Presenting video recordings of newborn resuscitations in debriefings for teamwork training

Izhak Nadler,¹ Penelope M Sanderson,² Coleen R Van Dyken,³ Peter G Davis,⁴ Helen G Liley⁵

Background: The Neonatal Resuscitation Program (NRP) and similar courses have been used to train clinicians. However, formal teamwork training was not included in these courses, and their effectiveness has been questioned. In adult resuscitation, debriefings using video recordings have improved outcomes, but recordings of neonatal resuscitation have been used primarily for research.

Aims: To test if debriefings that include video recordings of neonatal resuscitations can improve teamwork and clinical practice.

Methods: Over 9 months, clinicians voluntarily attended weekly debriefings in which recently performed resuscitations were presented, evaluated and then discussed. Discussions were focused on teamwork and were facilitated by an education nurse and a human factors expert with a training background. Subsequently, three experts, a neonatologist, a neonatal nurse educator and a midwifery educator from different organisations viewed and scored 19 recordings from the first 4.5 months and 19 recordings from the second 4.5 months. Experts were blinded to the recording period. The two sets of recordings were balanced for babies’ gestation and recordings were presented, evaluated and then discussed. Discussions were focused on teamwork and were facilitated by an education nurse and a human factors expert with a training background.

Results: All ‘teamwork’ items improved between the two periods and one item improved significantly (p<0.05). Scores for ‘Following guidelines’ showed little change except ‘Intubation’ which deteriorated (p<0.05). There was no significant interaction between periods and raters. Cronbach’s alpha indicated inter-rater reliability ranging from 0.54 to 0.86.

Conclusions: Voluntary debriefings had some positive effect on neonatal resuscitation teamwork. Future studies are warranted to determine the best methods for debriefing combined with other training methods.

INTRODUCTION

In team training, and especially in simulator-based training, debriefings are considered one of the most valuable aspects.¹ ² We sought such benefits in a broad-based, weekly debriefing process within a community of neonatal clinicians, facilitated by recordings of actual neonatal resuscitations rather than of simulated ones.

The American Academy of Pediatrics and American Heart Association Neonatal Resuscitation Program (NRP) was introduced in 1987 and has since been used in more than 120 countries.³ The NRP guidelines⁴ and 2007 NRP textbook⁵ stipulate clinical aspects of resuscitation, but although anticipation of the need for additional personnel is encouraged, teamwork aspects are not specified.⁶ This has led to questions about the effectiveness of the NRP curriculum.⁷ ⁸ Other courses, such as the Resuscitation Council UK Newborn Life Support program, have similar characteristics. Meanwhile, in the US, where NRP training is universal, the Joint Commission on Accreditation of Healthcare Organizations⁹ has recommended that in order to prevent perinatal mortality and morbidity, maternity hospitals should (1) ‘conduct team training…’ and (2) ‘…conduct debriefings to evaluate team performance and identify areas for improvement’.

Many studies have tested the effectiveness of neonatal resuscitation training for improving performance in the training environment and for trainees’ perceptions of their experience.⁶ ¹⁰–¹³ One large retrospective cohort study¹⁴ measured the impact of NRP training on the Apgar scores of high risk newborns and found improvements after the NRP was implemented. However, doubts about the reliability of Apgar scoring¹⁵–¹⁷ and potential confounds suggest that these results do not reflect a true impact of NRP on improving clinical practice and outcomes.
In simulator training, scenarios are usually recorded, presented and then discussed in facilitated debriefing sessions. Debriefings give trainees feedback about their performance and behaviour, and can also highlight effective or ineffective performance and behaviours exhibited by others.\textsuperscript{18–20} Recordings of actual rather than simulated performance are seldom shown but, when they are, results are mixed. Hoyt \textit{et al}\textsuperscript{21} reported that after team members reviewed recordings of actual resuscitation, team performance in adult resuscitation improved, but Carbine \textit{et al}\textsuperscript{8} found that viewing neonatal resuscitations in quality assurance sessions did not greatly improve clinical performance. No previously published study has tested whether presenting recordings of actual neonatal resuscitations in debriefings improves teamwork in subsequent resuscitations. Furthermore, several studies have found gaps between training and performance in neonatal resuscitation, with potentially serious consequences for patient safety.\textsuperscript{3,22,23}

In this study, we tested whether presenting recorded neonatal resuscitations in regular voluntarily attended debriefing sessions would change clinicians’ teamwork and clinical performance at subsequent resuscitations, in an attempt to bridge the gap between training and practice.

**METHODS**

**Recording resuscitations and selecting cases for presentation**

A motion-activated recording system was installed in the resuscitation room adjacent to the obstetric operating theatres at Mater Mothers’ Hospital (MMH) in Brisbane, Australia. Recordings were of infants born by caesarean section who required resuscitation and survived to be admitted to a neonatal unit or post-natal ward. Later, resuscitation episodes were downloaded from the recording system if manoeuvres included more than providing suction and free flow of oxygen. Parents’ informed consent was sought for use of their baby’s recording in the study. The study was approved by the Human Research Ethics Committees of Mater Health Services and The University of Queensland.

From 122 cases that were recorded between November 2008 and November 2009, and for which consent was granted, 45 recordings were selected for presentation in debriefing sessions. The recordings were selected as showing good or bad examples of teamwork and clinical performance and of potential interest to neonatal clinicians. Each presentation included three channels (figure 1): ‘cot’ view, ‘ceiling’ view and ‘data’ view. Each recording started when the baby was placed in the open care resuscitation station (hereafter the ‘cot’) and ran until the baby was about to be transferred from the resuscitation room. Recording lengths varied between 3:05 min and 27:25 min (mean 10:42 min). The resuscitation teams included doctors, nurses and midwives, and resuscitation team size varied from two to five clinicians. Recordings were edited to reduce the likelihood of identifying staff and babies, and they were not presented in chronological order.

Weekly debriefing sessions started in mid-February 2009 and ran until mid-November 2009. MMH clinicians were invited to participate on a voluntary basis between their clinical duties and attendance was not restricted to those who had performed the resuscitations. To increase opportunities for attendance, the same recordings were presented in three sessions on each day. Each session started with an explanation from an education nurse about the purpose of the debriefings, the visual format of the recording and the evaluation process (presented below). Attendees were also instructed that the identities, if evident, of those who performed the resuscitations should not be discussed. During the debriefings period, the 45 recordings were presented over 37 weeks (about 110 sessions), with an average of six clinicians attending each individual session.

**Figure 1** Format of presentation of the recordings, showing three views. Left video view shows the cot with staff (faces were unmasked for presentation to the panel of experts), top right shows a general view and lower right view shows data downloaded from the cot monitor.
After viewing each recording, clinicians completed an anonymous evaluation form (online appendix 1). The first section asked about the clinician’s profession and workplace area. The second section asked for their impression of the appropriateness of team size in the resuscitation and about the number of the team members who acted as leaders during the resuscitation. The third section asked 18 questions divided into three areas: (1) ‘teamwork’ (seven items), (2) ‘following resuscitation guidelines’ (seven items) and (3) ‘temporal control of the resuscitation procedure’ (four items). Clinicians used a fiv-point Likert scale to respond to each question. Attendees were told that there were no ‘right’ or ‘wrong’ answers and that they should indicate their own impression about each scored item. Altogether, 711 forms were completed by doctors, nurses and midwives attending debriefing sessions.

A discussion facilitated by the education nurse followed the evaluation. Teamwork aspects were emphasised and explained by the education nurse and by a research team member who was a human factors expert with a professional background in training in aviation. The main focus of the discussions was making the clinicians aware of activities and behaviours that can affect (positively or negatively) teamwork performance. The theme of the discussions was ‘How can we (the viewers) perform better’ and not ‘Why did they (the resuscitation team) perform poorly’. Clinicians were encouraged to air opinions mostly about teamwork, but also about clinical aspects regarding the resuscitation presented.

To encourage participation, encourage trust in the activity and support morale, we did not collect any identifying data about clinicians who attended the debriefing sessions other than profession and workplace (eg, neonatal unit, theatre). Furthermore, we did not conduct any analysis based on the identity of clinicians who had performed the resuscitations.

**Evaluation of recordings with a panel of experts**

The evaluation was based on changes in overall performance of teams during resuscitation and not on changes in the skills and behaviours of individuals. To assess whether the debriefing sessions had an overall effect on teamwork and clinical performance, we presented a subset of the downloaded recordings to three experts, each from a different organisation and discipline, in a two-day evaluation panel. One expert was a neonatal nursing educator with 10 years experience teaching the NRP curriculum, the second was an experienced neonatologist who had conducted research on neonatal resuscitation, and the third was a midwife who was also a simulator instructor. We preferred a multidisciplinary panel to a uniform one, estimating that a variety of opinions about team performance provide a more robust test of the debriefing initiative, while acknowledging that diverse opinions might reduce the statistical power of the study.

The three experts were familiar with the viewing and scoring procedures because they had provided advice about the scoring questionnaire and had participated a year earlier in a preliminary review of a different set of recordings, using the same presentation format and questionnaire. We presented two ‘calibration’ recordings at the start of the panel to re-familiarise each expert with the scoring, and to let them calibrate their scores. An ‘excellent’ and a ‘poor’ recording were selected based on the evaluations of clinicians during debriefings. After viewing and scoring these two recordings, the experts and the research team reviewed and discussed the different items in the evaluation form. The experts were asked to base their scores on overall performance of the team and not on performance exhibited by a single individual or during one short instance.

Unlike the recordings presented in the debriefing sessions, the recordings presented in the evaluation panel were not masked, letting the experts use information such as facial expressions, head movements and eye directions to assist scoring. Each recording started when the baby was placed in the resuscitation cot and ran until no additional interventions were applied. This made the presentations of the recordings in the evaluation panel somewhat shorter than the presentations in the debriefings (where the termination was ‘baby is about to be transferred from the resuscitation room’). Durations of selected recordings varied between 2:55 min and 11:20 min (mean 5:38 min).

Thirty-eight recordings were presented to the panel; 19 from the first 4.5 month period of the debriefing period and 19 from the second 4.5 months. The experts were blinded to the period of each recording. To reduce confounding, the researchers selected recordings so as to achieve the following objectives: (1) approximately balance the sets for gestation and birth weight, (2) achieve the same number of recordings in the two sets (table 1) and (3) achieve a total presentation time compatible with the two-day schedule. Within these constraints, the recordings for the two sets were selected randomly from the set of recordings downloaded, regardless of content or the scores given by the clinicians during the debriefing. We excluded recordings that for technical reasons did not include monitoring data (such as heart rate and oxyhaemoglobin saturation).

All recordings were presented in random (non-chronological) order, scored individually and then discussed by the panel. Recordings were presented on two consecutive days, in five sessions of 1:45 h with breaks between sessions. In a concluding session, we asked the experts their opinions about the debriefing method we had used and ways to increase its effectiveness.
**Analysis**

For each item in the evaluation form, a two way ANOVA was used to test the experts’ scores for significant differences between experts (n=3), between periods (first vs second 4.5 months) and the interaction of experts and periods (no Bonferroni correction). A Wald–Wolfowitz Runs Test (adjusted for continuity due to small sample size) was used to determine whether scores for items relating to teamwork showed greater improvement over periods than did scores for items unrelated to teamwork. Internal consistency of the experts’ scores was tested with Cronbach’s alpha (\( \alpha \)) statistic for each scored item.

We did not analyse changes over time in the scores given by clinicians in the debriefing sessions for the present paper. First, the recordings presented in the debriefings were selected to demonstrate specific issues and were potentially less representative of all downloaded recordings than those shown to the expert panel. Second, because of the voluntary attendance and anonymity of scoring, we could not know whether characteristics of the clinicians (such as level of experience) attending debriefings changed over time. Third, most clinicians in the debriefings experienced no prior discussion about how to grade performance and teamwork.

**RESULTS**

All items relating to teamwork showed some level of improvement (table 2). One item improved significantly (p<0.05) (‘Was information sought?’) and two other items showed a similar trend (p<0.10) (‘Were human resources used effectively?’ and ‘Were next steps planned and discussed/anticipated?’). Results for items about clinical guidelines were mixed and one item even deteriorated significantly (p<0.05) (‘Intubation’). Results for items about temporal control of the resuscitation procedure were also mixed, but showed a slight tendency to deteriorate rather than improve. The Wald–Wolfowitz Runs Test was significant (Z_{adj}=2.3, p<0.05), indicating that scores for items about teamwork showed greater improvement over periods than did scores for items unrelated to teamwork. Events relating to 2 of the 18 scored items (‘chest compressions’ and ‘administering IV fluids/medications’) were rare, so these items were excluded from the analysis.

Scores of the three experts were significantly different for ten items, but there were no significant interactions between experts and periods. Cronbach’s \( \alpha \) statistic indicated that the internal consistency of the expert’s scores ranged between 0.54 and 0.86.

**DISCUSSION**

The experts’ scores revealed some improvement in clinicians’ teamwork over the duration of the study, suggesting beneficial effects of presenting resuscitations in debriefing sessions focused on improving teamwork. However, there were no changes in items relating to the teams’ temporal control of procedure, and like Carbine et al\(^8\) we did not find measurable improvements in the performance of clinical procedures.

The fact that significant improvements were seen only in teamwork items, and that the teamwork items improved significantly more than the non-teamwork items, was not surprising, given that the debriefing sessions focused on teamwork. Nonetheless, it is a remarkable finding in a study that exercised no experimental control over who attended the debriefing sessions, over who performed the resuscitations in the recordings or the relationship between the two. The study was designed to protect clinicians’ privacy and morale, and to encourage trust in the recording process and participation in the debriefings, and it may not have allowed the full contribution of the debriefings on teamwork practice to emerge.\(^{24,25}\)

**Table 1** Characteristics of resuscitated babies in the downloaded recordings, in recordings presented in debriefings, and in recordings shown to the evaluation panel. Recordings from the first and second 4.5 months of the debriefing period are compared.

<table>
<thead>
<tr>
<th>Date recorded (all are 2009)</th>
<th>Number</th>
<th>Gestation in weeks and days Mean (SD)</th>
<th>Birth weight (g) Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recordings downloaded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid February—June</td>
<td>54</td>
<td>31:0 (3:6)</td>
<td>1598 (711)</td>
</tr>
<tr>
<td>July—mid November</td>
<td>62</td>
<td>32:0 (4:0)</td>
<td>1700 (802)</td>
</tr>
<tr>
<td>Recordings presented in debriefings for clinicians</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid February—June</td>
<td>22</td>
<td>31:2 (3:5)</td>
<td>1515 (569)</td>
</tr>
<tr>
<td>July—mid November</td>
<td>23</td>
<td>31:6 (4:3)</td>
<td>1806 (940)</td>
</tr>
<tr>
<td>Recordings presented to the expert panel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid February—June</td>
<td>19</td>
<td>30:3 (3:1)</td>
<td>1518 (767)</td>
</tr>
<tr>
<td>July—mid November</td>
<td>19</td>
<td>30:3 (3:0)</td>
<td>1511 (744)</td>
</tr>
</tbody>
</table>
Table 2  Scores given by the experts for the 16 items used to evaluate the recordings. Scores are shown for resuscitations in the first versus second 4.5 months of the debriefing period. Also presented are changes between the two periods in percentages relative to the scores in the first period, and inter-observer scoring consistency expressed as Cronbach's $\alpha$. (p) refers to p-value from ANOVA.

<table>
<thead>
<tr>
<th>Score</th>
<th>Mean (SD)</th>
<th>Change between periods</th>
<th>Difference between periods (p)</th>
<th>Consistency (Cronbach $\alpha$)</th>
<th>Difference between experts (p)</th>
<th>Interaction of experts and periods (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First period</td>
<td>Second period</td>
<td>%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were the activities coordinated?</td>
<td>3.14 (0.99)</td>
<td>3.33 (0.89)</td>
<td>6.1</td>
<td>0.28</td>
<td>0.70</td>
<td>0.44</td>
</tr>
<tr>
<td>Was information sought?</td>
<td>3.29 (1.07)</td>
<td>3.66 (0.80)</td>
<td>11.2</td>
<td>0.04</td>
<td>0.57</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Was information shared?</td>
<td>3.47 (1.08)</td>
<td>3.66 (0.80)</td>
<td>5.5</td>
<td>0.21</td>
<td>0.65</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Were problems identified AND pointed out clearly?</td>
<td>3.16 (1.30)</td>
<td>3.41 (0.92)</td>
<td>7.9</td>
<td>0.46</td>
<td>0.65</td>
<td>0.09</td>
</tr>
<tr>
<td>Was each team member supported by the other team members?</td>
<td>3.30 (0.98)</td>
<td>3.43 (0.88)</td>
<td>3.9</td>
<td>0.45</td>
<td>0.80</td>
<td>0.09</td>
</tr>
<tr>
<td>Were human resources used effectively?</td>
<td>3.01 (1.12)</td>
<td>3.35 (0.95)</td>
<td>11.3</td>
<td>0.09</td>
<td>0.77</td>
<td>0.07</td>
</tr>
<tr>
<td>Were next steps planned and discussed/anticipated?</td>
<td>3.11 (1.21)</td>
<td>3.57 (0.86)</td>
<td>14.8</td>
<td>0.09</td>
<td>0.83</td>
<td>0.03</td>
</tr>
<tr>
<td>Following guidelines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providing warmth/avoidance of heat loss</td>
<td>2.82 (1.15)</td>
<td>2.87 (0.98)</td>
<td>1.8</td>
<td>0.79</td>
<td>0.80</td>
<td>0.04</td>
</tr>
<tr>
<td>Positioning/maintaining airway patency</td>
<td>3.17 (0.98)</td>
<td>3.19 (0.86)</td>
<td>0.6</td>
<td>0.90</td>
<td>0.54</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Intubation</td>
<td>3.20 (1.05)</td>
<td>2.48 (0.91)</td>
<td>-22.5</td>
<td>0.01</td>
<td>0.86</td>
<td>0.25</td>
</tr>
<tr>
<td>Monitoring vital signs</td>
<td>3.19 (1.10)</td>
<td>3.41 (0.88)</td>
<td>6.9</td>
<td>0.18</td>
<td>0.64</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Prioritising activities</td>
<td>3.19 (1.00)</td>
<td>3.22 (1.00)</td>
<td>0.9</td>
<td>0.85</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>Control of procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessing neonate's needs</td>
<td>3.12 (1.07)</td>
<td>3.10 (1.02)</td>
<td>-0.6</td>
<td>0.92</td>
<td>0.60</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Planning required actions</td>
<td>2.98 (1.13)</td>
<td>3.04 (1.10)</td>
<td>2.0</td>
<td>0.57</td>
<td>0.65</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Implementing required actions</td>
<td>3.03 (0.98)</td>
<td>2.89 (0.94)</td>
<td>-4.6</td>
<td>0.37</td>
<td>0.57</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Evaluating performed activities</td>
<td>3.29 (0.96)</td>
<td>3.20 (0.86)</td>
<td>-2.7</td>
<td>0.53</td>
<td>0.63</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
In the debriefings, we emphasised teamwork because it is currently a minor component of neonatal resuscitation training. In previous research, Thomas showed a positive link between team training in neonatal resuscitation and performance in the training environment. Our results extend Thomas’ findings to suggest that debriefings using recordings of neonatal resuscitation can also enhance teamwork of clinical practice.

Previous studies have shown that performance of clinical procedures does not change after the presentation of recorded resuscitations. We found the same, except that scores for intubation performance alone were lower for the second period. There is no clear explanation for this, but we note that intubation procedures were performed only in about half of the recordings and the number of the intubations was not balanced between the two periods. Causes for such deterioration should be examined in future studies.

There are several limitations of the study. First, the experts received relatively little training compared with other studies. We let each expert ‘calibrate’ his or her scoring while not enforcing a unified scale that could have impaired the experts’ ability to bring different disciplinary perspectives to the task. Taken together, the low agreement between experts, the average consistency (correlation) and the absence of evidence for any interaction between individual experts and study periods indicate that the experts typically assigned scores of different magnitude but did not differ in their interpretation of trends between the two periods. Our experts had heterogeneous backgrounds; a more uniform group of experts with tighter scoring criteria or more extensive scoring training might produce results with higher levels of significance across the study periods. This is an important area for future research. An ideal assessment tool would have good reliability, would assess individual skills and team function separately and would correlate with important clinical outcome variables, but no such instrument yet exists.

Second, our study design did not let us explore changes in individuals’ procedural performance which are related to behaviours, cognitive and social skills. The overall performance that we captured suggests what may have happened but not why it happened. Future research should also investigate how debriefings affect the practice of individual clinicians.

Third, the results depend on holistic ratings rather than analysis of measurable teamwork behaviours. However, the relationship of measurable teamwork behaviours to the effectiveness of team performance has not been fully specified by researchers, and many evaluations of teamwork still rely on forms of expert judgement.

CONCLUSIONS

Debriefings that provide clinicians with feedback about their practice and that of their peers create learning opportunities and so can help to improve the effectiveness of healthcare teams. Effective teamwork is required for improving the quality of any clinical intervention and for enhancing patient safety. This study provides evidence that debriefings using recordings of actual practice can enhance teamwork in neonatal resuscitation. As in previous studies, the experts specifically noted the value of using recordings as a training tool in debriefings and also for evaluation.

Further studies, controlling more variables, are needed to determine how to make facilitated debriefings using clinical recordings as effective as possible. We also need to identify the best combination of training methods for making neonatal resuscitation a safer practice. Further studies are also needed to extend debriefings into other areas of clinical work where teamwork is important and where patient simulator-based training is limited or impractical on the scale that is needed.

Acknowledgements We are grateful to the medical, nursing and midwifery staff and visiting Medical Officers of Mater Mothers’ Hospital’s Neonatal Critical Care Units and Obstetric Theatres. We thank Pauline Lyon and Clair Doherty for taking part in the experts’ panel. We acknowledge the support given to the study by the parents of the babies who were recorded.

Funding The research was supported by an unrestricted donation (gift towards Nadler’s research to The University of Queensland) from Laerdal Inc. and by a grant from the Mater Foundation. This paper was written as part of Izhak Nadler’s PhD studies at The University of Queensland. During this period Nadler held Endeavour IPRS and UQILAS scholarships from The University of Queensland. The paper has not been previously published or submitted for publication elsewhere.

Competing interests As authors, we declare that we have no conflict of interest associated with the preparation of this paper.

Ethics approval This study was conducted with the approval of Mater Health Services HREC, Brisbane, Australia. The University of Queensland HREC, Brisbane, Australia.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


Presenting video recordings of newborn resuscitations in debriefings for teamwork training

Izhak Nadler, Penelope M Sanderson, Coleen R Van Dyken, et al.

*BMJ Qual Saf* 2011 20: 163-169 originally published online January 6, 2011
doi: 10.1136/bmjqs.2010.043547

References

This article cites 29 articles, 6 of which can be accessed free at:
http://qualitysafety.bmj.com/content/20/2/163.full.html#ref-list-1

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/