or life expectanc*)).tw.)
32	((quality of life or qol).tw. and cost-benefit analysis.sh.)
33	or/30-32 use ppez
34	(eq-5d* or eq5d* or eq-5* or eq5* or euroqual* or euro qual* or euroqual 5d* or euro qual 5d* or euro qol* or euroqol*or euro quol* or euroquol* or euro quol5d* or euroquol5d* or eur qol* or eurqol* or eur qol5d* or eurqol5d* or eur?qul* or eur?qul5d* or euro* quality of life or european qol).tw.
35	(euro* adj3 (5 d* or 5d* or 5 dimension* or 5dimension* or 5 domain* or 5domain*)).tw.
36	(hui or hui2 or hui3).tw.
37	(illness state* or health state*).tw.
38	(multiattibute* or multi attribute*).tw.
39	(qaly* or qal or qald* or qale* or qtime* or qwb* or daly).tw.
40	(quality adjusted or quality adjusted life year*).tw.
41	(sf36 or sf 36 or sf thirty six or sf thirtysix).tw.
42	sickness impact profile.sh.
43	(time trade off*1 or time tradeoff*1 or tto or timetradeoff*1).tw.
44	(utilit* adj3 (score*1 or valu* or health* or cost* or measur* or disease* or mean or gain or gains or index*)).tw.
45	utilities.tw.
46	((qol or hrqol or quality of life).tw. or *quality of life/) and ((qol or hrqol* or quality of life) adj2 (change*1 or declin* or decreas* or deteriorat* or effect or effects or high* or impact*1 or impacted or improve* or increas* or low* or reduc* or score or scores or worse)).ab.
47	quality of life.sh. and ((health-related quality of life or (health adj3 status) or ((quality of life or qol) adj3 (chang* or improv*)) or ((quality of life or qol) adj (measure*1 or score*1))).tw. or (quality of life or qol).ti. or ec.fs.)
48	or/29,33-47

#	Search
49	or/25,48
50	19 and 50
51	limit 50 to english language
52	(animals/ not humans/) or exp animals, laboratory/ or exp animal experimentation/ or exp models, animal/ or exp rodentia/
53	52 use ppez
54	(animal/ not human/) or nonhuman/ or exp animal experiment/ or exp experimental animal/ or animal model/ or exp rodent/
55	54 use emczd, emcr
56	(rat or rats or mouse or mice).ti.
57	or/53,55-56
58	51 not 57

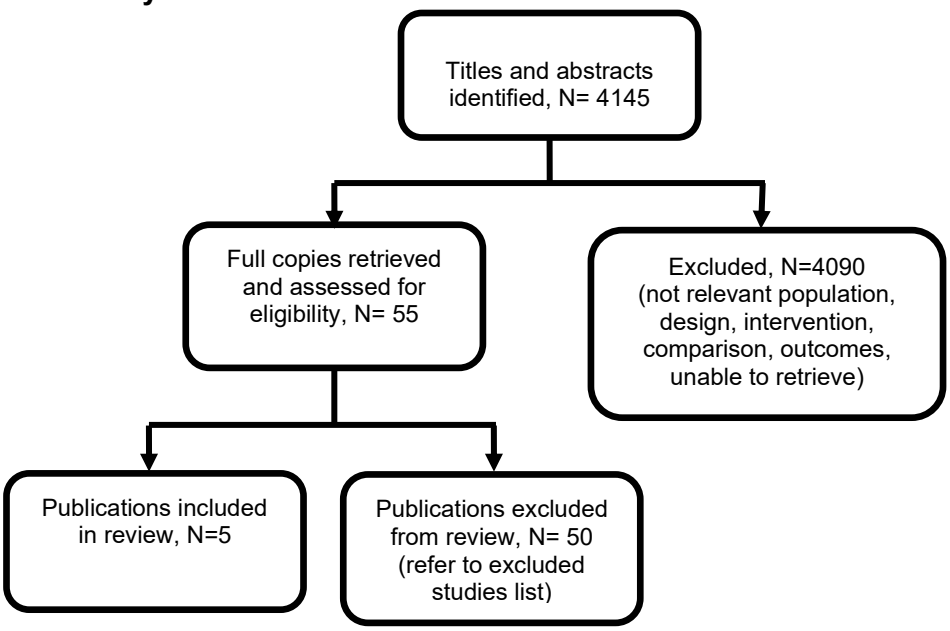
Database: HTA, NHS EED (global) [CRD Web]

#	Search
1	mesh descriptor postpartum period in hta, nhs eed
2	mesh descriptor peripartum period in hta, nhs eed
3	mesh descriptor postnatal care hta, nhs eed
4	(nullipara* or peri natal* or perinatal* or postbirth or post birth or postdelivery or post delivery or postnatal* or post natal* or postpartum* or post partum* or primipara* or puerpera* or puerperium* or ((after or follow*) near2 birth*)) hta, nhs eed
5	#1 or #2 or #3 or #4
6	mesh descriptor breast feeding explode all trees hta, nhs eed
7	mesh descriptor lactation hta, nhs eed
8	(breastfeed* or breast feed* or breastfed* or breastfeed* or breast fed or breastmilk or breast milk or expressed milk* or lactat* or (nursing next (baby or infant* or mother* or neonate* or newborn*))) hta, nhs eed
9	#6 or #7 or #8
10	mesh descriptor bottle feeding hta, nhs eed
11	mesh descriptor infant formula hta, nhs eed
12	(((bottle or formula or synthetic) near2 (artificial or fed or feed* or infant* or milk*)) or (artificial next (formula or milk)) or bottlefed or bottlefeed or cup feeding or (milk near2 (substitut* or supplement*)) or ((infant or milk or water or glucose or dextrose or formula) next supplement) or formula supplement* or supplement feed or milk feed or ((baby or babies or infant* or neonate* or newborn*) next (formula* or milk)) or formulafeed or formulated or (milk near2 powder*) or hydrolyzed formula* or (((feeding or baby or infant) next bottle*) or infant feeding or bottle nipple* or milk pump*)) hta, nhs eed
13	#10 or #11 or #12
14	#5 or #9 or #13

Appendix C – Clinical evidence study selection

Clinical study selection for: Which scoring systems are accurate in identifying or predicting illness severity in babies?

Figure 1: Study selection



Appendix D – Clinical evidence tables

Evidence table for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

Table 4: Evidence table

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td>Full citation

Morley, C. J., Thornton, A. J., Cole, T. J., Hewson, P. H., Fowler, M. A., Baby Check: A scoring system to grade the severity of acute systemic illness in babies under 6 months old, Archives of Disease in Childhood, 66, 100-106, 1991

Ref Id

1255710

Country/ies where the study was carried out

Australia and UK

Study type</td>
<td>Sample size
N=1007
n=298 assessed at home;
n=709 assessed in hospital

Characteristics
Of the babies seen at home: 290/298 (98%) well or mildly ill; None considered to be seriously ill.
Of the babies seen in hospital:
165/709 (23%) were seriously ill; 239/709 (34%) moderately ill; 305/709 (43%) mildly ill or well.
Babies had a wide range of conditions; upper respiratory tract infection (n=81); lower respiratory tract infection (n=135); diarrhoea and vomiting (n=64); feeding</td>
<td>Tests
<u>Baby check scoring system</u>
7 symptoms and 12 signs that in combination graded illness severity best. 2 parallel version designed, one for healthcare professionals and the other for parents. Score groups are interpreted for parents as:
Score 0-7 - 'Your baby is well or only a little unwell and is not likely to need medical attention at the moment'.
Score 8-12 - 'Your baby is unwell, but is not likely to be seriously ill. Contact your Dr, health visitor, or midwife for advice. Watch your baby closely, if you</td>
<td>Methods
<u>Sample size and enrolment</u>
For accurate determination of the symptoms and signs associated with serious illness requires the rarest to be recorded at least 5 times. A study in the community would need to enrol around 30 000 babies to ensure this. The only practical way to collect data from seriously ill babies would be to enrol them when they presented to hospital, where the incidence of serious illness is much higher than at home. At least 600 babies were needed to ensure that the rarest symptoms and signs were recorded at least 5 times. At home, 300 babies were sufficient to quantify the incidence of</td>
<td>Results
<u>Babies seen at home</u>
Outcome: Seriously ill (defined as a score of 13 or more)
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment - ve</th><th>Total</th></tr>
<tr><td>Scoring system +ve</td><td>0</td><td>3</td><td>3</td></tr>
<tr><td>Scoring system - ve</td><td>0</td><td>295</td><td>295</td></tr>
<tr><td>Total</td><td>0</td><td>298</td><td>298</td></tr>
</table>
Sensitivity: not calculable
Specificity: 98.99% (95% CI 97.1 to 9.8%)*
Positive likelihood ratio: not calculable
Negative likelihood ratio: not calculable
Prevalence of seriously ill babies: 0%*</td>
<td>Limitations
<u>Babies assessed at home</u>
(assessed using QUADAS-II for diagnostic accuracy studies)

Patient selection
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? Yes
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes
Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: HIGH

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes, Baby check</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td>Prospective cohort study

Aim of the study
Aim of the study was to grade systemic illness in babies under 6 months’ old

Study dates
April 1986- April 1987

Source of funding
Baby Illness Research Project Appeal of the Foundation for the Study of Infant Deaths, the Australian Institute of Health, the Ross Trust, Felton Bequests, the H L Hecht Trust, the Percy Baxter Charitable Trust, and A Williams Private Fund.</td>
<td>problems (n=20); apnoea (n=16); colic (n=27); intussusception (n=11); meningitis (n=10); urinary infection (n=16); eczema and dermatitis (n=92).

Inclusion Criteria
<u>Babies seen at home</u>
Full term babies from 0-25 weeks old in the community (Cambridge)
<u>Babies seen in hospital</u>
Full term babies under 26 weeks’ old presenting to hospital (Majority recruited in Melbourne)

Exclusion Criteria
not reported</td>
<td>think your baby is getting worse to the score again'.
Score 13-19 - 'Your baby is ill and needs to be seen by a Dr. Contact your Dr now and arrange for your baby to be seen'.
Score 20 or more - 'Your baby maybe seriously ill and needs to be seen by a Dr straight away'.

For further details of the Baby Check scoring system please see table 2 in the original paper.

<u>Grading of systemic illness</u>
No 'gold standard' for grading of systemic illness, assessors subjectively graded babies into categories of well, mildly ill, moderately ill, seriously ill. In hospital, each baby's illness was also graded, where possible, using other criteria: positive investigation results (n=200), a review of the notes by 3</td>
<td>all but the rarest symptoms and signs. During busy times, the paediatricians biased enrolment towards the more ill of the babies presented, this ensured a wide spectrum of illness in the babies taking part in the study.
<u>The assessment of baby</u>
Mother was asked 28 predefined symptoms, their duration and her impression of the severity. Only symptoms present <24 hours were recorded. Some were clarified with additional questions. The baby was then examined for 47 physical signs. 2 observers used an identical history questionnaire and examination procedure and practiced the assessment to minimise interobserver error. At the end of the study they compared notes to look for systematic differences. As a result 2 signs, mottling of the skin and mucousy breathing, were found to have been assessed differently and were excluded from further analyses.</td>
<td>Outcome: Moderately ill (defined as a score of 8-12)
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr>
<tr><td>Scoring system +ve</td><td>3</td><td>3</td><td>6</td></tr>
<tr><td>Scoring system -ve</td><td>5</td><td>287</td><td>292</td></tr>
<tr><td>Total</td><td>8</td><td>290</td><td>298</td></tr>
</table>
Sensitivity: 37.5% (95% CI 8.5% to 75.5%)*
Specificity: 98.97% (95% CI 97.0% to 99.8%)*
Positive likelihood ratio: 36.25 (95% CI 8.61 to 152.68)*
Negative likelihood ratio: 0.63 (95% CI 0.37 to 1.08)*
Prevalence of moderately ill babies: 2.7%

Outcome: well or mildly ill (defined as a score of 0-7)
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr>
<tr><td>Scoring system +ve</td><td>286</td><td>3</td><td>289</td></tr>
</table>
</td>
<td>scoring system was conducted first
2. If a threshold was used, was it pre-specified? Yes, thresholds for different health states were pre-specified
Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD
A. RISK OF BIAS
1. Is the reference standard likely to correctly classify the target condition? No, a research nurse graded the babies, an experienced paediatrician would be considered the reference standard to classify the target condition
2. Were the reference standard results interpreted without knowledge of the results of the index test? Unclear, no details on the methods
Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: HIGH

B. CONCERNS REGARDING APPLICABILITY Is there concern</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td></td>
<td></td>
<td>independent paediatricians (n=248), and investigation results (n=682). Whereas, in the community a research nurse on 2 weekdays per week from 9-5pm graded the babies. The observers acted independently of paediatricians in charge of admissions and took no part in the decision to admit patients. All the babies were also followed up for 3 days to ensure no serious diseases had been missed. After comparing each of these criteria, the assessor's impression of the illness was chosen as the grading for subsequent analysis it was recorded for all babies at the time they were seen, and there was a high level of concordance between this and the independent paediatrician's review (x=0.62, p<0.001). This level</td>
<td><u>Grading the severity of illness</u>
For details see assessment.
<u>Exploratory analyses</u>
Exploratory analyses showed that symptoms best discriminated illness severity if they were present only during the preceding 3 days. Continuous variables were investigated as linear and quadratic trends. Most, including respiratory rate, pulse rate, weight, and weight change did not contribute to the prediction of illness severity in the presence of other variables. 2 exceptions were rectal temperature and vomiting. These were converted to present/absent variables using the cut offs >38.2 degrees Celsius and 'vomits of at least half the feed after each of the last three feeds'.
<u>Identification of the best combination of symptoms and signs for grading illness</u>
Logistic and ordinal regression analyses were used to identify the best combination of</td>
<td>
<table>
<tr><td>Scoring system -ve</td><td>4</td><td>5</td><td>9</td></tr>
<tr><td>Total</td><td>290</td><td>8</td><td>298</td></tr>
</table>
Sensitivity: 98.6% (95% CI 96.5% to 99.6%)*
Specificity: 62.5% (95% CI 24.5% to 91.5%)*
Positive likelihood ratio: 2.63 (95% CI 1.07 to 6.43)*
Negative likelihood ratio: 0.02 (95% CI 0.01 to 0.07)*
Prevalence of well or mildly ill babies: 97.3%*

<u>Babies seen in hospital</u>
Outcome: Seriously ill (defined as a score of 13 or more)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>152</td><td>164</td><td>316</td></tr>
<tr><td>Scoring system -ve</td><td>13</td><td>380</td><td>398</td></tr>
<tr><td>Total</td><td>165</td><td>544</td><td>709</td></tr>
</table>
Sensitivity: 92.12% (95% CI 86.9% to 95.7%)*
Specificity: 69.85% (95% CI 65.8% to 73.7%)*
</td>
<td>that the target condition as defined by the reference standard does not match the review question? CONCERN: HIGH

DOMAIN 4: FLOW AND TIMING
A. RISK OF BIAS
1. Was there appropriate interval between index tests and reference standard? Yes
2. Did all patients receive a reference standard? Yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? Yes
Could the patient flow have introduced bias? RISK: LOW

<u>Babies seen in hospital</u>
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? No, during busy times, the paediatrician biased enrolment towards the more ill of the babies presented, this ensured a wide spectrum of illness in the babies taking part in the study.
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td></td>
<td></td>
<td>of agreement is similar to other studies comparing clinical judgement. Using a 7-point scale from 'normal baby requiring hospital investigations or treatment' (point 7) through to 'needs urgent hospital attention for a life threatening condition' (point 1). Scores between 1 and 4 ('need to be admitted for observation) were subsequently used to define serious illness.</td>
<td>symptoms and sins to differentiate the 3 illness groups (well plus mildly ill, moderately ill, and seriously ill). The regression coefficient for each symptoms and sign represents the increased chance of a baby being ill when that symptom or sign is present. Thus compared with an asymptomatic baby, the chance of a symptomatic baby being ill is the sum of the coefficients for the symptoms which are present. The regression coefficient for the symptoms which are present. The regression coefficients were converted to integers by multiplying by 3-93 to make a manageable score. The total score could then be calculated for each baby by identifying which of the 19 symptoms and signs the baby had, and adding the corresponding scores.
<u>The scores in a theoretical community population</u>
The sample was weighted towards hospital babies. To find the scores likely to occur</td>
<td>Positive likelihood ratio: 3.06 (95% CI 2.67 to 3.5)*
Negative likelihood ratio: 0.11 (95% CI 0.07 to 0.19)*
Prevalence of seriously ill babies: 23.3%*

Outcome: Moderately ill (defined as a score of 8-12)
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr>
<tr><td>Scoring system +ve</td><td>67</td><td>82</td><td>149</td></tr>
<tr><td>Scoring system -ve</td><td>172</td><td>388</td><td>560</td></tr>
<tr><td>Total</td><td>239</td><td>470</td><td>709</td></tr>
</table>
Sensitivity: 28.03% (95% CI 86.9% to 95.7%)*
Specificity: 82.55 (95% CI 78.8% to 85.9%)*
Positive likelihood ratio: 1.61 (95% CI 1.21 to 2.13)*
Negative likelihood ratio: 0.87 (95% CI 0.80 to 0.95)
Prevalence of moderately ill babies: 33.7%*

Outcome: Well or mildly ill (defined as a score of 0-7)
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr>
</table>
</td>
<td>Could the selection of patients have introduced bias? RISK: HIGH
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: HIGH

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes, Baby check scoring system was conducted first
2. If a threshold was used, was it pre-specified? Yes, thresholds for different health states were pre-specified
Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD
A. RISK OF BIAS
1. Is the reference standard likely to correctly classify the target condition? Yes, however there is no gold standard for</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>in the community, a theoretical cohort of 10 000 babies was calculated by taking the babies seen at home as 98% of the population and those seen in hospital as 2%. These proportions were based on the assumption that the hospital babies represented the illest 2% of the population.

<u>Sensitivity, Specificity, and predictive values</u>

The scoring system is designed to grade the severity of a baby's illness, with increasing scores identifying sicker babies. Specificity (the accuracy with which the score identifies well or mildly unwell babies) and sensitivity (the accuracy with which the score identifies seriously ill babies) are calculated for groups of scores as an illustration of the accuracy of the scoring system. Predictive values (the chance of a baby with a given score having a given grade of</td>
<td>
<table>
<tr><td>Scoring system +ve</td><td>189</td><td>55</td><td>244</td></tr>
<tr><td>Scoring system -ve</td><td>116</td><td>349</td><td>465</td></tr>
<tr><td>Total</td><td>305</td><td>404</td><td>709</td></tr>
</table>
Sensitivity: 61.97% (95% CI 56.3% to 67.4%)*
Specificity: 86.39% (95% CI 82.7% to 89.6%)*
Positive likelihood ratio: 4.55 (95% CI 3.51 to 5.91)
Negative likelihood ratio: 0.44 (95% CI 0.38 to 0.51)*
Prevalence of well or mildly ill babies: 43%*

<u>Theoretical cohort of 10,000 community babies, weighted 98:2 home: hospital</u>

Outcome: Seriously ill (defined as a score of 13 or more)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>43</td><td>145</td><td>188</td></tr>
<tr><td>Scoring system -ve</td><td>3</td><td>9809</td><td>9812</td></tr>
</table>
</td>
<td>severity of illness assessment and although the assessment was conducted by a paediatrician there is subjectivity in the diagnosis.
2. Were the reference standard results interpreted without knowledge of the results of the index test? Yes, observers acted independently of paediatricians in charge of admissions and took no part in the decision to admit patients.
Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW

DOMAIN 4: FLOW AND TIMING
A. RISK OF BIAS
1. Was there appropriate interval between index tests and reference standard? Yes
2. Did all patients receive a reference standard? Yes
3. Did patients receive the same reference standard? No, for comparison each baby's illness was also graded, where possible, using other criteria: positive investigation results (n=200); a review of the notes by 3 independent paediatricians</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td>illness) are shown for individual scores.</td><td>
<table>
<tr><td>Total</td><td>46</td><td>9954</td><td>10000</td></tr>
</table>
Sensitivity: 93.48% (95% CI 82.1% to 98.6%)*
Specificity: 98.54% (95% CI 98.3% to 98.8%)*
Positive likelihood ratio: 64.17 (95% CI 53.67 to 76.73)*
Negative likelihood ratio: 0.07 (95% CI 0.02to 0.2)*
Prevalence of seriously ill babies: 0.46%

Outcome: Moderately ill (defined as a score of 8-12)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>118</td><td>121</td><td>239</td></tr>
<tr><td>Scoring system -ve</td><td>213</td><td>9548</td><td>9761</td></tr>
<tr><td>Total</td><td>331</td><td>9669</td><td>10000</td></tr>
</table>
Sensitivity: 35.65% (95% CI 30.5% to 41.1%)*
Specificity: 98.75% (95% CI 98.5% to 99.0%)*
Positive likelihood ratio: 28.49 (95% CI 22.66 to 35.81)*
Negative likelihood ratio: 0.65 (0.60 to 0.71)*
Prevalence of moderately ill babies: 3.3%*</td><td>(n=248); and investigation results (n=682)
4. Were all patients included in the analysis? Yes
Could the patient flow have introduced bias? RISK: LOW

<u>Theoretical community cohort</u>
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? No, data extrapolated from 2 of the study cohorts using assumptions.
2. Was a case-control design avoided? No, infants assessed at home were the control group
3. Did the study avoid inappropriate exclusions? Yes
Could the selection of patients have introduced bias? RISK: HIGH
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: HIGH

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes, Baby check scoring system was conducted first</td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>Outcome: Well or mildly ill (defined as a score of 0-7)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>9459</td><td>114</td><td>9573</td></tr>
<tr><td>Scoring system -ve</td><td>164</td><td>263</td><td>427</td></tr>
<tr><td>Total</td><td>9623</td><td>377</td><td>10000</td></tr>
</table>
Sensitivity: 98.30% (95% CI 98% to 98.5%)*
Specificity: 69.76% (95% CI 64.5% to 74.4%)*
Positive likelihood ratio: 3.25 (95% CI 2.79 to 3.79)*
Negative likelihood ratio: 0.02 (95% CI 0.02 to 0.03)*
Prevalence of well or mildly ill babies: 96.3%*

*Calculated by the NGA technical team</td><td>2. If a threshold was used, was it pre-specified? Yes, thresholds for different health states were pre-specified
Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD
A. RISK OF BIAS
1. Is the reference standard likely to correctly classify the target condition? Yes, however there is no gold standard for severity of illness assessment and although the assessment was conducted by a paediatrician there is subjectivity in the diagnosis.
2. Were the reference standard results interpreted without knowledge of the results of the index test? Yes, observers acted independently of paediatricians in charge of admissions and took no part in the decision to admit patients.
Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW</td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW

DOMAIN 4: FLOW AND TIMING
A. RISK OF BIAS
1. Was there appropriate interval between index tests and reference standard? Yes
2. Did all patients receive a reference standard? No, data extrapolated from 2 of the study cohorts using assumptions
3. Did patients receive the same reference standard? No, data extrapolated from 2 of the study cohorts using assumptions
4. Were all patients included in the analysis? Yes
Could the patient flow have introduced bias? RISK: HIGH</td></tr>
<tr><td>Full citation

Thornton, A. J., Morley, C. J., Cole, T. J., Green, S. J., Walker, K. A., Rennie, J. M., Field trials of the Baby Check score card in hospital, Archives of Disease in Childhood, 66, 115-120, 1991</td><td>Sample size
N=357 babies presented to casualty and eligible for study
n=262 babies scored by house officer
n=196 babies graded by registrar
n=259 babies graded by consultant A
n=260 babies graded by consultant B
n=193 babies graded by registrar, consultant A, and consultant B</td><td>Tests
<u>Baby check scoring system</u>
7 symptoms and 12 signs that in combination graded illness severity best.
<u>Grading of systemic illness</u>
As soon after presentation as possible, without knowledge of the score, the duty paediatric registrar</td><td>Methods
<u>Sample size</u>
not reported
<u>The assessment of the baby</u>
13 paediatric house officers at Addenbrooke's Hospital were asked to score every baby under 26 weeks old presenting for assessment of an acute illness. They received no instruction on the use of</td><td>Results

<u>Babies seen by consultant A in hospital</u>
Outcome: Serious illness needing hospital treatment (defined as a score of 20 or more, grade 1)<table><tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr><tr><td>Scoring system +ve</td><td>43</td><td>13</td><td>56</td></tr></table></td><td>Limitations

Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? Yes
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes</td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td>Ref Id

1255716

Country/ies where the study was carried out

UK

Study type

Prospective cohort study

Aim of the study
To report a field trial in which Baby Check was used to score babies presenting with an acute illness.

Study dates
not reported

Source of funding
Baby illness research project appeal of the foundation for the study of infant deaths</td>
<td>Characteristics
Of the babies score, 227 (87%) were admitted (7 to intensive care) and 34 (13%) were sent home (1 not recorded). The median stay was 2 days, ranging from a few hours to 99 days (none died). The babies had a broad range of diagnoses, from minor complaints such as nappy rash to serious illnesses such as meningitis.
The scores ranged from 0-57, with a median of 12 (10th and 90th centiles 0, 34). The median score for babies to be sent home on was 3 and for those admitted to paediatric wards and to intensive care, 13 (3, 34) and 30, respectively.
Of the 262 scores, 100 (38%) were between 0 and 7, 40 (15%) 8 to 12, 51 (20%) 13 to 20, and 71 (27%) more than 20.</td>
<td>graded each baby's illness on a 7-point scale, ranging from: 'Baby needs urgent hospital treatment for a life threatening condition' to 'Well baby not requiring any special care or treatment'. The registrars grading reflected the baby's state at the time of presentation. 2 consultant paediatricians reviewed each baby's notes after discharge, using the same scale and without knowledge of the score. Their gradings took into account the investigation results, diagnosis, treatment, and outcome. For the analyses, these gradings were simplified into 4 categories: 1. Had a serious illness requiring hospital treatment; 2. Requires hospital admission for observation due to uncertainty about the severity of illness; 3. Needs careful</td>
<td>the Baby Check score card.

<u>Grading of severity of illness</u>
For details see grading of systemic illness

<u>Sensitivity, Specificity, and predictive values</u>
Differences in sensitivity, specificity, and predictive accuracy between the score and the registrars grading were explored through chi-squared analyses.</td>
<td>
<table>
<tr><td>Scoring system -ve</td><td>51</td><td>86</td><td>137</td></tr>
<tr><td>Total</td><td>94</td><td>99</td><td>193</td></tr>
</table>
Sensitivity: 45.74% (95% CI 35.4% to 56.3%)*
Specificity: 86.9% (95% CI 78.59% to 92.82%)*
Positive likelihood ratio: 3.48 (95% CI 2.01 to 6.05)*
Negative likelihood ratio: 0.62 (95% CI 0.51 to 0.76)*
Prevalence of babies that are serious illness needing hospital treatment: 48.7%*

Outcome: Serious illness needing hospital treatment or requires hospital admission for observation due to uncertainty about severity of illness (defined as a score of 13 or more, grade 1 and 2)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>82</td><td>14</td><td>96</td></tr>
<tr><td>Scoring system -ve</td><td>56</td><td>41</td><td>97</td></tr>
<tr><td>Total</td><td>138</td><td>55</td><td>193</td></tr>
</table>
</td>
<td>Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes, Baby Check scoring system was conducted first
2. If a threshold was used, was it pre-specified? Yes, thresholds for different health states were pre-specified
Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD
A. RISK OF BIAS
1. Is the reference standard likely to correctly classify the target condition? Yes, however there is no gold standard for</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td></td>
<td>Inclusion Criteria
Baby under 26 weeks old presenting for assessment of acute illness to Addenbrookes hospital.

Exclusion Criteria
not reported</td>
<td>observation and treatment. Could be managed at home by a capable mother; 4. Mildly ill or well. Could be managed at home by any mother.</td>
<td></td>
<td>Sensitivity: 59.42% (95% CI 50.74% to 67.69%)*
Specificity: 74.55% (95% CI 61.0% to 85.3%)*
Positive likelihood ratio: 2.33 (95% CI 1.45 to 3.75)*
Negative likelihood ratio: 0.54 (95% CI 0.42 to 0.70)*
Prevalence of babies that require hospital admission for observation: 71.5%*

Outcome: Requires hospital admission for observation due to uncertainty about severity of illness (defined as a score of 13-19, grade 2)
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr>
<tr><td>Scoring system +ve</td><td>10</td><td>30</td><td>40</td></tr>
<tr><td>Scoring system -ve</td><td>34</td><td>119</td><td>153</td></tr>
<tr><td>Total</td><td>44</td><td>149</td><td>193</td></tr>
</table>
Sensitivity: 22.7% (95% CI 11.5% to 37.8%)*
Specificity: 79.9% (95% CI 72.5% to 86%)*
Positive likelihood ratio: 1.13 (95% CI 0.6 to 2.12) *
Negative likelihood ratio: 0.97 (95% CI 0.81 to 1.16) *
Prevalence of babies that require hospital admission for observation: 22.8*</td>
<td>severity of illness assessment and although the assessment was conducted by 2 paediatricians there is subjectivity in the diagnosis.
2. Were the reference standard results interpreted without knowledge of the results of the index test? Yes, as soon after presentation as possible, without knowledge of the score, the duty paediatric registrar graded each baby's illness on a seven-point scale, ranging from: 'Baby needs urgent hospital treatment for a life threatening condition' to 'Well baby not requiring any special care or treatment'. The registrars grading reflected the baby's state at the time of presentation. 2 consultant paediatricians reviewed each baby's notes after discharge, using the same scale and without knowledge of the score.
Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW

DOMAIN 4: FLOW AND TIMING
A. RISK OF BIAS</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>Outcome: Needs careful observation and treatment. Could be managed at home by a capable mother (defined as a score of 8-12, grade 3)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>8</td><td>21</td><td>29</td></tr>
<tr><td>Scoring system -ve</td><td>35</td><td>129</td><td>164</td></tr>
<tr><td>Total</td><td>43</td><td>150</td><td>193</td></tr>
</table>
Sensitivity: 18.6% (95% CI 8.39% to 33.4%)*
Specificity: 86.0% (95% CI 79.4% to 91.12%)*
Positive likelihood ratio: 1.33 (95% CI 0.63 to 2.79)*
Negative likelihood ratio: 0.95 (95% CI 0.81 to 1.11)*
Prevalence of babies that need careful observation and treatment: 22.3%*

Outcome: Mildly ill or well. Could be managed at home by any mother (defined as a score 0-7, grade 4)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
</table></td><td>1. Was there appropriate interval between index tests and reference standard? Yes
2. Did all patients receive a reference standard? Yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? No, 54% (193/357) of babies who presented to hospital and eligible for the study were analysed.
Could the patient flow have introduced bias? RISK: LOW</td></tr>
</table>

Bibliographic details	Participants	Scoring systems and assessment	Methods	Outcomes and results	Comments

Scoring system +ve	11	57	68
Scoring system -ve	1	124	125
Total	12	181	193

Sensitivity: 91.67% (95% CI 61.52% to 99.79%)*
Specificity: 68.51% (95% CI 61.20% to 75.20%)*
Positive likelihood ratio: 2.91 (95% CI 2.21 to 3.83)*
Negative likelihood ratio: 0.12 (95% CI 0.02 to 0.80)*
Prevalence of mildly ill or well babies: 6.2%*

Babies seen by consultant B in hospital
Outcome: Serious illness needing hospital treatment (defined as a score of 20 or more, grade 1)

	Reference assessment +ve	Reference assessment -ve	Total
Scoring system +ve	38	18	56
Scoring system -ve	25	112	137

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>
<table>
<tr><td>Total</td><td>63</td><td>130</td><td>193</td></tr>
</table>
Sensitivity: 60.32% (95% CI 47.20% to 72.43%)*
Specificity: 86.15% (95% CI 79.0% to 91.58%)*
Positive likelihood ratio: 4.36 (95% CI 2.71 to 6.99)
Negative likelihood ratio: 0.46 (0.34 to 0.63)*
Prevalence of babies that are serious illness needing hospital treatment: 32.6%*

Outcome: Serious illness needing hospital treatment or requires hospital admission for observation due to uncertainty about severity of illness (defined as a score of 13 or more, grade 1 and 2)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>83</td><td>13</td><td>96</td></tr>
<tr><td>Scoring system -ve</td><td>43</td><td>54</td><td>97</td></tr>
<tr><td>Total</td><td>126</td><td>67</td><td>193</td></tr>
</table>
Sensitivity: 65.87% (95% CI 56.90% to 74.08%)*
Specificity: 80.60% (95% CI 69.11% to 89.24%)*
</td><td></td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>Positive likelihood ratio: 3.39 (95% CI 2.05 to 5.62)*
Negative likelihood ratio: 0.42 (95% CI 0.32 to 0.55)*
Prevalence of babies that require hospital admission for observation: 65.3%*

Outcome: Requires hospital admission for observation due to uncertainty about severity of illness (defined as a score of 13-19, grade 2)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>15</td><td>25</td><td>40</td></tr>
<tr><td>Scoring system -ve</td><td>48</td><td>105</td><td>153</td></tr>
<tr><td>Total</td><td>63</td><td>130</td><td>193</td></tr>
</table>
Sensitivity: 23.8% (95% CI 14% to 36.2%)*
Specificity: 80.8% (95% CI 72.9% to 87.2%)*
Positive likelihood ratio: 1.24 (95% CI 0.7 to 2.18)*
Negative likelihood ratio: 0.94 (95% CI 0.8 to 1.11)*
Prevalence of babies that require hospital admission for observation: 32.6%*

Outcome: Needs careful observation and treatment. Could be managed at home by a capable mother (defined as a score of 8-12, grade 3)</td><td></td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr>
<tr><td>Scoring system +ve</td><td>8</td><td>21</td><td>29</td></tr>
<tr><td>Scoring system -ve</td><td>42</td><td>122</td><td>164</td></tr>
<tr><td>Total</td><td>50</td><td>143</td><td>193</td></tr>
</table>
Sensitivity: 16% (95% CI 7.17% to 29.11%)*
Specificity: 85.31% (95% CI 78.43% to 90.67%)*
Positive likelihood ratio: 1.09 (95% CI 78.43% to 90.67%)*
Negative likelihood ratio: 0.98 (95% CI 0.86 to 1.13)*
Prevalence of babies that need careful observation and treatment: 25.9%*

Outcome: Mildly ill or well. Could be managed at home by any mother (defined as a score 0-7, grade 4)
<table>
<tr><th></th><th>Reference assessment +ve</th><th>Reference assessment -ve</th><th>Total</th></tr>
<tr><td>Scoring system +ve</td><td>14</td><td>54</td><td>68</td></tr>
</table>
</td><td></td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>
<table>
<tr><td>Scoring system -ve</td><td>3</td><td>122</td><td>125</td></tr>
<tr><td>Total</td><td>17</td><td>176</td><td>193</td></tr>
</table>
Sensitivity: 82.35% (95% CI 56.57% to 96.20%)*
Specificity: 69.32% (95% CI 61.94% to 76.04%)*
Positive likelihood ratio: 2.68 (95% CI 1.96 to 3.67)*
Negative likelihood ratio: 0.25 (95% CI 0.09 to 0.71)*
Prevalence of mildly ill or well babies: 8.8%*

*Calculated by NGA technical team
</td><td></td></tr>
<tr><td>Full citation

Chen, C. K., Chen, S. J., Hwang, B., Evaluation of the severity of illness in infants by the Baby Check Score, Zhonghua yi xue za zhi = Chinese medical journal; Free China ed, 59, 15-20, 1997

Ref Id

1255842</td><td>Sample size
N=495 babies <6 months old presented at emergency department.
n=394 babies scored and graded
n=134 babies graded retrospectively by 2 senior independent paediatricians

Characteristics

The Baby Check Scores were subdivided into four groups: 182</td><td>Tests
<u>Baby check scoring system</u>
7 symptoms and 12 signs that in combination graded illness severity best. Translated into Chinese.
<u>Grading of systemic illness</u>

A total of 15 on-duty residents, in ignorance of the score recorded by the interns, graded each baby's illness into four groups:</td><td>Methods
<u>Sample size</u>
Not reported
<u>The assessment of the baby</u>
16 interns scored the babies who were under 6 months of age using a Baby Check Score which was translated into Chinese.

<u>Grading of severity of illness</u>

See assessment

The concordance between the Baby Check</td><td>Results
<u>Babies seen by paediatrician A in hospital</u>
Unable to calculate data from 2 x 2 table in paper as missing data

<u>Babies seen by paediatrician B in hospital</u>
Outcome: Serious illness needing hospital treatment (defined as a score of 20 or more)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>8</td><td>8</td><td>16</td></tr>
</table>
</td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? Yes
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes
Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not</td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td>Country/ies where the study was carried out

Taiwan

Study type

Prospective cohort study

Aim of the study

To identify the concordance between the Baby Check Score System and the clinical evaluation of disease severity by paediatricians.

Study dates
April 1992 - July 1992

Source of funding
not reported</td>
<td>(46.2%) were between 0 and 7, 108 (27.4%) were between 8 and 12, 71 (18.0%) were between 13 and 19; 33 (8.4%) were 20 or more.

Inclusion Criteria

Babies who were under six months of age, brought to the Paediatric Emergency Room of Veterans General Hospital (VGH)-Taipei

Exclusion Criteria
Not reported</td>
<td>well, mildly ill, moderately ill and seriously ill babies. The definitions were as follows: (1) well babies: babies who could be managed at home, (2) mildly ill babies: babies who required careful observation and treatment, but could be managed at home by a capable person, (3) moderately ill babies: babies who required hospital admission for observation when there was uncertainty about the severity of the illness, (4) seriously ill babies: babies with a serious illness who needed hospital treatment. Two senior paediatricians (third-year paediatric residents A and B) reviewed the medical records after the babies were discharged, and graded the severity of the illness with the same definitions as on-duty residents did, in ignorance of</td>
<td>Score and the illness gradins used the same cut off point between mildly and moderately ill, and a score of 13. With the senior paediatricians' gradings as the standard, the sensitivity, specificity, and predictive accuracy of the score and the on-duty resident' ratings were compared with Chi-square analysis.</td>
<td>
<table>
<tr><td>Scoring system -ve</td><td>2</td><td>116</td><td>118</td></tr>
<tr><td>Total</td><td>10</td><td>124</td><td>134</td></tr>
</table>
Sensitivity: 80.0% (95% CI 44.4% to 97.5%)*
Specificity: 93.55% (95% CI 87.7% to 97.2%)*
Positive likelihood ratio: 12.4 (95% CI 5.93 to 25.95)*
Negative likelihood ratio: 0.21 (95% CI 0.06 to 0.74)*
Prevalence of babies that are serious illness needing hospital treatment: 7.5%

Outcome: Serious illness needing hospital treatment or moderate illness that requires hospital admission for observation due to uncertainty about severity of illness (defined as a score of 13 or more)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>28</td><td>14</td><td>42</td></tr>
<tr><td>Scoring system -ve</td><td>14</td><td>78</td><td>92</td></tr>
<tr><td>Total</td><td>42</td><td>92</td><td>134</td></tr>
</table>
</td>
<td>match the review question? CONCERN: LOW

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes, Baby Check scoring system was conducted first
2. If a threshold was used, was it pre-specified? Yes, thresholds for different health states were pre-specified
Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD
A. RISK OF BIAS
1. Is the reference standard likely to correctly classify the target condition? Yes, however there is no gold standard for severity of illness assessment and although the assessment was conducted
by 2 senior paediatricians there is subjectivity in the diagnosis.</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td>the score and the previous gradings.</td><td></td><td>Sensitivity: 66.67% (95% CI 50.5% to 80.4%)*
Specificity: 84.78% (95% CI 75.8% to 91.4%)*
Positive likelihood ratio: 4.38 (95% CI 2.58 to 7.43)*
Negative likelihood ratio: 0.39 (95% CI 0.25 to 0.61)*
Prevalence of babies that are seriously ill or moderately ill: 31.3%*

Outcome: Moderate illness that requires hospital admission for observation due to uncertainty about severity of illness (defined as a score of 13-19)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment - ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>12</td><td>14</td><td>26</td></tr>
<tr><td>Scoring system - ve</td><td>20</td><td>88</td><td>108</td></tr>
<tr><td>Total</td><td>32</td><td>102</td><td>134</td></tr>
</table>
Sensitivity: 37.5% (95% CI 21.1% to 56.3%)*
Specificity: 86.27% (95% CI 78.0% to 92.3%)*
Positive likelihood ratio: 2.73 (95% CI 1.4 to 5.29)*
Negative likelihood ratio: 0.72 (95% CI 0.55 to 0.96)*
Prevalence of babies that are moderately ill: 23.9%*</td><td>2. Were the reference standard results interpreted without knowledge of the results of the index test? Yes, 2 senior paediatricians (third-year paediatric residents A and B) reviewed the medical records after the babies were discharged, and graded the severity of the illness with the same definitions as on-duty residents did, in ignorance of the score and the previous gradings.
Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW

DOMAIN 4: FLOW AND TIMING
A. RISK OF BIAS
1. Was there appropriate interval between index tests and reference standard? Yes
2. Did all patients receive a reference standard? Yes
3. Did patients receive the same reference standard? Yes
4. Were all patients included in the analysis? No, 34% (134/495) of babies who presented to hospital and eligible for the study were analysed
Could the patient flow have introduced bias? RISK: LOW</td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>Outcome: Mildly ill needing careful observation and treatment. Could be managed at home by a capable mother (defined as a score of 8-12)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>24</td><td>10</td><td>34</td></tr>
<tr><td>Scoring system -ve</td><td>49</td><td>51</td><td>100</td></tr>
<tr><td>Total</td><td>73</td><td>61</td><td>134</td></tr>
</table>
Sensitivity: 32.89% (95% CI 22.4% to 44.8%)*
Specificity: 83.61% (95% CI 71.9% to 91.9%)*
Positive likelihood ratio: 2.01 (95% CI 1.04 to 3.86)*
Negative likelihood ratio: 0.80 (95% CI 0.66 to 0.98)
Prevalence of babies that are mildly ill: 54.5%*

Outcome: Well. Could be managed at home by any mother (defined as a score 0-7)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
</table></td><td></td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>
<table>
<tr><td>Scoring system +ve</td><td>17</td><td>41</td><td>58</td></tr>
<tr><td>Scoring system -ve</td><td>2</td><td>74</td><td>76</td></tr>
<tr><td>Total</td><td>19</td><td>115</td><td>134</td></tr>
</table>
Sensitivity: 89.47% (95% CI 66.9% to 98.7%)*
Specificity: 64.35% (95% CI 54.88% to 73.06%)*
Positive likelihood ratio: 2.51 (95% CI 1.88 to 3.35)*
Negative likelihood ratio: 0.16 (95% CI 0.04 to 0.61)*
Prevalence of babies that are well: 14.2%

*Calculated by the NGA technical team
</td><td></td></tr>
<tr><td>Full citation

Chandran, S., Sunita, K., Nair, A. K., Elbualy, M. S., A trial of baby check scoring system to identify high-risk infants in a polyclinic in Oman, Journal of Tropical Pediatrics, 44, 218-221, 1998

Ref Id</td><td>Sample size
N=90

Characteristics
The baby check scores ranged from 0 to 41 with a median of 10 and mean of 14.1 (SD ± 11.34).
Diagnoses in the group of well or mildly ill babies (n=43): upper respiratory tract infection n=21;</td><td>Tests
<u>Baby check scoring system</u>
7 symptoms and 12 signs that in combination graded illness severity best.
<u>Grading of systemic illness</u>
The patients were examined by the junior doctors and the working diagnosis was recorded in each case. They filled in</td><td>Methods
<u>Sample size</u>
not reported
<u>The assessment of baby</u>
For details see grading of systemic illness. The paediatric patients in this clinic are first attended by junior physicians with limited training in paediatrics.
<u>Grading of severity of illness</u>
for details see grading of systemic illness.</td><td>Results
<u>Babies seen in a polyclinic</u>
Outcome: Seriously ill needing immediate referral to tertiary care (defined as a score of 20 or more)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>6</td><td>16</td><td>22</td></tr>
</table>
</td><td>Limitations
Risk of bias assessed using QUADAS-II
DOMAIN 1: PATIENT SELECTION
A. RISK OF BIAS
1. Was a consecutive or random sample of patients enrolled? Unclear, no details on patient recruitment
2. Was a case-control design avoided? Yes
3. Did the study avoid inappropriate exclusions? Yes</td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr>
<td>1255867

Country/ies where the study was carried out

Oman

Study type

Prospective cohort study

Aim of the study
To test the usefulness of Baby Check score in a busy polyclinic in the Sultanate of Oman.

Study dates
February-June 1995

Source of funding
not reported</td>
<td>pyoderma n=11; diarrhoea n=4; Eczema n=5; nappy rash n=1; oral thrush n=1.
Diagnoses in the group of moderately ill babies (n=41): upper respiratory tract infection n=31; diarrhoea n=7; dysentery n=2; pyoderma n=1.
Diagnoses in the group of seriously ill babies (n=6); bronchiolitis n=2; pneumonia n=3; diarrhoea with severe dehydration n=1.

Inclusion Criteria
Infants in the age range of 1 to 6 months who presented to the clinic

Exclusion Criteria
not reported</td>
<td>the score card and also identified cases as well or mildly ill (patients needing minor medication and/or reassurance), moderately ill (patients needing observation and consultation), and seriously ill (those needing immediate referral for tertiary care). The scores were then assigned to the various signs and symptoms by the investigators, following the guidelines set by the original authors . The majority of these cases were reviewed 1 week later by the same physician or the specialist (paediatrician with specialised training).</td>
<td></td>
<td>
<table>
<tr><td>Scoring system -ve</td><td>0</td><td>68</td><td>68</td></tr>
<tr><td>Total</td><td>6</td><td>84</td><td>90</td></tr>
</table>
Sensitivity: 100% (95% CI 54.07% to 100%)*
Specificity: 80.95% (95% CI 70.9% to 88.72%)*
Positive likelihood ratio: 5.25 (95% CI 3.38 to 8.16)*
Negative likelihood ratio: 0.0
Prevalence of babies that are seriously ill: 6.7%

Outcome: Seriously ill needing immediate referral to tertiary care or Moderately ill needing observation and consultation (defined as a score of 13 or more)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>35</td><td>5</td><td>40</td></tr>
<tr><td>Scoring system -ve</td><td>12</td><td>38</td><td>50</td></tr>
<tr><td>Total</td><td>47</td><td>43</td><td>90</td></tr>
</table>
Sensitivity: 74.47% (95% CI 59.6% to 86.1%)*
</td>
<td>Could the selection of patients have introduced bias? RISK: LOW
B. CONCERNS REGARDING APPLICABILITY Is there concern that the included patients do not match the review question? CONCERN: LOW

DOMAIN 2: INDEX TESTS
A. RISK OF BIAS
1. Were the index test results interpreted without knowledge of the results of the reference standard? Yes, Baby Check scoring system was conducted first
2. If a threshold was used, was it pre-specified? Yes, thresholds for different health states were pre-specified
Could the conduct or interpretation of the index test have introduced bias? RISK: LOW

B. CONCERNS REGARDING APPLICABILITY Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD
A. RISK OF BIAS
1. Is the reference standard likely to correctly classify the target condition? Unclear, although there is no gold</td>
</tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>Specificity: 88.37% (95% CI 74.9% to 96.1%)*
Positive likelihood ratio: 6.40 (95% CI 2.76 to 14.85)*
Negative likelihood ratio: 0.29 (95% CI 0.18 to 0.48)*
Prevalence of babies that are seriously ill or moderately ill: 52%*

Outcome: Moderately ill needing observation and consultation (defined as a score of 13-19)
<table><tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr><tr><td>Scoring system +ve</td><td>15</td><td>3</td><td>18</td></tr><tr><td>Scoring system -ve</td><td>26</td><td>46</td><td>72</td></tr><tr><td>Total</td><td>41</td><td>49</td><td>90</td></tr></table>
Sensitivity: 36.59% (95% CI 22.1% to 53.0%)*
Specificity: 93.88% (95% CI 83.13% to 98.72%)*
Positive likelihood ratio: 5.98 (95% CI 1.86 to 19.22)*
Negative likelihood ratio: 0.68 (95% CI 0.53 to 0.86)*
Prevalence of babies that are moderately ill: 45.6%*</td><td>standard for severity of illness assessment, an experienced paediatrician would be an appropriate assessor, in the study the assessor may have been a junior doctor or a paediatrician, thus introducing uncertainty to the diagnosis.
2. Were the reference standard results interpreted without knowledge of the results of the index test? No, assessor may be a junior doctor that assessed the infant using Baby check rather than an independent paediatrician. Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: HIGH

B. CONCERNS REGARDING APPLICABILITY Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: HIGH

DOMAIN 4: FLOW AND TIMING
A. RISK OF BIAS
1. Was there appropriate interval between index tests and reference standard? Yes
2. Did all patients receive a reference standard? Yes
3. Did patients receive the same reference standard? No, the reference standard was either a junior doctor or a Paediatrician
4. Were all patients included in the analysis? Yes</td></tr>
</table>

<table>
<tr><th>Bibliographic details</th><th>Participants</th><th>Scoring systems and assessment</th><th>Methods</th><th>Outcomes and results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>Outcome: Mildly ill or well needing minor medication and/or reassurance (defined as a score <13)
<table>
<tr><td></td><td>Reference assessment +ve</td><td>Reference assessment -ve</td><td>Total</td></tr>
<tr><td>Scoring system +ve</td><td>38</td><td>12</td><td>50</td></tr>
<tr><td>Scoring system -ve</td><td>5</td><td>35</td><td>40</td></tr>
<tr><td>Total</td><td>43</td><td>47</td><td>90</td></tr>
</table>
Sensitivity: 88.37% (95% CI 74.9% to 96.1%)*
Specificity: 74.47% (95% CI 59.7% to 86.1%)*
Positive likelihood ratio: 3.46 (95% CI 2.10 to 5.71)*
Negative likelihood ratio: 0.16 (95% CI 0.07 to 0.36)*
Prevalence of mildly ill or well babies: 47.80%*

*Calculated by the NGA technical team</td><td>Could the patient flow have introduced bias? RISK: LOW</td></tr>
<tr><td>Full citation

T. J. Cole, C. J. Morley, A. J. Thornton, M. A. Fowler, P. H. Hewson, A Scoring System to</td><td>Sample size
For details see Morley 1991

Characteristics</td><td>Tests
For details see Morley 1991</td><td>Methods
For details see Morley 1991</td><td>Results
For details see Morley 1991</td><td>Limitations
For details see Morley 1991</td></tr>
</table>

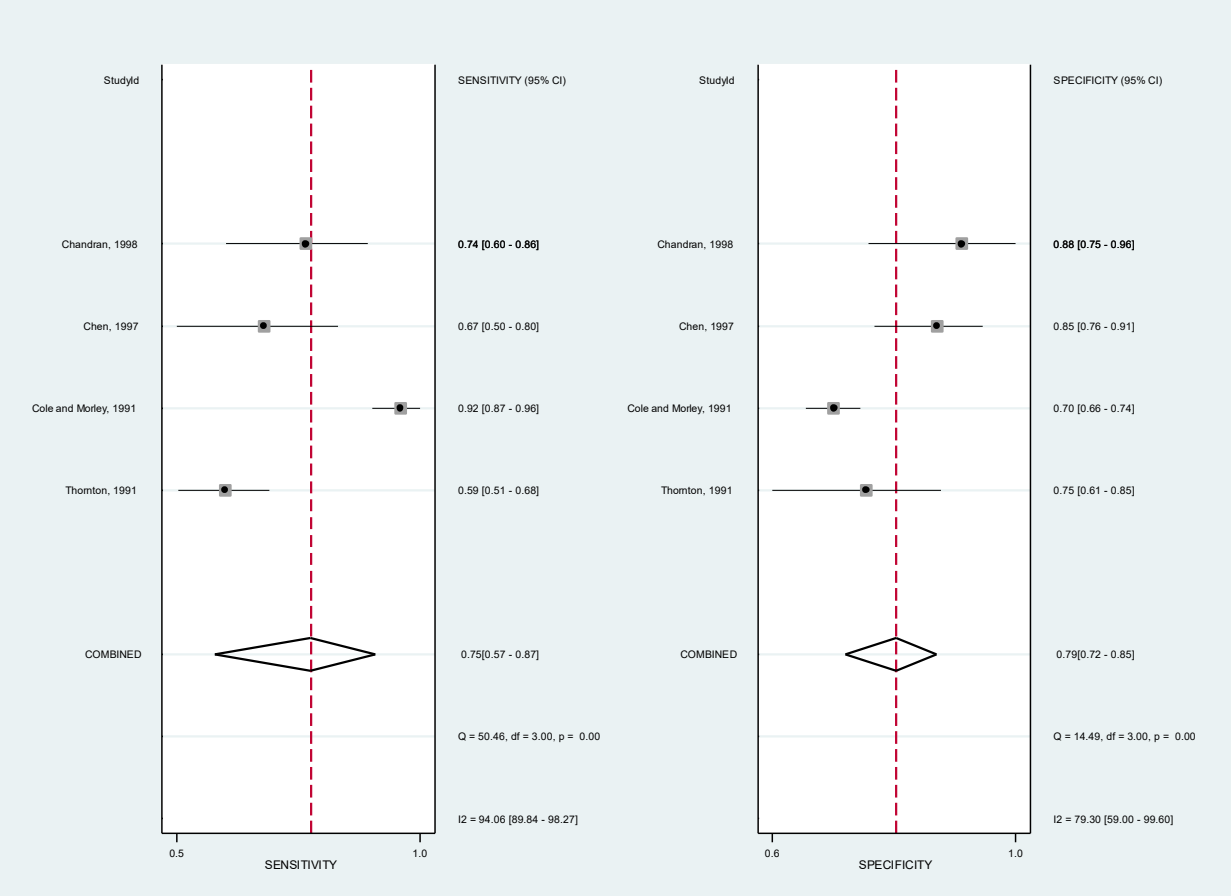
Bibliographic details	Participants	Scoring systems and assessment	Methods	Outcomes and results	Comments
Quantify Illness in Babies under 6 Months of Age, Journal of the Royal Statistical Society. Series A (Statistics in Society), 154, 287-304, 1991 **Ref Id** 1267924 **Country/ies where the study was carried out** **Study type** **Aim of the study** For details see Morley 1991 **Study dates** **Source of funding**	**Inclusion Criteria** **Exclusion Criteria**				

NGA: National Guideline Alliance; SD: standard deviation

Appendix E – Forest plots

Forest plots for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

Figure 2: Forest plot of Baby Check scoring system for grade 1 and 2 illness (score 13 or more) assessed in secondary care



Appendix F – GRADE tables

GRADE tables for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

Table 5: Clinical evidence profile for Baby Check scoring system in grading severity of illness in secondary care

Scoring system	No of studies	No of participants	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality of evidence (GRADE)	LR+ (95% CI)	LR- (95% CI)
Baby Check for grade 1 illness (score 20 or more)	1 (Chandran 1998)	90	Sensitivity = 1.00 (0.54 to 1.00)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Very serious imprecision[2]	VERY LOW	5.25 (3.38 to 8.16)	0.00
			Specificity = 0.81 (0.71 to 0.89)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
	1 (Chen 1997)	134	Sensitivity = 0.8 (0.44 to 0.97)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Very serious imprecision[2]	VERY LOW	12.4 (5.93 to 25.95)	0.21 (0.06 to 0.74)
			Specificity = 0.94 (0.88 to 0.97)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
	1 (Thornton 1991)[4]	193	Sensitivity = 0.46 (0.35 to 0.56)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	3.48 (2.01 to 6.05)	0.62 (0.51 to 0.76)
			Specificity = 0.87 (0.79 to 0.93)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
Baby Check for grade 1 and 2 illness (score 13 or more)	4[5,6]	1,126	Sensitivity= 0.75 (0.57 to 0.87)	No serious risk of bias	Very serious[7]	Serious indirectness[1]	Serious imprecision[3]	VERY LOW	2.82 (2.47 to 3.22)	0.33 (0.28 to 0.40)
			Specificity= 0.79 (0.72 to 0.85)	No serious risk of bias	Serious[8]	Serious indirectness[1]	Serious imprecision[3]	VERY LOW		
Baby Check for grade 2 illness	1 (Chandran 1998)	90	Sensitivity= 0.37 (0.22 to 0.53)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	5.98 (1.86 to 19.22)	0.68 (0.53 to 0.86)

(score 13-19)			Specificity= 0.94 (0.83 to 0.99)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
	1 (Chen 1997)	134	Sensitivity= 0.33 (0.22 to 0.45)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	2.73 (1.40 to 5.29)	0.72 (0.55 to 0.96)
			Specificity= 0.84 (0.72 to 0.92)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Very serious imprecision[2]	VERY LOW		
	1 (Thornton 1991)[9]	193	Sensitivity= 0.23 (0.11 to 0.38)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	1.13 (0.6 to 2.12)	0.97 (0.81 to 1.16)
			Specificity= 0.80 (0.73 to 0.86)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
Baby Check for grade 3 illness (score 8-12)	1 (Chen 1997)	134	Sensitivity= 0.33 (0.22 to 0.45)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	2.01 (1.04 to 3.86)	0.80 (0.66 to 0.98)
			Specificity= 0.84 (0.72 to 0.92)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Very serious imprecision[2]	VERY LOW		
	1 (Cole and Morley 1991)	709	Sensitivity= 0.28 (0.22 to 0.34)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	1.61 (1.21 to 2.13)	0.87 (0.80 to 0.95)
			Specificity= 0.83 (0.79 to 0.86)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE		
	1 (Thornton 1991)[10]	193	Sensitivity= 0.19 (0.08 to 0.33)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	1.33 (0.63 to 2.79)	0.95 (0.81 to 1.11)
			Specificity= 0.86 (0.79 to 0.91)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
Baby Check for grade 4 illness (score 0-7)	1 (Chen 1997)	134	Sensitivity= 0.89 (0.67 to 0.99)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Very serious imprecision[2]	VERY LOW	2.51 (1.88 to 3.35)	0.16 (0.04 to 0.61)
			Specificity= 0.64 (0.55 to 0.73)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE		
	1 (Cole and Morley 1991)	709	Sensitivity= 0.62 (0.56 to 0.67)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	No serious imprecision	MODERATE	4.55 (3.51 to 5.91)	0.44 (0.38 to 0.51)

			Specificity= 0.86 (0.83 to 0.9)	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
	1 (Thornton 1991)	193	Sensitivity= 0.92 (0.62 to 1.00)[11]	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Very serious imprecision[2]	VERY LOW	2.91 (2.21 to 3.83)[11]	0.12 (0.02 to 0.80)[11]
			Specificity= 0.69 (0.61 to 0.75) [11]	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		
			Sensitivity= 0.82 (0.57 to 0.96)[12]	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	VERY LOW	2.68 (1.96 to 3.67)[12]	0.25 (0.09 to 0.71)[12]
			Specificity= 0.69 (0.62 to 0.76) [12]	No serious risk of bias	No serious inconsistency	Serious indirectness[1]	Serious imprecision[3]	LOW		

CI: confidence interval; GRADE: Grading of Recommendations, Assessment, Development and Evaluation; Grade 1 illness: serious illness that requires hospital treatment; Grade 2 illness: infants need to be admitted to hospital for observation due to uncertainty about the severity illness; Grade 3 illness: infants that need careful observation and treatment. Could be managed at home by a capable mother; Grade 4 illness: infants are well and could be managed at home by any mother; LR+: positive likelihood ratio; LR-: negative likelihood ratio

[1]Quality of evidence downgraded by 1 as the population is indirect (babies up to 6 months of age included)

[2]Quality of evidence downgraded by 2 as the 95% confidence interval crosses 2 MID thresholds

[3]Quality of evidence downgraded by 1 as the 95% confidence interval crosses 1 MID threshold

[4]Data from consultant A in Thornton 1991 used in the analysis, sensitivity analysis was conducted using data from consultant B (sensitivity 0.63; specificity 0.86), difference in results doesn't cross any MID thresholds and wouldn't change overall conclusions.

[5]See corresponding forest plot for studies contributing to this outcome

[6]Data from consultant A in Thornton 1991 used in the meta-analysis, sensitivity analysis was conducted using data from consultant B in the meta-analysis (sensitivity 0.77; specificity 0.80), difference in results doesn't cross any MID thresholds and wouldn't change overall conclusions.

[7]Quality of evidence downgraded by 2 as the heterogeneity was very serious (I^2 statistic >80%)

[8]Quality of evidence downgraded by 1 as the heterogeneity was serious (I^2 statistic >50%)

[9]Data from consultant A in Thornton 1991 used in the analysis, sensitivity analysis was conducted using data from consultant B (sensitivity 0.24; specificity 0.80), difference in results doesn't cross any MID thresholds and wouldn't change overall conclusions.

[10]Data from consultant A in Thornton 1991 used in the analysis, sensitivity analysis was conducted using data from consultant B (sensitivity 0.16; specificity 0.85), difference in results doesn't cross any MID thresholds and wouldn't change overall conclusions.

[11]Data from consultant A in Thornton 1991

[12]Data from consultant B in Thornton 1991

Table 6: Clinical evidence profile for Baby Check scoring system in grading severity of illness in the community

Scoring system	No of studies	No of participants	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality of evidence (GRADE)	LR+ (95% CI)	LR- (95% CI)
Baby Check for grade 1		298	Sensitivity = NC	NA	NA	NA	NA	NA	NC	NC

and 2 illness (score 13 or more)	1 (Cole and Morley 1991)		Specificity = 0.99 (0.97 to 1.00)	Serious risk of bias[1]	No serious inconsistency	Serious indirectness[2]	No serious imprecision	LOW		
Baby Check for grade 3 illness (score 8-12)	1 (Cole and Morley 1991)	298	Sensitivity = 0.38 (0.085 to 0.76)	Serious risk of bias[1]	No serious inconsistency	Serious indirectness[2]	Serious imprecision[3]	VERY LOW	36.25 (8.61 to 152.68)	0.63 (0.37 to 1.08)
			Specificity = 0.99 (0.97 to 1.00)	Serious risk of bias[1]	No serious inconsistency	Serious indirectness[2]	No serious imprecision	LOW		
Baby Check for grade 4 illness (score 0-7)	1 (Cole and Morley 1991)	298	Sensitivity = 0.99 (0.97 to 1.00)	Serious risk of bias[1]	No serious inconsistency	Serious indirectness[2]	No serious imprecision	LOW	2.63 (1.07 to 6.43)	0.02 (0.01 to 0.07)
			Specificity = 0.63 (0.25 to 0.92)	Serious risk of bias[1]	No serious inconsistency	Serious indirectness[2]	Very serious imprecision[4]	VERY LOW		

CI: confidence interval; GRADE: Grading of Recommendations, Assessment, Development and Evaluation; Grade 1 illness: serious illness that requires hospital treatment; Grade 2 illness: infants need to be admitted to hospital for observation due to uncertainty about the severity illness; Grade 3 illness: infants that need careful observation and treatment. Could be managed at home by a capable mother; Grade 4 illness: infants are well and could be managed at home by any mother; LR+: positive likelihood ratio; LR-: negative likelihood ratio; NA: not applicable; NC: not calculable

[1]Quality of evidence downgraded by 1 due to serious risk of bias (research nurse's grading as reference standard)

[2]Quality of evidence downgraded by 1 as the population is indirect (babies up to 6 months of age included)

[3]Quality of evidence downgraded by 1 as the 95% confidence interval crosses 1 MID threshold

[4]Quality of evidence downgraded by 2 as the 95% confidence interval crosses 2 MID thresholds

Table 7: Clinical evidence profile for Baby Check scoring system in grading severity of illness in a theoretical community cohort

Scoring system	No of studies	No of participants	Effect size (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality of evidence (GRADE)	LR+ (95% CI)	LR- (95% CI)
Baby Check for grade 1 and 2 illness (score 13 or more)	1 (Cole and Morley 1991)	10,000 theoretical cohort	Sensitivity = 0.93 (0.82 to 0.99)	Serious risk of bias[1]	No serious inconsistency	Very serious indirectness[2]	Serious imprecision	VERY LOW	64.14 (53.67 to 76.73)	0.07 (0.02 to 0.2)
			Specificity = 0.99 (0.98 to 1.00)	Serious risk of bias[1]	No serious inconsistency	Very serious indirectness[2]	No serious imprecision	VERY LOW		
Baby Check for grade 3 illness (score 8-12)	1 (Cole and Morley 1991)	10,000 theoretical cohort	Sensitivity = 0.36 (0.31 to 0.41)	Serious risk of bias[1]	No serious inconsistency	Very serious indirectness[2]	No serious imprecision	VERY LOW	28.49 (22.6 to 35.81)	0.65 (0.60 to 0.70)
			Specificity = 0.99 (0.985 to 0.99)	Serious risk of bias[1]	No serious inconsistency	Very serious indirectness[2]	No serious imprecision	VERY LOW		

Baby Check for grade 4 illness (score 0-7)	1 (Cole and Morley 1991)	10,000 theoretical cohort	Sensitivity = 0.98 (0.98 to 0.99)	Serious risk of bias[1]	No serious inconsistency	Very serious indirectness[2]	No serious imprecision	VERY LOW	3.25 (2.79 to 3.79)	0.02 (0.02 to 0.03)
			Specificity = 0.70 (0.65 to 0.74)	Serious risk of bias[1]	No serious inconsistency	Very serious indirectness[2]	No serious imprecision	VERY LOW		

CI: confidence interval; GRADE: Grading of Recommendations, Assessment, Development and Evaluation; Grade 1 illness: serious illness that requires hospital treatment; Grade 2 illness: infants need to be admitted to hospital for observation due to uncertainty about the severity illness; Grade 3 illness: infants that need careful observation and treatment. Could be managed at home by a capable mother; Grade 4 illness: infants are well and could be managed at home by any mother; LR+: positive likelihood ratio; LR-: negative likelihood ratio.

[1]Quality of evidence downgraded by 1 due to serious risk of bias (research nurse's grading as reference standard)

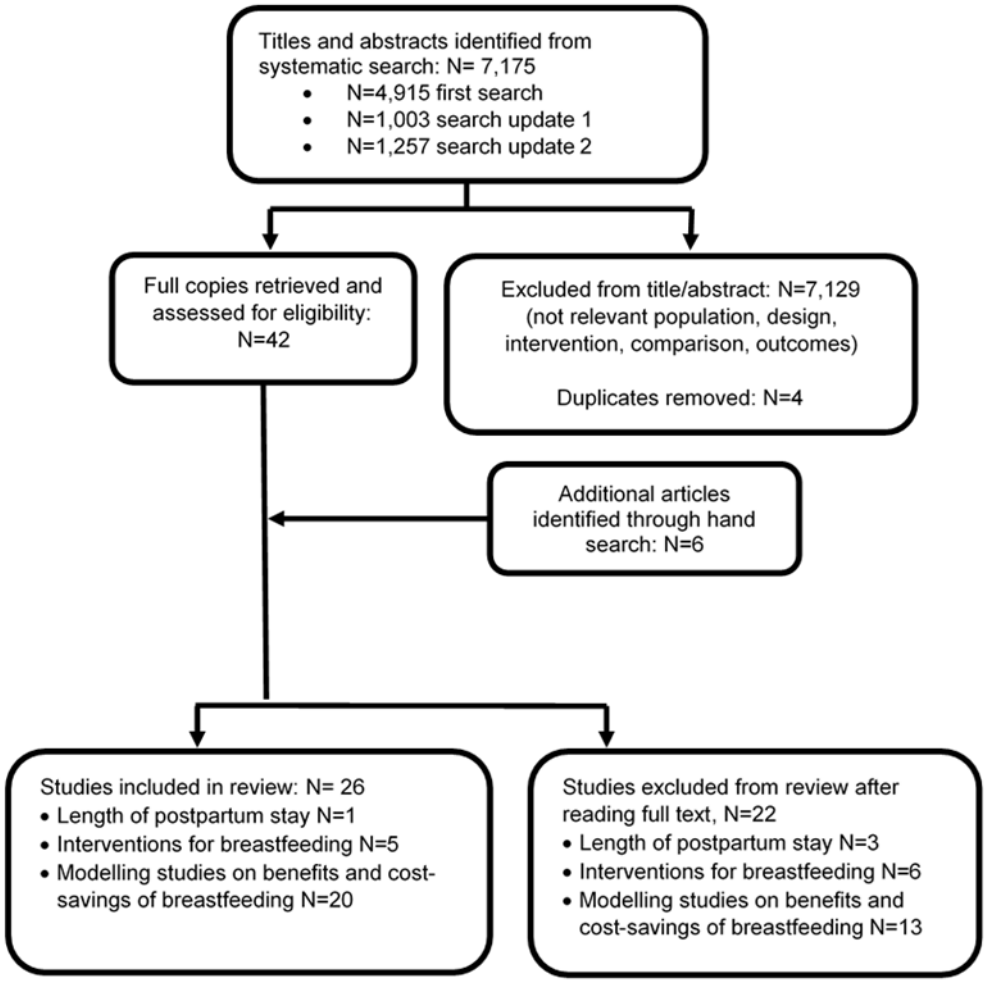
[2]Quality of evidence downgraded by 2 as the population is very indirect (babies up to 6 months of age included and theoretical cohort extrapolated from 2 cohorts in the study based on assumptions)

Appendix G – Economic evidence study selection

Economic evidence study selection for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

A global health economics search was undertaken for all areas covered in the guideline. Figure 2 shows the flow diagram of the selection process for economic evaluations of postnatal care interventions, including modelling studies on the benefits and cost-savings of breastfeeding.

Figure 2. Flow diagram of selection process for economic evaluations of postnatal care interventions and modelling studies on the benefits and cost-savings of breastfeeding



Appendix H – Economic evidence tables

Economic evidence tables for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

No economic evidence was identified that was applicable to this review question.

Appendix I – Economic evidence profiles

Economic evidence profiles for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

No economic evidence was identified that was applicable to this review question.

Appendix J – Economic analysis

Economic analysis for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

No economic evidence was identified that was applicable to this review question.

Appendix K – Excluded studies

Excluded studies for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

Clinical studies

Table 7: Excluded studies and reasons for their exclusion

Study	Reason for exclusion
Arora, R., Mahajan, P., Evaluation of child with fever without source: Review of literature and update, Pediatric Clinics of North America, 60, 1049-1062, 2013	Study design not of interest for review - literature review
Baker,M.D., Bell,L.M., Unpredictability of serious bacterial illness in febrile infants from birth to 1 month of age, Archives of Pediatrics and Adolescent Medicine, 153, 508-511, 1999	Population not of interest for review - fever was sole entry criterion for inclusion in study
Bell, D., Mac, A., Ochoa, Y., Gordon, M., Gregurich, M. A., Taylor, T., The Texas Children's Hospital Pediatric Advanced Warning Score as a predictor of clinical deterioration in hospitalized infants and children: a modification of the PEWS tool, Journal of pediatric nursing, 28, e2-9, 2013	Population not of interest for review - infants and children, 24.7% of population aged <1 year old (no stratification for age ranges).
Berry, M. A., Shah, P. S., Brouillette, R. T., Hellmann, J., Predictors of mortality and length of stay for neonates admitted to children's hospital neonatal intensive care units, Journal of Perinatology, 28, 297-302, 2008	Population not of interest for review - neonates in neonatal intensive care units
Bilan, N., Galehgolab, B. A., Emadaddm, A., Shiva, S. H., Risk of mortality in pediatric intensive care unit, assessed by prism-III, Pakistan Journal of Biological Sciences, 12, 480-485, 2009	Population not of interest for review - neonates in neonatal intensive care units
Bonafide, C. P., Holmes, J. H., Nadkarni, V. M., Lin, R., Landis, J. R., Keren, R., Development of a score to predict clinical deterioration in hospitalized children, Journal of Hospital Medicine, 7, 345-349, 2012	Outcome modelled not of interest for review - clinical deterioration
Broughton,S.J., Berry,A., Jacobe,S., Cheeseman,P., Tarnow-Mordi,W.O., Greenough,A., The mortality index for neonatal transportation score: A new mortality prediction model for retrieved neonates, Pediatrics, 114, e424-e428, 2004	Outcome modelled not of interest for review - mortality
Broughton,S.J., Berry,A., Jacobe,S., Cheeseman,P., Tarnow-Mordi,W.O., Greenough,A., An illness severity score and neonatal mortality in retrieved neonates, European Journal of Pediatrics, 163, 385-389, 2004	Outcome modelled not of interest for review - mortality
Chamberlain, J. M., Patel, K. M., Ruttimann, U. E., Pollack, M. M., Pediatric risk of admission (PRISA): a measure of severity of illness for assessing the risk of hospitalization from the	Population not of interest for review - infants, children and adults (not stratified by age)

Study	Reason for exclusion
emergency department, Annals of Emergency Medicine, 32, 161-9, 1998	
Chen, C. K., Chen, S. J., Hwang, B., Evaluation of the baby check score in emergency room, Zhonghua Minguo xiao er ke yi xue hui za zhi [Journal], Zhonghua Minguo xiao er ke yi xue hui. 36, 187-191, 1995	Same study as Chen 1997, however Chen 1997 included as dataset is more comprehensive for analysis
Choi, K. M., Ng, D. K., Wong, S. F., Kwok, K. L., Chow, P. Y., Chan, C. H., Ho, J. C., Assessment of the Pediatric Index of Mortality (PIM) and the Pediatric Risk of Mortality (PRISM) III score for prediction of mortality in a paediatric intensive care unit in Hong Kong, Hong Kong Medical Journal, 11, 97-103, 2005	Population not of interest for review - infants and children in paediatric intensive care units
Cole, T. J., Gilbert, R. E., Fleming, P. J., Morley, C. J., Rudd, P. T., Berry, P. J., Baby Check and the Avon infant mortality study, Archives of Disease in Childhood, 66, 1077-8, 1991	Study design not of interest - retrospective case control study in infants with sudden, unexpected infant death
Cole, T. J., Thornton, A. J., Green, S. J., Morley, C. J., Field trials of Baby Check: A scoring system to quantify illness in babies under 6 months, Medical Informatics, 15, 261-268, 1990	Overview of the Baby Check field studies. No additional data of interested for included studies Morley 1991; Thornton 1991; and Cole 1991.
De Leon, A. L., Romero-Gutierrez, G., Valenzuela, C. A., Gonzalez-Bravo, F. E., Simplified PRISM III score and outcome in the pediatric intensive care unit, Pediatrics International, 47, 80-3, 2005	Population not of interest for review - infants and children in paediatric intensive care units
Dean, N. P., Fenix, J. B., Spaeder, M., Levin, A., Evaluation of a Pediatric Early Warning Score Across Different Subspecialty Patients, Pediatric Critical Care Medicine, 18, 655-660, 2017	Outcome modelled not of interest for review - clinical deterioration
Deerojanawong, J., Prapphal, N., Udomittipong, K., PRISM score and factors predicting mortality of patients with respiratory failure in the pediatric intensive care unit, Journal of the Medical Association of Thailand, 84 Suppl 1, S68-75, 2001	Outcome modelled not of interest for review - mortality
Gorelick,M.H., Alessandrini,E.A., Cronan,K., Shults,J., Revised Pediatric Emergency Assessment Tool (RePEAT): a severity index for pediatric emergency care, Academic Emergency Medicine, 14, 316-323, 2007	Population not of interest for review - infants and children from 0-19 years old, only 2.1% <29 days old (no stratification of results by age)
Gravel, J., Gouin, S., Amre, D., Bergeron, S., Lacroix, J., Evaluation of the pediatric risk of admission score in a pediatric emergency department, Annals of Emergency Medicine, 41, 630-638, 2003	Population not of interest for review: infants and children 0-19 years old (no age stratification)
Gray, J. E., Richardson, D. K., McCormick, M. C., Workman-Daniels, K., Goldmann, D. A., Neonatal therapeutic intervention scoring system: a therapy-based severity-of-illness index, Pediatrics, 90, 561-7, 1992	Population not of interest for review - 70% of population preterm.
Harsha, S. S., Archana, B. R., SNAPPE-II (score for neonatal acute physiology with perinatal extension-II) in predicting mortality and morbidity	Country not of interest for review - India (not classified as a world bank high-income country)

Study	Reason for exclusion
in NICU, Journal of Clinical and Diagnostic Research, 9, SC10-SC12, 2015	
Henderson, A. J., Garland, L., Warne, S., Bailey, L., Weir, P., Edees, S., Risk adjusted mortality of critical illness in a defined geographical region, Archives of Disease in Childhood, 86, 194-9, 2002	Outcome modelled not of interest for review - mortality
Hewson, P. H., Gollan, R. A., A simple hospital triaging system for infants with acute illness, Journal of Paediatrics and Child Health, 31, 29-32, 1995	Accuracy of prediction values reported without confidence intervals. Furthermore, raw data unavailable to construct 2 x 2 predictive accuracy tables.
Hewson,P., Poulakis,Z., Jarman,F., Kerr,J., McMaster,D., Goodge,J., Silk,G., Clinical markers of serious illness in young infants: a multicentre follow-up study, Journal of Paediatrics and Child Health, 36, 221-225, 2000	Accuracy of prediction values reported without confidence intervals. Furthermore, raw data unavailable to construct 2 x 2 predictive accuracy tables.
Justicia-Grande, A. J., Pardo-Seco, J., Cebey-Lopez, M., Vilanova-Trillo, L., Gomez-Carballa, A., Rivero-Calle, I., Puente-Puig, M., Curros-Novo, C., Gomez-Rial, J., Salas, A., Martinon-Sanchez, J. M., Redondo-Collazo, L., Rodriguez-Tenreiro, C., Martinon-Torres, F., Development and validation of a new clinical scale for infants with acute respiratory infection: The resvinet scale, PLoS ONE, 11 (6) (no pagination), 2016	No outcomes of interest - no outcomes on model performance or predictive accuracy
Kamath-Rayne, B. D., MacGuire, E. R., McClure, E. M., Goldenberg, R. L., Jobe, A. H., Clinical algorithms for the identification of sick newborns in community-based settings, Acta Paediatrica, 101, 344-51, 2012	Country not of interest for review - included studies not classified as world bank high-income countries
Kanter,R.K., Edge,W.E., Caldwell,C.R., Nocera,M.A., Orr,R.A., Pediatric mortality probability estimated from pre-ICU severity of illness, Pediatrics, 99, 59-63, 1997	Outcome modelled not of interest for review - Mortality
Lee,S., Aziz,K., Dunn,M., Clarke,M., Kovacs,L., Ojah,C., Ye,X., Transport risk index of physiologic stability, version II (TRIPS-II): A simple and practical neonatal illness severity score, American Journal of Perinatology, 30, 395-400, 2013	Population not of interest for review - neonates in neonatal intensive care units
Mahale,R., Dutta,S., Ahluwalia,J., Kishore,S.S., Narang,A., Baseline illness severity does not alter accuracy of neonatal sepsis screen, American Journal of Perinatology, 27, 327-332, 2010	Country not of interest for review - India (not classified as a world bank high-income country)
Mahieu, L. M., De Dooy, J. J., Cossey, V. R., Goossens, L. L., Vrancken, S. L., Jespers, A. Y., Vandeputte, C. T., De Muynck, A. O., Internal and external validation of the NOSEP prediction score for nosocomial sepsis in neonates, Critical Care Medicine, 30, 1459-1466, 2002	Outcome modelled not of interest for review - nosocomial sepsis
Maulen-Radovan, I., Gutierrez Castrellon, P., Zaldo Rodriguez, R., Martinez Natera, O., PRISM score evaluation to predict outcome in	Outcome modelled not of interest for review - mortality

Study	Reason for exclusion
pediatric patients on admission at an emergency department, Archives of Medical Research, 27, 553-8, 1996	
Mittal, K., Gupta, V., Khanna, P., Kaushik, J. S., Sharma, A., Evaluation of Integrated Management of Neonatal and Childhood Illness (IMNCI) algorithm for diagnosis and referral in under-five children, Indian Journal of Pediatrics, 81, 797-799, 2014	Country not of interest for review - India (not classified as a world bank high-income country)
Morley, C. J., Thornton, A. J., Green, S. J., Cole, T. J., Field trials of the Baby Check score card in general practice, Archives of Disease in Childhood, 66, 111-114, 1991	No outcomes of interest for review - no model performance or predictive accuracy outcomes
Muktan, D., Singh, R. R., Bhatta, N. K., Shah, D., Neonatal mortality risk assessment using SNAPPE-II score in a neonatal intensive care unit, BMC Pediatrics, 19 (1) (no pagination), 2019	Outcome modelled not of interest for review - mortality
Nigrovic, L. E., Mahajan, P. V., Blumberg, S. M., Browne, L. R., Linakis, J. G., Ruddy, R. M., Bennett, J. E., Rogers, A. J., Tzimenatos, L., Powell, E. C., Alpern, E. R., Casper, T. C., Ramilo, O., Kuppermann, N., Febrile Infant Working Group of the Pediatric Emergency Care Applied Research, Network, The Yale Observation Scale Score and the Risk of Serious Bacterial Infections in Febrile Infants, Pediatrics, 140, 2017	Outcome modelled not of interest for review - serious bacterial infection
Orr, R. A., Venkataraman, S. T., Cinoman, M. I., Hogue, B. L., Singleton, C. A., McCloskey, K. A., Pretransport Pediatric Risk of Mortality (PRISM) score underestimates the requirement for intensive care or major interventions during interhospital transport, Critical Care Medicine, 22, 101-107, 1994	Outcome modelled not of interest for review - mortality
Orr, R. A., Venkataraman, S. T., McCloskey, K. A., Janosky, J. E., Dragotta, M., Bills, D., King, W. D., Measurement of pediatric illness severity using simple pretransport variables, Prehospital Emergency CarePrehosp Emerg Care, 5, 127-33, 2001	Outcome modelled not of interest for review - mortality
Pollack, M. M., Patel, K. M., Ruttimann, U. E., PRISM III: An updated pediatric risk of mortality score, Critical Care Medicine, 24, 743-752, 1996	Outcome modelled not of interest for review - mortality
Pollack,M.M., Patel,K.M., Ruttimann,U.E., The Pediatric Risk of Mortality III--Acute Physiology Score (PRISM III-APS): a method of assessing physiologic instability for pediatric intensive care unit patients, Journal of Pediatrics, 131, 575-581, 1997	Outcome modelled not of interest for review - mortality
Ponce-Ponce De Leon, A. L., Romero-Gutierrez, G., Aldana, C. V., Gonzalez-Bravo, F. E., Simplified PRISM III score and outcome in the pediatric intensive care unit, Pediatrics International, 47, 80-83, 2005	Outcome modelled not of interest for review - mortality

Study	Reason for exclusion
Radfar, M., Hashemieh, M., Fallahi, M., Masihi, R., Utilization of SNAP II and SNAPPE II Scores for Predicting the Mortality Rate Among a Cohort of Iranian Newborns, Archives of Iranian Medicine, 21, 153-157, 2018	Population not of interest for review - neonates in neonatal intensive care units
Richardson, D. K., Corcoran, J. D., Escobar, G. J., Lee, S. K., SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores, Journal of Pediatrics, 138, 92-100, 2001	Population not of interest for review - neonates in neonatal intensive care units
Richardson, D. K., Shah, B. L., Frantz, I. D., 3rd, Bednarek, F., Rubin, L. P., McCormick, M. C., Perinatal risk and severity of illness in newborns at 6 neonatal intensive care units, American Journal of Public Health, 89, 511-6, 1999	Population not of interest for review - preterm babies
Roukema, J., Steyerberg, E. W., van Meurs, A., Ruige, M., van der Lei, J., Moll, H. A., Validity of the Manchester Triage System in paediatric emergency care, Emergency Medicine Journal, 23, 906-10, 2006	Population not of interest for review - infants and children 0-16 years old (no age stratifications)
Slater, A., Shann, F., Anzics Paediatric Study Group, The suitability of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand, Pediatric Critical Care Medicine, 5, 447-54, 2004	Outcome modelled not of interest for review - mortality
Taori, R. N., Lahiri, K. R., Tullu, M. S., Performance of PRISM (Pediatric Risk of Mortality) score and PIM (Pediatric Index of Mortality) score in a tertiary care Pediatric ICU, Indian Journal of Pediatrics, 77, 267-271, 2010	Country not of interest for review - India (not classified as a world bank high-income country)
Thomson, H., Ross, S., Wilson, P., McConnachie, A., Watson, R., Randomised controlled trial of effect of Baby Check on use of health services in first 6 months of life, British Medical Journal, 318, 1740-1744, 1999	Study design not of interest for review - randomised controlled trial
Thornton, A. J., Morley, C. J., Green, S. J., Cole, T. J., Walker, K. A., Bonnett, J. M., Field trials of the Baby Check score card: Mothers scoring their babies at home, Archives of Disease in Childhood, 66, 106-110, 1991	No outcomes of interest for review - no model performance or predictive accuracy outcomes
Wiebe, R. A., Rosen, L. M., Triage in the emergency department, Emergency Medicine Clinics of North America, 9, 491-505, 1991	Study design not of interest for review: literature review
Young Infants Clinical Signs Study, Group, Clinical signs that predict severe illness in children under age 2 months: a multicentre study, Lancet, 371, 135-42, 2008	Country not of interest for review - included studies are not classified as world bank high-income countries
Zobel, G., Kuttnig, M., Grubbauer, H. M., Rodl, S., Evaluation of clinical scoring systems in critically ill infants and children, Clinical Intensive Care, 1, 202-6, 1990	Population not of interest for review: infants and children (no age stratification)

Economic studies

No economic evidence was identified for this review.

Appendix L – Research recommendations

Research recommendations for review question: Which scoring systems are accurate in identifying or predicting illness severity in babies?

No research recommendations were made for this review question.