

The Quality of Routine Examinations of the Newborn performed by Midwives and SHOs; an Evaluation using Video-recordings

Authors:

Linda Bloomfield ¹
Cathy Rogers ²
Joy Townsend ¹
Dieter Wolke ³
Emmanuel Quist-Therson ⁴

- 1 University of Hertfordshire, Centre for Research in Primary and Community Care
- 2 University of Hertfordshire, Department of Midwifery and Child
- 3 University of Hertfordshire, Department of Psychology, DWRU and University of Bristol, Division of Child Health, ALSPAC
- 4 West Hertfordshire Hospitals NHS Trust

Running Head:

Quality of Routine Newborn Examinations

Keywords:

newborn examination, midwife, SHO,
video analysis, inter-rater reliability

Correspondence:

Linda Bloomfield
Centre for Research in Primary and Community Care
University of Hertfordshire
College Lane
Hatfield, Herts
AL10 9AB

Telephone: 01707 285992
e-mail: l.j.bloomfield@herts.ac.uk

Acknowledgements:

The work was funded as part of the UK NHS Health Technology Assessment Programme (grant 94/40/05). Opinions and conclusions are those of the authors and not the funders. Thanks are due to, Julie Hayes, Anne Monument, David Messer, Shreya Dave, Powatti Ramchand and Maggie Tomlin, for comments and participation in the study and to John Bain for editing the videotapes. We are indebted to the consultant paediatricians and senior midwives who rated the videos, and the midwives, junior paediatricians, mothers and babies who were video recorded.

Abstract

Objective: To evaluate the quality of the routine examination of the newborn as carried out by senior house officers (SHOs) and midwives.

Design: Randomised controlled trial. Eligible babies were randomised to a midwife or SHO who were then video recorded undertaking the routine newborn examination.

Setting: District general hospital in Southeast England.

Participants: 11 Midwives and 8 SHOs.

Main outcome measures: Quality of 62 observed technical and communication components of the newborn examination, as agreed between independent consultant paediatrician and senior midwife raters.

Results: Major differences were found in the rated quality of examinations between midwives and SHOs for; the examinations of the heart and lungs; for the overall quality of the examination; and in areas of communication skills. Quality of examination of the hips was assessed as poor for both professional groups. Where there were significant differences between examiners, the quality of the midwives' examinations was higher. Inter-rater agreement between the consultant paediatricians and senior midwives ranged from excellent to poor for different items of the examination with a mean kappa value of 0.42 across all items indicating moderate agreement.

Conclusions: The quality of midwife examinations exceeded that of SHOs. All midwives who examine receive formalised training in the examination of the newborn; SHOs may benefit from similar specific training. Training for both professional groups may need to be reviewed, particularly in relation to agreement on "gold standards" for the screening of developmental dysplasia of the hip.

KEY MESSAGES

- ‘Gold standards’ for components of the routine newborn examination need to be agreed within and between professional groups.
- The examination of the hips was assessed as poor and training should be reviewed.
- The quality of midwife examinations exceeded that of SHOs for hearts, lungs and overall examination.
- Midwives examining have formalised training (ENB N96) in the newborn examination.
- Greater emphasis in training should be placed on communication skills and health education.
- The use of video recordings may be useful in training examiners of the newborn.

Introduction

A detailed examination of the newborn is recommended as a core component of Child Health Surveillance^{1 2} and is usually carried out by a senior house officer (SHO). Recommended components include history taking, physical examination, health education and parental reassurance.^{1 2} Recent changes in the organisation of maternity care and training, and the reduction of junior doctors' hours have resulted in some midwives, who have received specific training, taking responsibility for the examinations. This development has been facilitated by the introduction of the (ENB N96) course in neuro-behavioural physiological assessment of the newborn, as defined by the English Nursing Board for nurses, midwives or health visitors and developed in 1995.³⁻⁵ There is consensus in the literature that, provided they receive adequate training, midwives are well placed to undertake the examination of the newborn,^{1 2 5-9} nevertheless there is currently no evidence to support this view, although mothers have been shown to be more satisfied with midwife examinations than those carried out by SHOs.^{10 11} The introduction of newborn care by advanced neonatal nurse practitioners (ANNPs) at Ashington has attracted considerable interest and is of a high standard,¹² and Lee and colleagues have also shown that ANNPs were significantly more effective in detecting abnormalities in hips, heart and eyes than were SHOs.¹³ Given appropriate guidelines, SHOs have been shown to have the skills to assess the significance of and decide on appropriate management for neonatal murmurs.¹⁴ However several studies have reported that the examination does not perform well as a screening tool.¹⁵⁻²⁰

An appropriate test of quality would be how far practitioners adhere to agreed best practice in administering the examination and this could be assessed with the aid of video recordings. In psychological research, video-recordings are often used for assessments

with subsequent structured evaluation of objectivity and checking of quality of administration by examiners.²¹⁻²³ Video analysis is also increasingly used in training for self-evaluation and for discussion purposes.^{22 24 25} However, where they have been used to make comparisons between different personnel, for example in carrying out screening for hip instability, variations between raters as well as differences in levels of performance between the groups of examiners have been reported.²³ This suggests that 'gold standards' for correct assessment are needed.

The study reported here is part of a programme of evaluation of the routine examination of the newborn with regard to UK practice, including a randomised controlled trial of 826 examinations to assess safety and appropriateness of referral, parental satisfaction with the examination and quality of examination.²⁶ This study reports the quality assessment part of the randomised controlled trial which used video recording as an objective audio-visual record of examinations. A proforma was developed and validated for use in the evaluation in relation to the quality of the physical components of the examination, health education and parental reassurance. The aim was to identify any differences in quality of examination between midwives and SHOs.

Methods

Setting

The study was carried out in Southeast England in a district general hospital, which has approximately 3000 deliveries per year, and where the routine examination of the newborn was conducted by either an SHO or a community midwife. All community

midwives conducting the examination had completed the English National Board N96 post-registration course in physiological assessment of the newborn, one requisite for midwives to be allowed to carry out the examination. Training of the SHOs had been more informal and was not standardised. The hospital Trust had exclusion criteria for babies on the postnatal ward who could not be examined by a midwife.¹⁰ The protocol for this study was approved by the NHS Trust's Ethical Committee.

Subjects

Eleven midwives and eight SHOs participating in a randomised controlled trial of the examination of the newborn,¹⁰ gave signed consent to be videoed while performing examinations. Babies were included in the study if they fulfilled the Trust's inclusion criteria for midwife examination, and if mothers had agreed to participate in the randomised controlled trial.¹⁰ Babies were therefore comparable between midwife and SHO examinations. Signed consent specifically to video the baby's examination was obtained also from the mother. Each examiner was videoed on two separate occasions. In total 39 newborn examinations were recorded, including one pair of twins; 22 were conducted by midwives and 17 by SHOs.

Procedure

All video-taped examinations were carried out in the hospital, on the post-natal ward at the mother's bedside. All sessions were videoed with a hand held video camera focused on the baby and the examiner's hands. Care was taken to ensure that the camera and operator did not interfere with the examination and that the identity of the examiner was kept anonymous. The video recordings were edited to remove any verbal or visual

reference to the examiners' identities and professions before the examinations were rated by the independent observers. Each videoed examination was randomised to one of two tapes. A consultant paediatrician and a senior midwifery lecturer on the research team piloted the proforma using a number of videoed newborn examinations. The scoring format was adjusted and instructions clarified for items where there was rating disagreement.

Two consultant paediatricians and two senior midwives, with extensive and current experience of the newborn examination, rated the recordings using the written proforma developed by the research team. The raters were from three different hospitals, not including the study hospital, and none knew the videoed staff. One consultant and one midwife independently assessed 20 examinations of which 13 were conducted by midwives and seven by SHOs, and the other consultant and midwife rated the remaining 19 examinations of which nine were conducted by midwives and ten by SHOs. Each examination was therefore independently rated by one consultant and one midwife. Each videoed examination was randomised to one of two tapes, therefore it was not predetermined whether or not both videos for each examiner would be reviewed by the same raters. Four midwives and four SHOs were reviewed on both occasions by the same raters and seven midwives and four SHOs by different raters on both occasions.

The written proforma included criteria for rating each physical component of the examination, each aspect of communication and the examiner's response and sensitivity to the mother (appendix 1). It included 61 items to be observed and coded. Fifty four of these items required a response from the raters of 'yes' (it was done), 'no' (it was not

done) or 'unable to judge' according to whether the rater observed the item to have been carried out or not. 'Unable to judge' was selected if the behaviour was not observable, for example due to background noise, or being obscured by the examiner's body. Six items, including 'how much did the baby cry or fuss during the examination?', required a rating on a four-point Likert scale, with responses ranging from 'not at all' to 'most or all of the time'. One item, 'How would you judge the overall quality of the physical examination in terms of technical competence?', required rating on a seven-point scale from 'very poor' to 'very good'. A further item (62) was constructed from the comments of the raters about whether the Barlow's test for neonatal hip instability had been carried out or not. Raters were encouraged to record comments where appropriate. Items relating to the examination of the hips were adapted from a published form designed to highlight the essential components of the hip examination.²³ Guidelines, including instructions, diagrams and rating scales, were given to the raters to facilitate use of the proforma. The four independent raters attended a briefing day prior to assessing the videos; each rated the same two videotapes so that the rating criteria could be standardised.

Analysis

Evaluation of examinations

The differences between appropriate examination by the midwives and SHOs were tested for each item using Fisher's exact test. Items rated on a four point Likert scale were dichotomised as categories, for example 'how much did the baby cry or fuss during the examination' was recoded as either 'not at all/rarely' or 'frequently/most or all of the time'.

Identification of items with acceptable inter-rater reliability

For each item, the level of agreement between raters was assessed using the kappa coefficient, which is the standard measure of level of agreement.²⁷ Items with kappa values greater or equal to 0.4 were considered to have moderate to good agreement; those with kappa values less than 0.4 were interpreted as having poor to fair rater agreement.²³ (The percentage level of agreement between raters was also assessed, but not used as a criterion of agreement as this does not discriminate between actual agreement and agreement that arises due to chance, nor does it account for bias.) Where one rater had rated an item as ‘unable to judge’ or had failed to enter a rating, that item for that examination was excluded from further analysis.

Results

Comparisons of the observed skills and competence

The comparisons of the observed skills of the midwife and SHO examiners are shown in three tables as follows: Table 1a, items for which there were significant differences between the examiners and good agreement between the raters (Fisher’s exact test, $p \leq 0.05$ and $\text{kappa} \geq 0.4$); Table 1b, items for which there were no significant differences found between the examiners and there was good agreement between the raters (Fisher’s exact test, $p > 0.05$ and $\text{kappa} \geq 0.4$); and Table 1c, items for which there were significant differences between the examiners although not good agreement between the raters (Fisher’s exact test, $p \leq 0.05$ and $\text{kappa} < 0.4$). Tables differentiate comparisons by the

consultant paediatrician and senior midwife raters. For the remaining items there were not significant differences between examiners, nor good agreement between the raters, and these are not shown separately but are included in the numbers in Table 2.

TABLES 1a, 1b, 1c

For every item where significant differences between examiners were identified, the item received higher ratings for the midwife examinations than for the SHOs (Tables 1a and 1c). Major differences between midwives and SHOs were identified for; quality of the examination of both the heart and the lungs; for the overall quality of the examination; and in areas of communication skills, including discussing healthcare issues and soothing the baby. There were no significant differences identified between examiners for; components of the Ortolani's test of the hips; palpation of pulses; screening for neurological problems; examination of the eyes and spine; history taking; or discussion of baby issues (Table 1b). On the overall quality of the physical examination, midwives were rated as good or very good by the senior midwives for 72.7% of the examinations and by the consultant paediatricians for 22.7% of the examinations. SHOs were rated good or very good by senior midwives for 11.8% of the examinations and by consultant paediatricians 0% of the examinations (Table 1c).

Child health care issues were more frequently rated as discussed by midwives than by SHOs (Table 1a). Cord care, feeding, sticky eye and nail cutting were discussed by the SHOs, while a wider range of issues were discussed by the midwives including feeding, sleeping position, cord care, bathing, stools, cot death, skin care and jaundice.

Inter-rater agreement

Agreement between the raters for each item ranged from poor to excellent (kappa between zero and 1.0), with a mean kappa value of 0.42 across all items, indicating moderate agreement overall. The percentage agreement between raters ranged from 44% to 100%, with a mean of 81.5 %. Raters said that no Barlow's procedure for neonatal hip instability was observed for a number of examinations, resulting in 23% missing data for the six items relating to the Barlow's test. The Barlow's test was therefore excluded from further analysis.

For twenty seven (48.2%) of the remaining 56 items included in the analysis, the kappa values were greater or equivalent to 0.4, i.e. there was moderate to good agreement between raters (Table 2). Eighteen of these related to technical components of the examination, including elements of the Ortolani's test for neonatal hip instability and screening for neurological problems; nine items related to the communication skills of the examiner including explaining what they were doing during the examination, responding to the mother and soothing the baby. The twenty nine items with kappa values less than 0.4, poor to fair agreement, included technical elements such as the overall screening for hip problems, heart disease and cataracts, as well as communication aspects such as inviting questions and explaining any problems identified (Table 2).

TABLE 2

Discussion

There were two major findings from this video assessment of the newborn examination. Firstly, that all statistically significant differences in quality of examinations were in

favour of the midwives; secondly, that despite a training day and briefing, there was moderate to good agreement between the raters for only half the items on the videotapes. Midwife examinations were rated as being of higher quality than SHO examinations for both technical administration and communication skills, by both consultants and senior midwives. For certain components of the examination neither midwives nor SHOs were rated highly. In particular the screening for hip problems, particularly using the Barlow's test, was often poor, family history of problems was rarely discussed and the baby was often not relaxed during the hip examination or during auscultation of the heart. In many instances the Barlow's test was rated as not performed and the items had to be excluded from the analysis. This disagreement was due to the raters considering that for this test, each hip should be tested individually, whereas the practice of the examiners was to perform it on both hips simultaneously.

We consider that the lack of well defined 'gold standards' for procedures partly explains why only poor to fair agreement was found on half the items, on what was a relatively straightforward rating format. It was not the role of the study to develop 'gold standards', but during the training day for the raters, it was found that 'gold standards' for certain examination components, such as the Barlow's test, differed between the two consultant paediatricians, and between the consultant paediatricians and the senior midwives. Other studies have also reported significant differences in opinion between observers, particularly on what constitutes a good hip screening examination,²³ and have considered this a major obstacle to judging and improving the quality of certain assessments.²³ This indicates that for certain components of the examination, particularly for the Barlow's examination of the hips, tighter and clearer 'gold standards' need to be agreed within and between professional groups. It was clear that both consultant paediatricians and senior

midwives rated the quality of midwife examinations more highly than they rated the SHO examinations, although senior midwives tended to give higher ratings than did the consultant paediatricians. Even where differences in the quality of examinations were not statistically significant, the trend was still in favour of midwives for most aspects of the examination.

Analysis from videotapes allows objective assessments of the same behaviours by different observers, but it has limitations. It is possible that despite the removal from the tapes of all visual and verbal reference to the examiner's identity, the raters may have partly 'guessed' the examiner's profession on the assumption that most midwives are female. However many of the SHOs were also female, and as the midwives were rated higher we conclude that no bias was evident. The observer of the videotaped examinations is not necessarily able to see what the examiner sees, for example when screening for cataracts nor to hear what the examiner hears, for example, when screening for heart disease. Therefore for some aspects of quality control, additional methods of assessment, such as audio playing of different heart recordings to assess the correct detection rates of heart murmurs are required. A larger study could facilitate this and raise awareness of 'gold standards'.

These video analysis ratings accord with the results of parents' satisfaction with 826 newborn examinations on the postnatal ward¹⁰ and three months later.¹¹ Mothers reported higher satisfaction when a midwife rather than an SHO carried out the examination and they reported that midwives discussed physical and behavioural health care issues more often than did SHOs, again in accordance with the video recorded observations. The

direct observation and the mothers' views provide consistent results. The findings also accord with a recent comparison of trainee paediatricians with ANNPs concerning the detection rate for abnormalities, which reported the latter to be more effective.¹³ Early results from the Ashington study suggest that care of the newborn by ANNPs is of a satisfactory standard.¹² The findings of higher quality of examinations by midwives and more effective examinations by ANNPs, may be due to the more intensive and formal training they receive compared to that received by SHOs.

This is the only study to make such a comparison of SHOs and midwives; as midwives care for the large majority of the newborn, it would be practicable for them to undertake routine examinations. Current practice of the examination clearly confounds examiner's profession and training; whereas training (the ENB N96) is required for all examining midwives, formal training is not required for paediatric SHOs who carry out the examination. The conclusion is not that SHOs could not be as competent as midwives were they given sufficient training. However the results do demonstrate that from the point of view of the clients – the parents and baby- the examinations are as competently carried out by midwives as by SHOs.

Conclusions

This study suggests, in accordance with Hall,^{1 2} that with adequate training and support the examination of the newborn may be carried out satisfactorily by midwives. The findings strongly suggest that SHOs carrying out the routine examination of the newborn would benefit from a formalised introduction and training for the newborn examination similar to that provided for midwives. Furthermore greater emphasis in their training

could be placed on communication skills and health education. There is scope in the current training to enhance the quality of newborn assessments particularly for screening for developmental dysplasia of the hip and family history taking. The use of video recordings for purposes of training and supervision,^{24 25} and to ensure objectivity of assessors, could become an integral part of training.

References

1. Hall D. *Health for all children*. 3rd ed. Oxford: Oxford University Press, 1996.
2. Hall D. The role of the routine neonatal examination. *British Medical Journal* 1999;318:619-620.
3. DOH. Doctors in training: A new deal. London: Department of Health, 1991.
4. DOH. Changing childbirth: Report of the Expert Maternity Group. London: Department of Health: HMSO, 1993.
5. RCM. Vision 2000. Executive Summary. London: Royal College of Midwives, 2000.
6. Michaelides S. A deeper knowledge. *Nursing times* 1995;91(35):59-61.
7. Mackeith N. Who should examine the "normal" neonate? *Nursing Times* 1995;91:34-35.
8. Lomax A. Expanding the midwife's role in examining the newborn. *British Journal of Midwifery* 2001;9(2):1000-1002.
9. Seymour J. Who checks out? *Midwifery* 1995;5(4):201-202.
10. Wolke D, Dave S, Hayes J, Townsend J, Tomlin M. Routine examination of the newborn and maternal satisfaction: a randomised controlled trial. *Archives of Disease in Childhood Fetal and Neonatal Edition* 2002;86:F155-160.
11. Wolke D, Dave S, Hayes J, Townsend J, Tomlin M. A randomised controlled trial of maternal satisfaction with the routine examination of the newborn baby at three months. *Midwifery* 2002;18(2):145-154.
12. Hall D, Wilkinson A. Neonatal services provided by Advanced Neonatal Nurse Practitioners at Ashington - what have we learned in the first five years? London: RCPCH, 2001:1-5.
13. Lee T, Skelton R, Skene C. Routine neonatal examination: Effectiveness of trainee paediatrician compared with advanced neonatal nurse practitioner. *Archives of Disease in Childhood: Fetal Neonatal edition* 2001;85:f100-f104.
14. Farrer K, Rennie J. Neonatal murmurs: are senior house officers good enough. *Archives of Disease in Childhood Neonatal Edition* 2003;88:F147-151.
15. Abu-Harb M, Wyllie J, Hey E, Richmond S, Wren C. Presentation of left obstructive heart malformations in infancy. *Archives of Disease in Childhood* 1994;71:179-183.
16. Ainsworth S, Wyllie J, Wren C. Prevalence and clinical significance of cardiac murmurs in neonates. *Archives of Disease in Childhood. Fetal neonatal Edition* 1999;80:F43-45.
17. Arlettaz R, Archer N, Wilkinson. Natural history of innocent heart murmurs in newborn babies: controlled echocardiographic study. *Archives of Disease in Childhood. Fetal Neonatal Edition* 1998;78:F166-170.
18. Wren C, Richmond S, Donaldson L. Presentation of congenital heart disease in infancy: Implications for routine examination. *Archives of Disease in Childhood. Fetal neonatal edition* 1999;80:F49-53.
19. Glazener C, Ramsay C, Campbell M, et al. Neonatal examination and screening trial (NEST): a randomised, controlled, switchback trial of alternative policies for low risk infants. *British Medical Journal* 1999;318:627-631.
20. Dezateux C, Godward S. A national survey of screening for congenital dislocation of the hip. *Archives of Disease in Childhood Fetal and Neonatal Edition* 1996;74:445-448.

21. Gutbrod T, St.John K, Rust L, Wolke D. The mother-infant structured play assessment (MISPA): Very preterm infants and their mothers at 3 months. *Journal of Reproductive and Infant Psychology* 1999;17(3):222.
22. Trainor B, Haugh P, Kernohan G, Mollan R. Hip screening: are health visitors adequately prepared? *Health visitor* 1994;67(9):299-301.
23. El-Shazly M, Trainor B, Kernohan W, Turner I, Haugh P, Johnston A, et al. Reliability of the Barlow and Ortolani tests for neonatal hip instability. *Journal of Medical Screening* 1994;1:165-168.
24. Stranc M, McDiarmid J, Stranc L. Video assessment using surgical technique. *British Journal of Plastic Surgery* 1991;44:65-68.
25. Edwards A. Assessment of practical skills using video-recording. *Journal of the Royal College of General Practitioners* 1988;38:517.
26. Townsend J, Wolke D, Hayes J, Dave S, Rogers C, Bloomfield L, et al. Routine examination of the newborn by midwives: a randomised controlled trial (EMREN trial) report to the NHS National co-ordinating centre for Health Technology Assessment, 2002.
27. Carletta J. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics* 1996;22(2):249-254.