Research Repository UCD

Title	Instruments measuring practitioner performance of the complete examination and screening of neonates: A systematic review
Authors(s)	Greene, Elizabeth, Hegarty, Josephine, O'Connell, Rhona, Connaughton, Breda, Coveney, Kate, Hardie, Philip, Horton, Sally, Szafranska, Marcelina, Murphy, Margaret
Publication date	2023-04
Publication information	Greene, Elizabeth, Josephine Hegarty, Rhona O'Connell, Breda Connaughton, Kate Coveney, Philip Hardie, Sally Horton, Marcelina Szafranska, and Margaret Murphy. "Instruments Measuring Practitioner Performance of the Complete Examination and Screening of Neonates: A Systematic Review." Wiley, April 2023. https://doi.org/10.1111/jpc.16389.
Publisher	Wiley
Item record/more information	http://hdl.handle.net/10197/26919
Publisher's version (DOI)	10.1111/jpc.16389

Downloaded 2026-05-02 00:27:59

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

Instruments measuring practitioner performance of the complete examination and screening of neonates: a systematic review

Abstract

Aim: The complete examination and screening of the neonate is a recommended assessment of neonatal well-being conducted by appropriately trained medical, midwifery, and nursing personnel at specific intervals during the first six-weeks post-birth. Our aim was to identify and critically evaluate instruments that measure practitioner performance of this important assessment of neonatal health.

Methods: Using the **CO**nsensus-based **S**tandards for the selection of health **M**easurement **In**struments (COSMIN) methodology, a systematic review was undertaken.

Results: Four studies were identified as suitable for data extraction and analysis. This paper briefly describes the four instruments, discusses and compares the COSMIN analysis and ratings of each instrument. A recommendation for the instrument identified as the most suitable to measure practitioner performance is provided.

Conclusion: Most instruments were designed by educators to measure the performance of practitioners developing competence in the complete examination and screening of the neonate. Further development and piloting of instruments designed to measure the performance and continuing competence of qualified practitioners of the newborn examination are required.

Key Words: competence, neonates, physical examination, practitioners, COSMIN

KEY POINTS

- A systematic review was performed using the COSMIN methodology to evaluate instruments that measured practitioner performance of the complete physical examination and screening of the neonate.
- Four studies that described two peer-assessment and two self-assessment instruments were analysed. The peer-assessment tools achieved higher overall ratings in the ten COSMIN standards.
- One instrument is recommended for use by clinicians, though further development and psychometric testing is advised.

1 Introduction

It is recommended that all newly born infants undergo a formal, full and detailed head-to-toe physical examination which includes screening for a number of specific congenital disorders between 48 and 72 hours of birth and again between six to eight weeks following birth[1,2,3,4]. Traditionally, the **C**omplete (detailed) physical **E**xamination and **S**creening **o**f the **N**eonate (CESoN) was performed largely by medical practitioners, with healthy term gestation babies typically being examined by a junior doctor. The training that medical practitioners have received to perform newborn examination & screening has been identified as being informal[5,6], and sometimes consisting of a 'see one, do one, teach one' approach[7].

This is changing and increasingly midwives and nurses have completed postgraduate training to perform this vital aspect of neonatal care. Indeed, some undergraduate midwifery curricula now include education and training to perform the CESoN[8]. To date, although a significant number of qualified midwives undertake this postgraduate education and gain certification, many do not continue to regularly perform CESoN[9-12]. In addition, following completion of postgraduate training in the CESoN, some midwives and nurses can struggle to examine and screen babies on a

regular basis due to a variety of factors including workforce shortages and redeployment [7,11-12].

Maintenance of competence to assure patient/client safety is a core principle for healthcare practitioners[13-16] and health care professionals may be required to demonstrate continuing competence in the CESoN. Some key performance indicators have been established for certain elements of the CESoN[2] i.e., timelines for completing the examination and referring babies that have a positive hip screen, but there is no firm consensus in the available literature regarding either a set number of examinations or learning activities that all qualified practitioners must complete to maintain competence in this area of practice[17]. Hence, the need for an objective measurement tool to provide evidence of the practitioners' retention of knowledge and skills.

In advance of completing this systematic review, a thorough scoping of the literature [unpublished] was conducted by the lead author. At that stage, no practitioner performance measurement tools for the Complete Examination and Screening of the Neonate were identified. If identified, such a measurement tool could assist practitioners to measure their performance of the CESoN against current best practice and then focus continuing professional development activities accordingly.

The aim of this review is to identify, describe, critically appraise, compare, and summarize the quality of the measurement properties of tools or instruments which measure practitioner performance of the detailed or complete physical examination and screening of the neonate (CESoN).

The primary objective is to provide evidence-based guidance to inform the selection of an instrument which measures practitioner performance of the CESoN using the **CO**nsensus-based **S**tandards for the selection of health **M**easurement **In**struments (COSMIN) standards[18-20]. The COSMIN methodology includes several key stages to

allow for rating and comparison of the tools. These include: sourcing of instruments; describing the characteristics of the included studies and instruments; determining the reliability, validity, responsiveness of each instrument using COSMIN risk of bias checklist; determining the quality of the instrument through rating the development and piloting of each instrument; combining risk of bias ratings to determine an overall rating on reliability, validity and responsiveness of each instrument and determining a GRADE rating for the overall quality of the instrument.

2 Methods

2.1 Identification and retrieval of papers to source instruments

A PICO[21,22] (Patient/Person/Problem; Intervention; Comparator; Outcome) framework was used to identify the search terms and subject headings which were used for title and abstract searching in the following databases: MEDLINE (via PubMed); EMBASE; CINAHL (via EBSCO); ERIC International (via ProQuest); British Education Index (via EBSCO); ProQuest Dissertation and Theses; The Cochrane Central Register of Controlled Trials (CENTRAL); Web of Science, CABI Global Health, Open Grey.

The search commenced with papers published in 2000 as this timeline coincided with the introduction of the first graduate educational programmes for midwives, nurses and other non-medical healthcare practitioners in the Complete Examination and Screening of the Neonate. Databases were initially searched from January 2000, to 17th September 2019 with searches repeated to update the search taking in the timeframe of September 2019 to 30th October 2021. All citations were exported by the lead author (EG) to EndNote© (x8)[23] and then to Covidence©[24]. Initial title and abstract screening and subsequent full text screening was done by eight authors (EG, ROC, MM, BC, KC, PH, SH, MS). This systematic review was accepted for registration on PROSPERO[25]- registration number: CRD42020151657.

2.2 Eligibility Criteria

Eligibility criteria were any human studies or articles published from January 2000 to October 2021, in English. The inclusion criteria were any type of study or structured review that described the development of, or use of an instrument, which measured practitioner performance of or competence in the complete examination and screening of the live-born human neonate. Instruments using any assessment or measurement method were included (e.g., self-reported, peer-assessment). Also included were papers with an instrument measuring practitioner performance of the neonatal examination and screening that included most of the fundamental elements of the complete examination and screening of the neonate, as per National Institute of Health and Care Excellence-NICE guideline NG194[1] and Public Health England (see **Figure 1**) guidance regarding newborn and infant physical examination-NIPE screening[2].

Papers were excluded if they pertained to physical examination and screening of an infant more than 8 weeks after birth. Papers were excluded if the instrument incorporated a neonatal examination and screening that did not include most of the fundamental elements of the complete examination and screening[1,2], ultrasound examination and screening of the foetus; newborn hearing screening; metabolic screening e.g. the newborn blood spot screening programme; tandem mass spectrometry.

Three reviewers (EG, ROC, MM) independently performed full text screening of papers. Any screening conflicts were resolved through discussion and consensus with a third author.

The seminal publications[18-20,26] regarding the COSMIN methodology for systematic reviews refer to Patient-Reported Outcome Measures or PROMs. Those authors suggest that their methodology for a systematic review of patient-reported outcome

measurement instruments can be adapted to a systematic review of performance-based outcome measures (PerBOMs)[18,27(pg.6)]. Where applicable, within the Risk of Bias (RoB) Checklist and content validity rating items, we substituted the term 'PerBOM' for 'PROM', the construct 'CESoN' for 'condition', and the population 'professional' for 'patient'. For this review, the term PerBOMs refers to the tools measuring practitioner performance of the complete examination and screening of the neonate being rated as per the COSMIN standards and taxonomy.

The content validity ratings are the most important measurement properties of a PerBOM[27(pg.16)]. There are three steps to the content validity evaluation process. The COSMIN Risk of Bias Checklist[28] begins with the content validity evaluations, with a rating to be assigned to the quality of the PerBOM development and piloting of each tool. A detailed user manual[29] provides descriptors for each item and rating. A 'worst score counts principle'[29(pg.15)] is applied for all COSMIN standards and rating items.

3 Results

3.1 Progress of papers through the review process

Figure 2 details the flow diagram of preferred reporting items for systematic reviews-PRISMA[30]. Following full text review of 212 papers, ten studies that described instruments that measured practitioner performance of the complete examination and screening of the neonate were identifed[10,31-39]. Of these ten studies, four[10,31-33] included the instruments described and were immediately identified as suitable for inclusion and data extraction. The authors of six studies[34-39] were contacted for further information regarding the instruments described in their papers to decide regarding inclusion for data extraction. The authors of three studies[34-36] did provide further information which indicated that those tools did not meet the inclusion criteria. The author of one study[37] could not provide sufficient additional information regarding the tool described, so this study was excluded. The authors of the two remaining

studies[38,39] were contacted on multiple occasions, but no additional information was provided, therefore these studies could not be included for data extraction.

Four studies[10,31-33] describing four instruments to measure practitioner performance of the complete examination and screening of the neonate-CESoN were rated using the COSMIN Risk of Bias Checklist[28]. Two studies[10,32] described peer-assessment instruments that measured practitioner performance of the CESoN, and two studies[31,33] described self-assessment instruments. The characteristics of the four instruments are detailed in **Table 1**, whilst the characteristics of the four included studies are outlined within **Table 2**. The Risk of Bias Checklist[29] permits each instrument to be assessed and rated as per the 10 measurement properties of the COSMIN Taxonomy[27] (**Table 3**).

3.2 Comparing results of self-assessment versus peer-assessment tools

Two distinct types of PerBOMs were identified in the full-text review and data extraction phase (**Table 1**). Practitioner performance of the complete examination and screening of the neonate was measured in two studies[31,33] by the practitioners completing a self-evaluation using a questionnaire-type instrument. The first two PerBOMs were: the Perceived Self-Efficacy Tool for Newborn Assessment (PSETNA)[31], and the Preventative Child Healthcare Nurses Competence Questionnaire Tool (PCHNCQ)[33].

In the remaining two studies[10,32] practitioners were evaluated performing the complete examination and screening of the neonate by their peers utilising one of two distinct evaluative PerBOMs: the video analysis proforma- baby examination study (VAPBES)[9], and the newborn examination checklist (NEC)[32].

The results of the COSMIN analysis shall be presented clearly distinguishing between the peer-assessment PerBOM tools versus the self-assessment PerBOM tools.

3.3 Content Validity – Quality of PerBOM development

The quality of PerBOM development is displayed in **Table 4**. As can be seen in Summary of Findings in **Table 5**, the self-assessment PerBOMs[31,33] and the peer-assessment[10,32] PerBOMs received overall 'sufficient' ratings for content validity. Although both the VAPBES[10] and the NEC[32] peer-assessment tools had overall (+) sufficient ratings for comprehensibility, the risk of bias checklist demonstrated that neither tool was adequately evaluated by the target populations using qualitative methods regarding either comprehensibility or comprehensiveness. COSMIN consider qualitative data gathered from the piloting phase of a PerBOM to be an integral element of the PerBOM development process to assure overall content validity. Both the PSETNA[31] and PCHNCQ[33] self-assessment tools had an overall '(?)' indeterminate rating for content validity that arose from either 'doubtful' or 'inadequate' ratings, particularly during the development phase of the PCHNCQ.

The VAPBES[10] PerBOM scored the highest overall COSMIN ratings for content validity. This comprehensive peer-assessment tool contains 62-items using a combination of response options (7-point and 4-point Likert scales, categorical/nominal measurements, open-ended questions).

3.4 Results of other measurement properties

Once content validity has been rated, each PerBOM is rated for the remaining COSMIN standards in boxes 3 to 10 of the Risk of Bias Checklist, as outlined in **Table 3**. For this review, cross-cultural validity was not applicable (N/A), as none of the tools had been translated from another language or jurisdiction.

3.5 COSMIN summary of findings

COSMIN recommend that where possible, for the structural validity, internal consistency, reliability, and measurement standards, that the results of multiple content validity studies for a PerBOM be pooled for meta-analysis[28(pg.31)]. This pooling

of data from multiple studies could be of, for example, the Cronbach's alpha scores of one unidimensional scale, or the weighted kappa of ordinal scores. It was not possible for a meta-analysis of pooled data to be performed within this systemic review, as there was only one published study for each of the four PerBOMs. Therefore, we have followed the COSMIN guidance that the results of the Risk of Bias ratings for each PerBOM and overall ratings are qualitatively pooled and reported in the discussion section. Overall ratings were described as 'sufficient' (+), 'insufficient' (-), indeterminant (?), or not reported (NR) (Table 5). The individual ratings represent a summary of the risk of bias ratings with 'sufficient' (+) awarded for ≥85% of the items of the instrument fulfilling the criterion; 'insufficient' (-) awarded for <85% of the items of the instrument not fulfilling the criteria and indeterminant (?), awarded for either no/ not enough information available or the quality of study or part of the study being inadequate. Not reported was given if no details at all were reported on the items.

The overall ratings of the PerBOMs for each of the ten COSMIN standards of the taxonomy is provided (**Table 5**). Accompanying these are the ratings for the overall quality of evidence for each COSMIN standard. The overall quality of evidence uses a modified GRADE approach[18-20] with an overall quality of evidence rating system (GRADE) score of High, Moderate, Low, Very Low being awarded.

4 Discussion

The relevance of the results of this systematic review for practitioners of the complete examination and screening of the neonate was the frequent 'adequate', 'doubtful', or 'inadequate' ratings in the PerBOM development process (**Table 4**). The is determined to be partly linked with the nature, setting, and target population of the studies. The PSETNA[31], NEC[32], and PCHNCQ[33] tools were all developed for studies that evaluated an educational package or intervention related to attaining competence in the complete examination and screening of the neonate-CESoN, not

to aid re-validation or continuing professional development for registered practitioners.

Neither the PSETNA[31] or PCHNCQ[33] self-assessment instruments scored as highly across the ten COSMIN standards as the VAPBES[10] or NEC[32] peer-assessment instruments. These results do not preclude the suitability of self-assessment instruments to measure practitioner performance of the CESoN, but rather that multiple RoB indicators in the PerBOM development stages of the PSETNA[31] and PCHNCQ[33] were rated as either 'doubtful' or 'inadequate'. However, if these issues were rectified, a modified version of either self-assessment tool could be piloted and rated as 'sufficient' in a subsequent COSMIN analysis. The only PerBOM identified that was specifically developed to evaluate the quality of performance of qualified, experienced practitioners of the neonatal examination & screening was the VAPBES[10] instrument.

The objective of this review was to identify a suitable instrument to measure practitioner performance of the CESoN for the purposes of assuring continuing competence e.g., revalidation, continuing professional development. We have demonstrated that the VAPBES[10] peer-assessment instrument, while lengthy, includes tailored rating scales to evaluate the quality of various aspects of practitioner performance of the CESoN.

In summary, the VAPBES[10] peer-assessment instrument measuring practitioner performance of the complete examination and screening of the neonate is recommended for use in view of the overall 'sufficient' ratings across six of the ten COSMIN standards which evaluated four performance-based outcome measurement instruments[10,31-33] and the narrative synthesis findings which identified the VAPBES[10] as the only instrument designed to objectively and adequately measure the performance of qualified practitioners from medical and midwifery/nursing disciplines.

5 Conclusions

All the instruments we identified and evaluated had limitations, particularly in the development and content validity testing phase, which COSMIN considers the most important standard of an outcome-measurement instrument. Given that revalidation of continuing competence can be assured through a combination of self and peer evaluation methods, we recommend that easy to administer peer and self-assessment instruments designed for registered practitioners of the complete examination and screening of the neonate currently employed in the clinical environment are further developed and piloted.

There is growing recognition internationally of the need for legislation that mandates activities that contribute to revalidation of professional standards and maintenance of competence for healthcare professionals. The development of instruments that measure and encourage high-quality practitioner performance of essential clinical assessments simultaneously protects both the well-being of our healthcare professions and the clients we care for every day.

Acknowledgements: with gratitude to my doctoral supervisors in the School of W, University X. Sincere thanks also to my colleagues in the School of Y, University Z.

Complete list of abbreviations used: CESoN, NICE, NIPE, COSMIN, VAPBES, PSETNA, PCHNCQ, NEC.

Explanations of Abbreviations used: CESoN= Complete Examination and Screening of the Neonate, NICE= National Institute of Health and Care Excellence, NIPE= Newborn and Infant Physical Examination, COSMIN= COnsensus-based Standards for the selection of health Measurement INstruments, VAPBES= Video Analysis Proforma Baby Examination Study, PSETNA= Perceived Self-Efficacy Tool for Newborn Assessment, PCHNCQ= Preventative Child Healthcare Nurses Competence Questionnaire Tool, NEC= Newborn Examination Checklist.

Statements of conflicts of interests

The authors have no conflicts of interest to declare.

Statement of funding: Doctoral studies of the lead author are being funded by the School of Y, University Z.

References

1. National Institute of Health and Care Excellence- NICE. *Postnatal Care* NICE guideline [NG194]. 1.3 Postnatal care of the baby: NICE; 2021. p. 18-21. https://www.nice.org.uk/guidance/ng194/resources/postnatal-care-pdf-66142082148037
2. Public Health England- PHE. *Neonate and infant physical examination screening programme handbook*. Public Health England; 2021. https://www.gov.uk/government/publications/newborn-and-infant-physical-examination-programme-handbook/newborn-and-infant-physical-examination-screening-programme-handbook
3. HSE. The Newborn Clinical Examination Handbook. Health Service Executive; 2018. https://www.hse.ie/eng/about/who/healthwellbeing/our-priority-programmes/child-health-and-wellbeing/neonate%20exam.pdf
4. Paediatrics & Child Health Division of The Royal Australasian College of Physicians- RACP. *Examination of the Newborn.* https://media.starship.org.nz/racp-examination-of-the-newborn/examinationofthenewborn.pdf
5. Hayes J, Dave S, Rogers C, Quist-Therson E, Townsend J. A national survey in England of the routine examination of the newborn baby. *Midwifery*. 2003;19(4):277-84.
6. McKinnon K. G338(P) How do junior doctors learn the newborn and infant physical examination? *Archives of disease in childhood*. 2017;102(Suppl 1):A133-A.
7. McDonald S, Allan H, Brown A. Perceptions of changing practice in the examination of the newborn, from holistic to opportunistic. *British journal of midwifery*. 2012;20(11):786-91.

8. Yearley C, Rogers C, Jay A. Including the newborn physical examination in the pre-registration midwifery curriculum: National survey. *British journal of midwifery*. 2017;**25**(1):26-32
9. Rogers C, Yearley C, Jay A. Education provision for the newborn physical examination as a postregistration module: National survey. *British journal of midwifery*. 2017;**25**(2):88-92
10. Townsend J, Wolke D, Hayes J, Dave S, Rogers C, Bloomfield L, et al. Routine examination of the newborn: the EMREN study. Evaluation of an extension of the midwife role including a randomised controlled trial of appropriately trained midwives and paediatric senior house officers - Introduction. *Health technology assessment* (Winchester, England). 2004;**8**(14):1.
11. Lanlehin R, Noble H, McCourt C. How well do midwives use skills and knowledge in examining newborns? *British journal of midwifery*. 2011;**19**(11):687-91.
12. Steele D. Examining the newborn: why don't midwives use their skills? *British journal of midwifery*. 2007;15(12):748-52.
13. Nursing and Midwifery Board of Ireland- NMBI. *Scope of Nursing & Midwifery Practice Framework*. 2015. https://www.nmbi.ie/Standards-Guidance/Scope-of-Practice/Considerations-in-Determining-Scope/Competence.aspx
14. Irish Medical Council. *Guide to Professional Conduct and Ethics for Registered Medical Practitioners* (Amended). 66 Maintaining competence. 2019. p. 42-3. https://www.medicalcouncil.ie/news-and-publications/reports/guide-to-professional-conduct-and-ethics-for-registered-medical-practitioners-amended-.pdf
15. General Medical Council- United Kingdom. *Good Medical Practice*. Domain 1: Knowledge skills and performance. 2019. https://www.gmc-uk.org/ethical-

guidance/ethical-guidance-for-doctors/good-medical-practice/domain-1---knowledge-skills-and-performance#paragraph-7

16. Nursing and Midwifery Council- United Kingdom. *What is revalidation? What revalidation means for you*. Nursing and Midwifery Council- NMC 2021. https://www.nmc.org.uk/revalidation/overview/what-is-revalidation/
17. Public Health England- PHE. *Newborn and infant physical examination (NIPE) screening programme handbook*. 12 Training and maintenance of competency: Public Health England; 2021. https://www.gov.uk/government/publications/newborn-and-infant-physical-examination-programme-handbook/newborn-and-infant-physical-examination-screening-programme-handbook#training-and-maintenance-of-competency
18. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research*. 2018; **27**(5):1147-57.
19. Mokkink LB, de Vet, H C. W, de Vet, H C. W, Prinsen CAC, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*. 2018; **27**(5):1171-9.
20. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*. 2018; **27**(5):1159-70.
21. Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, Hirsh J, et al. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA : the journal of the American Medical Association*. 1992; **268**(17):2420-5. doi:10.1001/jama.1992.03490170092032
22. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity

in three search tools for qualitative systematic reviews. *BMC health services research*. 2014; **14**(1):579-.

23. EndNote. Product Information 2020 [Available from: https://endnote.com/product-details/
24. Covidence systematic review software. 2019 [Available from: www.covidence.org. www.covidence.org
25. Greene E, O'Connell R, Murphy MM et al. A systematic review of tools available to measure clinician performance of the complete physical examination and screening of the neonate: a protocol. *PROSPERO International Prospective Register of Systematic Reviews*. 2020; CRD42020151657. Available from: https://www.crd.york.ac.uk/prospero/
26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*. 2010; **63**(7):737-45.
27. Mokkink L.B. PCAC, Patrick D.L., Alonso J., Bouter L.M., de Vet H.C.W., and Terwee C.B. *COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs) user manual*. 2018; Version 1.0. https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf
28. Mokkink LB, De Vet, H.C.W., Prinsen, C.A.C, Patrick, D.L., Alonso, J., Bouter, L.M., Chiarotto, A., Westerman, M. J. *COSMIN Risk of Bias checklist* 2018. https://www.cosmin.nl/wp-content/uploads/COSMIN-RoB-checklist-V2-0-v17_rev3.pdf
29. Terwee CB, Prinsen C.A.C., Chiarotto A., De Vet H.C.W., Bouter L.M., Alonso J., Westerman M.J., Patrick D.L., Mokkink L.B. *COSMIN methodology for*

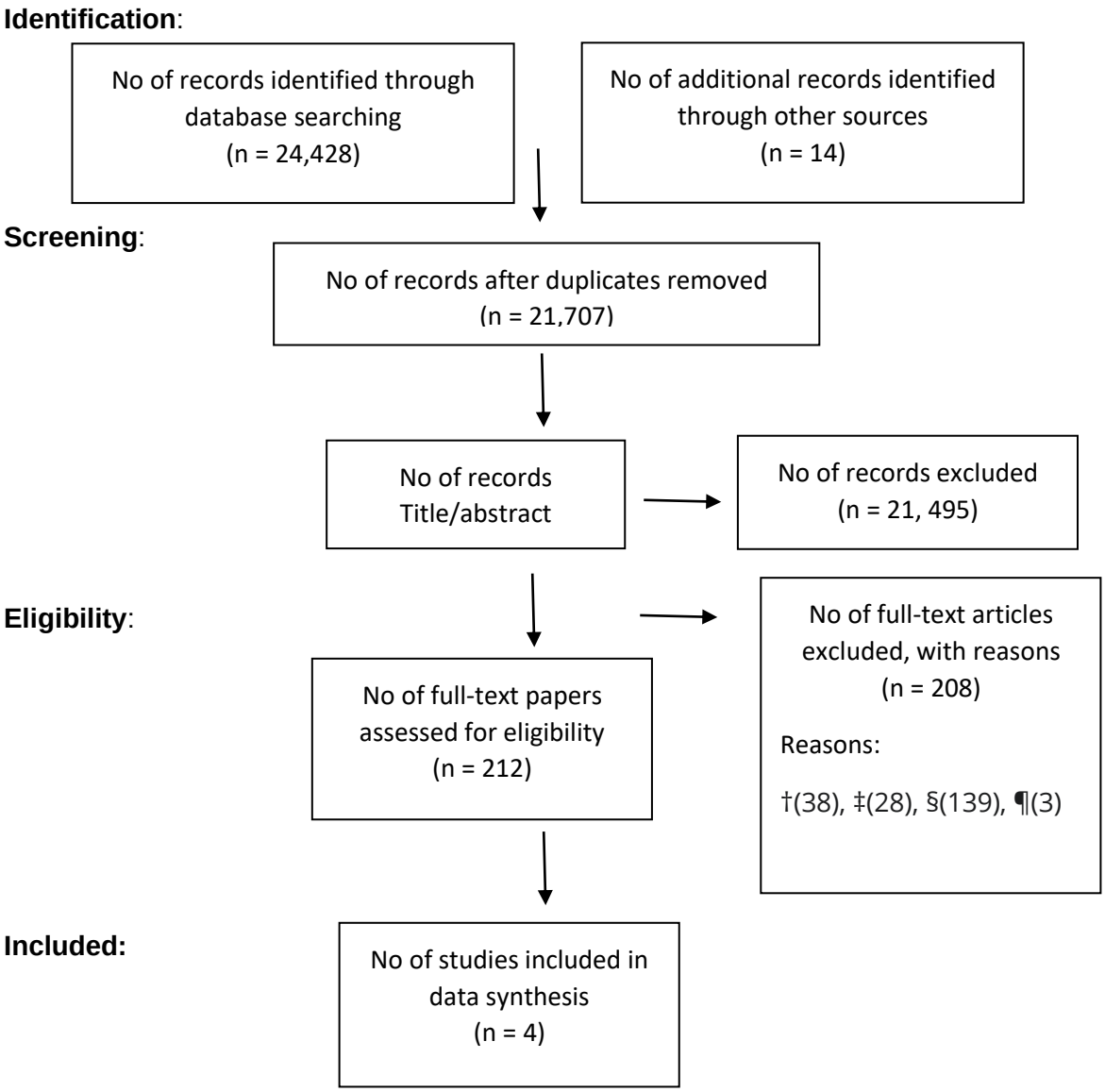
assessing the content validity of PROMs user manual version 1.0 2018. https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf

30. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P, for the PG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009; **339**(7716):332-6. doi:10.1136/bmj.b2535
31. Anderson M. Effect of integrated high-fidelity simulation in knowledge, perceived self-efficacy and satisfaction of nurse practitioner students in newborn assessment. A dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in health studies in the graduate school of the Texas Woman's University. 2007. https://twu-ir.tdl.org/handle/11274/9302
32. Stewart A, Inglis G, Jardine L, Koorts P, Davies MW. A randomised controlled trial of blended learning to improve the newborn examination skills of medical students. *Archives of disease in childhood Fetal and neonatal edition*. 2013; **98**(2):F141-4.
33. Benjamins SJ, Damen MLW, Van Stel HF. Feasibility and impact of doctor-nurse task delegation in preventive child health care in the netherlands, a controlled before-after study. *PloS one*. 2015; **10**(10):e0139187-e.
34. Hoque N, Wang S, Watson M. G543(P) Improving patient safety and quality of care for regional neonatal services. *Archives of disease in childhood*. 2015; **100**(Suppl 3):A241-A.
35. Hundley VA, Tucker JS, van Teijlingen E, Kiger A, Ireland JC, Harris F, et al. Midwives' competence: is it affected by working in a rural location? *Rural Remote Health*. 2007; **7**(3):764.
36. Green P. P063: Intertwining scenarios with hi-fidelity simulation to teach medical students peripartum clinical skills. *Journal of Paediatrics and Child Health*. 2011 (s1):77-8

37. Guest EM, Keatinge DR, Reed J, Johnson KR, Higgins HM, Greig J. Implementing and evaluating a professional practice framework in child and family health nursing: A pilot project. *Nurse education in practice*. 2013; **13**(5):393-9. doi:10.1016/j.nepr.2012.11.004
38. Petershack JA GA, Vasquez MM, Littlefield JH, Manwell-Jackson M. Improving trainee performance of the newborn physical examination. *Southern Regional Meeting Abstracts of the American Federation for Medical Research*. 2009. https://afmr.org/abstracts/2009/SR2009_abstracts/293.cgi
39. Sahana KS, Qadiri GJ, Saldanha PRM. End of Paediatric Posting Assessment of Medical Interns at a Teaching hospital from Coastal Karnataka. *Journal of Nepal Paediatric Society*. 2019; **38**(2):90-3.

Characteristics of the CESoN **C**omplete **E**xamination and **S**creening **o**f the **N**eonate- as per National Institute of Health and Care Excellence-NICE[1] and Newborn and Infant Physical Examination-NIPE[2] and Health Service Executive-HSE[3]
✓ Assess perinatal history and risk factors ✓ A detailed head-to-toe physical examination including: o Overall skin colour & texture, perfusion, breathing, behaviour, posture o Structure and symmetry of the head, face, ears o Proportions and symmetry of the neck, clavicles, limbs, hands, feet, and digits o Auscultate heart and lung sounds o Assess chest shape, respiratory effort o Assess and palpate abdomen to out rule organomegaly o Assess genitalia and anus completeness and patency o Assess and palpate spinal structures ✓ Elicit primitive reflexes ✓ Screening for congenital defects of the following systems ✓ Eyes -confirm red reflex is present, out rule opacities ✓ Heart- assess rate, rhythm, sounds, murmurs and confirm bilateral femoral pulses on palpation ✓ Hips- perform Barlow and Ortolani manoeuvers ✓ Genitalia & testes in males- palpate to out rule undescended testes ✓ Performance of anthropometric measurements ✓ measure the body length, occiptio-frontal head circumference, birth weight then plot on a growth chart ✓ Discharge advice to parents/caregivers ✓ How to recognize if baby is unwell ✓ Health promotion including optimising Infant Mental Health, Safe Sleep advice, Immunisations, Vitamin D supplementation.

Figure 1 Characteristics of the CESoN Complete Examination and Screening of the Neonate



†did not include the most of the fundamental elements of the complete examination and screening of the neonate[1,2]; **‡**Qualitative/Discussion about CESoN only, **§**physical examination and screening of an infant more than 8 weeks after birth ultrasound examination and screening of the foetus; newborn hearing screening; metabolic screening e.g. the newborn blood spot screening programme; tandem mass spectrometry, ¶CESoN tool not available

[30]Moher D, Liberati A, Tetzlaff J, Altman DG, Group P, for the PG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009; **339**(7716):332-6. doi:10.1136/bmj.b2535

Figure 2

PRISMA (Preferred reporting items for systematic review and meta-analyses flow diagram)[30]

Table 1. Characteristics of included instruments

PerBOM Name	Country of origin and (language of development)	Number of items	Response options	Range of Scores	Mode of Administration	Target Population	Average Completion Time	Content of Instrument: (Checklist as per NICE[1] and NIPE[2])
VAPBES- Video analysis proforma Baby Examination Study [10] ***Peer-Assessment Tool**	U.K. English	62 items:	54 items: Yes/No/Unable to Judge 3 items: 4-point Likert 1=Baby quiet 2=Baby whimpering 3=crying, slight resistance 4=crying 3 items: 4-point Likert 1=not at all 2=rarely 3=frequently 4=most or all of the time 1 item: 7-point Likert range: -3= Very Poor 0= Neither +3= Very good 1 item: open-ended question	**Total score across 7 items indicating lowest quality examination skills= 18. Total score across 7 items indicating highest quality examination skills= 12.	Maternity hospital setting. Senior midwives and consultant paediatricians completed the instrument to rate quality of junior doctors and midwives' performance of newborn examination via video recordings	Paediatricians, Midwives.	Not reported	✓ Assess history and risk factors ✓ Performance of the head-to-toe physical examination ✓ Elicit primitive reflexes ✓ Screening of ✓ eyes ✓ heart ✓ hips ✓ genitalia & testes in males for congenital defect ✓ Performance of anthropometric measurements ✓ Advice re: signs of ill health, health promotion.
NEC- Newborn Examination Checklist[32] ***Peer-Assessment Tool**	Australia English	40-item	40 items Yes/No responses (dichotomous) 8 items deemed 'essential'	Lowest score: =0 Maximum score = 40 Essential score= 8	Hospital: Week 4 of an 8-week rotation.	Medical Students	Not reported	✓ Assess history and risk factors ✓ Performance of the head-to-toe physical examination ✓ Elicit primitive reflexes ✓ Screening of ✓ eyes ✓ heart ✓ hips ✓ genitalia & testes in males for congenital defect ✓ Performance of anthropometric measurements ✓ Advice re: signs of ill health, health promotion.

PSETNA- Perceived Self-Efficacy Tool for Newborn Assessment[31] ***Self-Assessment Tool**	U.S.A English	18-items	4-point Likert scale 1= not at all true 2= hardly true 3= moderately true 4= exactly true	Lowest score= 18 Highest score = 72	University setting: Control Group: 1 questionnaire administered on 3 occasions: 1. pre-educational package 2. post education 3. post education Experimental Group: 1 questionnaire administered on 3 occasions: 1. pre-educational package 2. post education 3. post education	Nurse Practitioner Students	10 minutes approx.	✓ Assess history and risk factors ✓ Performance of the head-to-toe physical examination ✓ Elicit primitive reflexes ✓ Screening of ☒ eyes ✓ heart ✓ hips ✓ genitalia & testes in males for congenital defect ✓ Performance of anthropometric measurements ☒ Advice re: signs of ill health, health promotion.
PCHNCQ- Preventative Child Healthcare Nurses Competence Questionnaire Tool[33] ***Self-Assessment Tool**	Netherlands Dutch (translated by authors to English for publication)	53-items	3 items- demographic questions 50 items- ordinal response options: 1= I have not learned this 2= I have learned this but do not use in my work 3= I have learned this and use partly 4= I have learned this and use fully.	Competence Score: 1= Lowest 4= Highest. Lowest possible score for the 50 baby examination items= 50. Highest possible score for the 50 baby examination items= 200.	Community setting. x2 Competence questionnaires - nurses self-completed. 1st- To control and experimental groups pre and post education and training. 2nd one year after training.	Nurses	Not reported	✓ Assess history and risk factors ✓ Performance of the head-to-toe physical examination ✓ Elicit primitive reflexes ✓ Screening of ✓ eyes ✓ heart ✓ hips ✓ genitalia & testes in males for congenital defect ✓ Performance of anthropometric measurements ✓ Advice re: signs of ill health, health promotion.

**The scoring system assigned to the 54 '*Yes/No/Unable to Judge'* items were not reported, therefore the range of total scores indicating either high quality or poor quality Baby Examinations as per the VAPBES is unknown.

Table 2. Characteristics of included studies

Study Name	**Author(s) (year) (Country)**	**Study aim(s) Add Study Design**	**Sample Total *N***	**Sample characteristics**
Routine examination of the newborn: the EMREN~ study. Evaluation of an extension of the midwife role including a randomized controlled trial of appropriately trained midwives and paediatric senior house officers – Introduction. [10]	Townsend et al. 2004 U. K	Null hypothesis: Midwives & paediatric SHO£ equal performance of CESoN.	N=19	2 Consultant Paediatricians **&** 2 Senior Midwives assessors (4) 11 midwives + 8 SHO£ = 19 Each practitioner videoed twice = 39 neonatal examinations rated by assessors
Effect of integrated high-fidelity simulation in knowledge, perceived self-efficacy and satisfaction of nurse practitioner students in newborn assessment. [31]	Anderson 2007 U.S.A.	Null hypothesis: No difference n self-efficacy scores of student NPs^ in Newborn Assessment (control & intervention groups).	N= 68	Control 22 Experimental 46 Student Nurse Practitioners (22+46=68)
Feasibility and impact of doctor-nurse task delegation in preventive child health care in the Netherlands, a controlled before-after study. [33]	Benjamins et al. 2015 Netherlands	No hypothesis. Verify nurse competence in 'medical' newborn examination.	N= 18	Control & Experimental groups of nurses (13+5=18) Education levels similar, Previous experience levels lower in some nurses.
A randomised controlled trial of blended learning to improve the newborn examination skills of medical students.[32]	Stewart et al. 2013 Australia	Hypothesis: Can blended learning improve newborn examination skills of medical students.	N= 3	3 neonatologists. 71 (37+34) students assessed (control and experimental groups)

~ Evaluation of the **M**idwife's **R**ole extension in the **E**xamination of the **N**ewborn; £ Senior House Office (junior doctor); ^ student nurse practitioners.

Table 3. Details of COSMIN Taxonomy and Risk of Bias Checklist

The COSMIN Taxonomy	COSMIN Risk of Bias-RoB Checklist	Explanations of RoB Checklist Terms in Boxes 1 to 10[26-29]
Validity	Box 1. PROM development	Concerns a) standards for evaluating the quality of PerBOM design and concept elicitation, and b) standards for evaluating quality of pilot testing of PerBOM.
Content validity	Box. 2 Content Validity	Concerns standards for evaluating the quality of studies on the content validity of a PerBOM by rating standards for relevance, comprehensiveness, and comprehensibility.
Construct validity	Box. 3 Structural Validity	Standards to evaluate if the PerBOM scores can adequately reflect the dimensionality of the construct to be measured i.e., unidimensionality or structural validity.
Criterion validity	Box 4. Internal Consistency	Internal consistency refers to the degree of interrelatedness among the PerBOM items- applicable to unidimensional scales.
Reliability	Box 5. Cross cultural validity/ Measurement invariance	Cross-cultural validity evaluates the degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument.
Internal Consistency	Box 6. Reliability	Reliability evaluates the proportion of the total variance in the measurements which is due to 'true' differences between participants being measured by the PerBOM.
Test re-test, inter-rater, intra-rater	Box 7. Measurement Error	Measurement error evaluates the systematic and random error of an individual participant's score that is not attributed to true changes in the construct to be measured in the PerBOM.
Measurement error	Box 8. Criterion Validity	Criterion validity evaluates the degree to which the scores of a PerBOM are an adequate reflection of a 'gold standard' for the construct being measured.
Responsiveness	Box 9. Hypothesis testing for construct validity	Hypotheses testing for construct validity evaluates the degree to which the scores of a PROM are consistent with hypotheses of the construct being measured.
Interpretability	Box 10 Responsiveness	Responsiveness evaluates the ability of a PerBOM to detect change over time in the construct to be measured.

Table 4. COSMIN Ratings of Quality of PerBOM Development[27-29]

PerBOM	PerBOM design							Cognitive interview (CI) study[2]				TOTAL PerBOM DEVELOPMENT
	General design requirements					Concept elicitation [1]	Total PerBOM design	General design requirements	Comprehen-sibility	Comprehen-siveness	Total CI study	
	Clear construct	Clear origin of construct	Clear target population for which the PROM was developed	Clear context of use	PerBOM developed in sample representing the target population			CI study performed in sample representing the target population				
Townsend VAPBES[10]	V	V	V	V	V	D	D	V	D	D	D	D
Stewart NEC[32]	V	V	V	V	I	D	D	I	D	D	I	I
Anderson PSETNA[31]	V	V	V	V	V	D	D	V	I	D	I	I
Benjamins PCHNCQ[33]	V	V	V	V	I	I	I	I	I	D	I	I

V = very good; A = adequate; D = doubtful; I = inadequate; NA = not applicable

[1] When the PROM was not developed in a sample representing the target population, the concept elicitation was not further rated

[2] Empty cells indicate that a CI study (or part of it) was not performed

Data informing Table 4 is available as supplementary files 1 and 2

Table 5. Summary of Findings: Overall COSMIN Ratings of PerBOMs and Quality of Evidence Ratings[27-29]

Table 5. Overall Ratings of PerBOMs and Quality of the evidence for PerBOMS	**VAPBES**[10]		**PSETNA**[31]		**PCHNCQ**[33]		**NEC**[32]	
	OVERALL RATING	**QUALITY OF EVIDENCE**	**OVERALL RATING**	**QUALITY OF EVIDENCE**	**OVERALL RATING**	**QUALITY OF EVIDENCE**	**OVERALL RATING**	**QUALITY OF EVIDENCE**
	+ / - / ?	**High, moderate, low, very low**	**+ / - / ?**	**High, moderate, low, very low**	**+ / - / ?**	**High, moderate, low, very low**	**+ / - / ?**	**High, moderate, low, very low**
Content validity	+	Moderate	?	Low	?	Moderate	+	Moderate
Relevance	+	Moderate	+	Moderate	+	Moderate	+	Moderate
Comprehensiveness	+	Moderate	?	Moderate	-	Low	+	Moderate
Comprehensibility	+	Moderate	?	Moderate	+	Moderate	+	Moderate
Structural validity	?	Moderate	-	Moderate	?	Moderate	?	Moderate
Internal consistency	?	Moderate	+	Moderate	-	Low	+	Moderate
Cross-cultural validity	NR		NR		NR		NR	
Measurement invariance	+	Moderate	+	Moderate	?	Moderate	+	Moderate
Reliability	?	Moderate	-	Moderate	-	Low	-	Low
Measurement error	+	Moderate	?	Moderate	-	Low	-	Low
Criterion validity	+	Moderate	+	Moderate	+	Moderate	-	Low
Construct validity	+	Moderate	+	Moderate	?	Moderate	+	Moderate
Responsiveness	+	Moderate	+	Moderate	?	Moderate	+	Moderate

Overall Ratings System: 'sufficient' (+), 'insufficient' (-), indeterminant (?), Not Reported (NR). Overall Quality of Evidence Rating System (GRADE): High, Moderate, Low, Very Low. Data informing Table 5 is available as supplementary file 1 and 2