

CHAPTER 4

Feasibility and reliability of PRISMA-Medical for specialty-based incident analysis

C. Snijders, T.W. van der Schaaf, H. Klip,
R.A. van Lingen, W.P.F. Fetter, A. Molendijk
on behalf of the NEOSAFE study group

Qual Saf Health Care 2009;18:486–91

ABSTRACT

Aims and objectives: To examine the feasibility and reliability of the PRISMA-Medical method for systematic, specialty-based analysis and classification of incidents in the neonatal intensive care unit (NICU).

Methods: Following the introduction of a Neonatology System for Analysis and Feedback on medical Events (NEOSAFE) in eight tertiary care NICUs and one paediatric surgical ICU, multidisciplinary unit patient safety committees started using PRISMA-Medical to identify root causes of voluntary reported incidents. Committee members were PRISMA-trained and familiar with the department and its processes. We describe the results of PRISMA-analysis of incidents reported during the first year. At $t=3$ months and $t=12$ months after introduction, test cases were performed to measure agreement at three levels of root cause classification using PRISMA-Medical. Interrater reliability was determined by calculating generalised κ values for each level of classification.

Results: During the study period, 981 out of 1786 eligible incidents (55%) were analysed for underlying root causes. In total, 2313 root causes were identified and classified, giving an average of 2.4 root causes for every incident. Although substantial agreement (κ 0.70–0.81) was reached at the main level of root cause classification of the testcases (discrimination between technical, organisational and human failure), and agreement among the committees at the second level (discrimination between skill-based, rule-based and knowledge-based errors) was acceptable (κ 0.53–0.59), discrimination between rule-based errors (the third level of classification) was more difficult to assess (κ 0.40–0.47).

Conclusion: With some restraints, PRISMA-Medical proves to be both feasible and acceptably reliable to identify and classify multiple causes of medical events in the NICU.

INTRODUCTION

In the industrial sector it has been acknowledged that human errors occur and that therefore systems should be designed in such a way that errors are prevented or detected before they develop into a true accident.^{1,2} In clinical practice, there is increasing interest in the development of such systems, since several large studies have confirmed the frequent occurrence of errors in medicine resulting in (possible) patient harm.³⁻⁶ In 2005, a Neonatology System for Analysis and Feedback on medical Events (NEOSAFE) was introduced in The Netherlands to establish specialty-based learning from incidents. Specialty-based reporting systems can be used to collect incidents on a grand scale, to conduct benchmarking, and to identify areas for specialty-based improvement.⁷⁻¹¹ On the other hand, specialty-based systems also require standardised and reliable methods for the collection and analysis of incidents across different units. Little is known on the reliability of methods for systematic incident analysis.^{8,12-14} This is an important issue to address if we want to use these methods as a diagnostic technique to expose specialty-based system weaknesses.¹⁵ The objective of this study was to examine the feasibility and reliability of the PRISMA-Medical method for systematic, specialty-based analysis and classification of incidents in the NICU.

METHODS

PRISMA-Medical

PRISMA (Prevention Recovery Information System for Monitoring and Analysis) was originally developed to manage human error in the chemical process industry, but in the last decade it was also applied in the transportation sector, as well as in health care (PRISMA-Medical).^{8,12,13,16,17} The main goal of PRISMA is to build a quantitative database of incidents (including near misses) and process deviations, in order to facilitate the development and evaluation of system-based preventive strategies. Three main steps can be identified in the PRISMA-Medical method: (1) the Causal Tree incident description method; (2) classification of root causes by the Eindhoven Classification Model (ECM); and (3) formulation of structural measures for improvement. In this study we focus on the first two steps.

Causal trees provide a visual interpretation of the chain of events leading to an incident, without hypothesising about possible causes. They present critical activities and decisions during the development of an incident in chronological order, and show

how activities and decisions are logically related to each other. Causal trees support the fact that nearly all incidents have more than one cause. By continuing to ask “why” of each event (beginning with the top event), a structure of causes and consequences arises, until the root causes are identified at the bottom of the tree. These root causes are subsequently classified by linking them to one of the categories of the ECM (Appendix A, pp 132-133). In some incidents, recovery factors can also be identified. In this study we focus on the failure factors.

The ECM includes both active failures and latent conditions. Active failures are mainly represented by human error. The human section of the model is based on the SRK-model developed by Rasmussen, which distinguishes three levels of behaviour: (1) Skill-based behaviour; (2) Rule-based behaviour; and (3) Knowledge-based behaviour.¹⁸ The medical version of PRISMA also distinguishes Patient-Related Factors. The latent conditions in the ECM involve technical and organisational errors.¹ In total, the ECM distinguishes 20 different types of failure factors, which have been linked to 20 classification codes (Appendix A, pp 132-133). Technical and organisational factors are considered first when classifying root causes, and human failures are considered last. This sequence helps to counteract the tendency to start and stop analysis at the level of the end-user and leave the technical and organisational context of an incident unquestioned. The standardised classification of the causes of incidents through a coding system enables the analysis of multiple incident types or incidents from multiple units at the same time.^{16,17}

Implementation of NEOSAFE

From February through June 2005, NEOSAFE has been implemented in eight of the ten Dutch level III NICUs (14–24 beds per NICU) and one paediatric surgical ICU. In these units, a total of approximately 3500 neonates are admitted each year. Voluntary, non-punitive incident reporting was introduced to establish specialty-based learning.⁹ An incident was defined as ‘any event which could have reduced, or did reduce the safety margin for the patient’.¹⁹ Before the introduction of NEOSAFE, only severe or catastrophic NICU incidents were collected through mandatory reporting to a central hospital committee, without performing systematic analysis routinely.

A multidisciplinary patient safety committee, consisting of at least one physician and three nurses, was formed in each unit. Committee members were recruited on a voluntary basis. Unit employees were encouraged to communicate openly following incidents, and to report incidents to the committee non-anonymously, to enable contact if any additional information was needed during analysis. Personnel were asked to fill in a two-paged incident report form (containing both closed and open items)

immediately after the discovery of an incident. Incidents were either self-reported or reported by personnel who discovered the incident.

In December 2004, members of the committees participated in a two-day PRISMA-Medical course. With the introduction of NEOSAFE, patient safety committees started using PRISMA-Medical to identify root causes of incidents pertaining to medication, arterial and venous lines, and mechanical ventilation. Based on previous literature and expert opinion, these incident categories were thought to have the greatest influence on patient safety in the NICU.¹⁹⁻²¹ To increase reliability of results, each analysis was conducted by two members of the committee, who were PRISMA-trained and familiar with the department and its processes. Committees were encouraged to analyse incidents within two weeks after reporting. Due to the organisational structure in the NICU, the composition of these couples was subject to continuous changes. The patient safety committees managed an electronic database (MS Access) of reported incidents and results of subsequent PRISMA-analysis. Patient and staff confidentiality was ensured by excluding personal identification from the electronic database. To stimulate incident reporting and safety awareness, the committees provided all NICU employees with a summary of incident reports and planned preventive actions on a regular basis. From July 2005 (t=0), incident reports have been aggregated for specialty-based analysis. The local medical research ethics committee (METC Zwolle) waived the need for ethical approval as the study only involved the registration of incidents and their causes. The present study describes the results of the PRISMA-analysis of incidents reported between 1 July 2005 and 30 June 2006.

Reliability testing

In a pilot study in September 2005 (Test 1, t=3 months), the committees applied ECM classification codes to a total of 9 root causes in predefined causal trees, sent by e-mail by the central investigator (CS). The predefined causal trees were constructed by an expert couple (authors CS and TvdS), and were based on randomly selected incidents reported to one of the NICUs (one from each incident category). In November 2005 (t=5 months), a PRISMA-Medical refresher course and consensus meeting was held to evaluate implementation of PRISMA-Medical. As a result of this evaluation, the ECM was further illustrated with examples specific to the NICU ('ECM-NICU').

At the next NEOSAFE-evaluation meeting in June 2006 (Test 2, t=12 months), ECM classifications were performed on new test cases. The test cases consisted of two sets of 45 root causes (15 root causes from each incident category) randomly selected from 1240 root causes identified in the central NEOSAFE database during the first 6 months. Analyst couples were formed within each committee. Couples were divided

into group A and group B. Using the ECM-NICU, each couple in group A classified the first set of 45 root causes, and each couple in group B classified the second set, while an expert couple (authors CS and TvdS) classified all causes.

In both tests, pairwise agreement for each rater couple versus the expert, and agreement among all raters was measured at the three levels of ECM classification (Appendix A, 132-133). Interrater reliability was determined by calculating generalised κ values for each level of ECM classification.^{22,23} We used a macro in SPSS 12.0.1 for Windows to define the number of cases and number of raters. The classification by Landis and Koch was used for interpretation of κ values ($\kappa < 0.00$ = poor; $0.00-0.20$ = slight; $0.21-0.40$ = fair; $0.41-0.60$ = moderate; $0.61-0.80$ = substantial; $0.81-1.00$ = (almost) perfect).²⁴

RESULTS

For organisational reasons, one committee did not supply PRISMA-results during the study period, and was therefore excluded from further analysis. Several of the remaining eight units reported shortage of time in handling the large number of incidents reported after the introduction of the voluntary reporting system. Therefore, units expecting time-management problems were instructed to analyse every third report in order to get a representative sample of PRISMA-analyses. During the study period, 981 out of 1786 eligible incidents (55%) were analysed for underlying root causes. In total, 2313 root causes were identified and classified, giving an average of 2.4 root causes for every incident. Table 1 shows the results of PRISMA-analysis for each incident category.

On average, 64% of all identified causes were classified as human failure, almost one-third of root causes represented technical and organisational failure (9% and 22%, respectively), and another 3% were patient-related factors. Most technical failures were found among incidents with mechanical ventilation, whereas the causes of incidents with medication were most often classified as human failure (Figure 1). As can be seen from Figure 2, verification failures were most often found in this incident category, followed by failures in monitoring and intervention. Besides human failures, a peak in protocol failures was found among incidents with IV lines (over 10% of all identified causes), and another 8% of causes of incidents with IV lines were classified as patient-related factors. Among incidents with mechanical ventilation, 9% of the causes were classified as technical failures that were beyond the control and responsibility of the unit.

Table 1. Results of PRISMA-analysis for each incident category^a

	Medication		Mechanical ventilation		IV lines		Total	
	n	%	n	%	n	%	n	%
Number of analysed incidents								
2005	352 / 619	(57)	79 / 193	(41)	59 / 119	(50)	490 / 931	(53)
2006	332 / 543	(61)	108 / 211	(51)	51 / 101	(50)	491 / 855	(57)
Total	684 / 1162	(59)	187 / 404	(46)	110 / 220	(50)	981 / 1786	(55)
Number of root causes / incident^b								
2005	2.5 (870 / 352)		2.7 (210 / 79)		2.7 (160 / 59)		2.5 (1240 / 490)	
2006	2.1 (698 / 332)		2.4 (259 / 108)		2.3 (116 / 51)		2.2 (1073 / 491)	
Total	2.3 (1568 / 684)		2.5 (459 / 187)		2.5 (276 / 110)		2.4 (2313 / 981)	

^a As identified in the central NEOSAFE database (July 2005 – June 2006).

^b As identified through the Causal Tree incident description method.

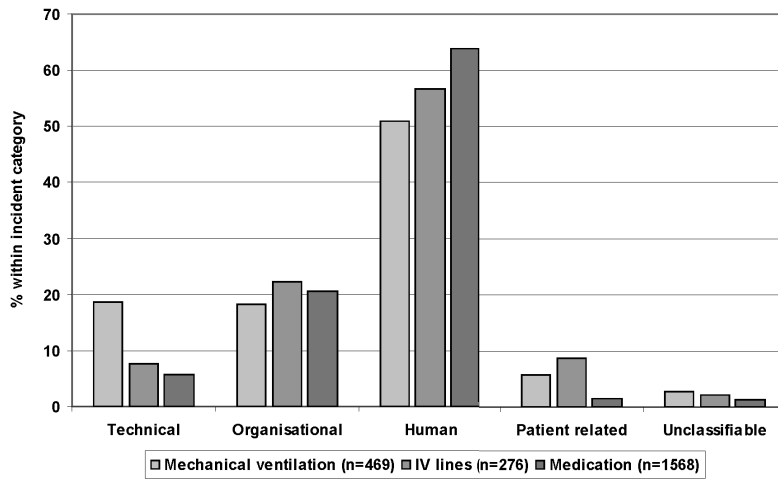


Figure 1. Distribution of root causes (main level) for each incident category (July 2005 – June 2006)

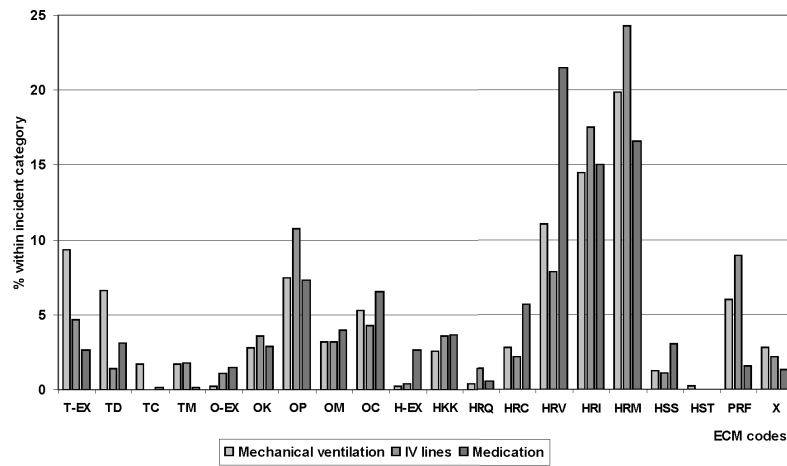


Figure 2. PRISMA-profile of root causes (sub-level) identified in each incident category (July 05 – June 06)*

*See Appendix for an explanation of the ECM codes.

Table 2. ECM classification^a in Test 1

Test cases	ECM classification codes ^b							
	Expert	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7
Accidental cutting of a central venous line								
1. Rules for fixation are unclear	OP	OP	OP	OP	OP	OP	OP	OP
2. Inexperienced nurse	HKK	HRQ	HKK	HRQ	HRQ	HRQ	HKK	HKK
3. Nurse inattention	HRM	HRV	HSS	HRM	HRM	HRI	HRM	HRM
Wrong connection of ventilation tubes								
4. Checks were done incorrectly	HRI	HRI	HRV	HRI	HRI	HRI	OP	HRI
5. Design of machine makes exchange of inspiration and expiration tubes possible	TD	TC	TD	TD	TD	TD	TD	TD
6. Colleague did not check equipment	HRV	OP	HRC	OC	–	HRV	OC	HRV
Medication overdose								
7. Protocol is confusing (Strength used in protocol differs from ampoule strength)	OP	OP	OP	OP	–	OP	HRV	OP
8. Inexperienced nurse	HKK	HRQ	HKK	HRV	HRQ	HKK	HKK	HKK
9. Colleague did not check dose	HRV	OC	HRV	HRC	HRI	HRC	OC	HRI

^aEindhoven Classification Model.

^bCodes relate to different aspects of technical, organisational, or human failure (see Appendix).

In Test 1, seven out of eight committees were able to return the results of testcases in time (Table 2). Multirater agreement was moderate at the sub-level and SRK-level, and substantial at the main level of ECM classification (Table 3).

In Test 2, twenty members from seven patient safety committees participated in the test. One committee was unable to join the NEOSAFE-meeting and therefore did not participate in this test. Each group of testcases was performed by five pairs of raters. In both group A and B, although the extent of agreement between each couple and the expert varied, substantial agreement was reached among all analyst couples at the main level of classification, whereas the SRK-level showed moderate agreement. At the sub-level of classification, moderate agreement was reached among analyst couples in group A. In group B, however, agreement at this level was fair (Table 4).

Table 3 Interrater reliability of ECM classification in Test 1

Level of ECM classification ^a	Raters (patient safety committees)	No. of raters	Agreement in cases (n=9) ^b	% agreement	κ ^c
Main level	1&2&3&4&5&6&7	7	5/7	71	0.81
	1&2&3&4&5&6&7& expert	8	5/7	71	0.79
	1 & expert	2	7/9	78	0.61
	2 & expert	2	9/9	100	1.00
	3 & expert	2	8/9	89	0.79
	4 & expert	2	7/7	100	1.00
	5 & expert	2	9/9	100	1.00
	6 & expert	2	5/9	56	0.22
7 & expert	2	9/9	100	1.00	
SRK-level	1&2&3&4&5&6&7	7	1/7	14	0.53
	1&2&3&4&5&6&7&expert	8	1/7	14	0.49
	1 & expert	2	4/9	44	0.21
	2 & expert	2	8/9	89	0.85
	3 & expert	2	6/9	67	0.50
	4 & expert	2	5/7	71	0.53
	5 & expert	2	8/9	89	0.83
	6 & expert	2	5/9	56	0.42
7 & expert	2	9/9	100	1.00	
Sub-level	1&2&3&4&5&6&7	7	1/7	14	0.47
	1&2&3&4&5&6&7&expert	8	1/7	14	0.41
	1 & expert	2	3/9	33	0.21
	2 & expert	2	6/9	67	0.60
	3 & expert	2	5/9	56	0.49
	4 & expert	2	4/7	57	0.49
	5 & expert	2	6/9	67	0.61
	6 & expert	2	5/9	56	0.47
7 & expert	2	8/9	89	0.86	

^a See Appendix for classification codes in each level.

^b Cases with missing values were excluded from generation of sums of agreement.

^c Classification by Landis and Koch.

Table 4. Interrater reliability of ECM classification in Test 2.

Level of ECM classification ^a	Raters (couples)	No. of raters	Group A			Group B		
			Agreement in cases (n=45) ^b	% agreement	κ^c	Agreement in cases (n=45) ^b	% agreement	κ^c
Main level	1&2&3&4&5	5	18 / 26	69	0.74	18 / 26	69	0.70
	1&2&3&4&5&expert	6	18 / 26	69	0.77	18 / 26	69	0.74
	1 & expert	2	36 / 40	90	0.83	34 / 36	94	0.87
	2 & expert	2	32 / 39	82	0.70	35 / 42	83	0.66
	3 & expert	2	33 / 37	89	0.84	34 / 45	76	0.52
SRK-level	4 & expert	2	38 / 41	93	0.88	34 / 43	79	0.60
	5 & expert	2	30 / 35	86	0.78	32 / 36	89	0.77
	1&2&3&4&5	5	12 / 26	46	0.58	13 / 26	50	0.59
	1&2&3&4&5&expert	6	12 / 26	46	0.60	13 / 26	50	0.61
	1 & expert	2	26 / 40	65	0.54	30 / 36	83	0.73
Sub-level	2 & expert	2	23 / 39	59	0.47	38 / 42	90	0.52
	3 & expert	2	26 / 37	70	0.65	29 / 45	64	0.50
	4 & expert	2	29 / 41	71	0.61	27 / 43	63	0.47
	5 & expert	2	24 / 35	69	0.60	28 / 36	78	0.64
	1&2&3&4&5	5	6 / 26	23	0.44	4 / 26	15	0.40
Sub-level	1&2&3&4&5&expert	6	6 / 26	23	0.48	4 / 26	15	0.45
	1 & expert	2	24 / 40	60	0.56	20 / 36	56	0.50
	2 & expert	2	21 / 39	54	0.49	25 / 42	60	0.55
	3 & expert	2	19 / 37	51	0.47	21 / 45	47	0.40
	4 & expert	2	23 / 41	56	0.51	24 / 43	56	0.49
5 & expert	2	21 / 35	60	0.56	19 / 36	53	0.46	

^a See Appendix for classification codes in each level.

^b Cases with missing values were excluded from generation of sums of agreement.

^c Classification by Landis and Koch.

DISCUSSION

Our results show that the PRISMA-Medical method has a great potential to contribute to the identification of system failures that lead to incidents in the NICU. As both active failures (human failure) and latent conditions (technical and organisational failures) that caused the incident were discovered, the total profile of root causes identified through PRISMA-analysis can be used to provide a more realistic view of how the system is actually working. The variation in causal coding between incident categories suggests that different approaches are needed for each incident category to prevent recurrence. However, our practical experience shows that feasibility of PRISMA-Medical for specialty-based analysis strongly depends on the availability of time and PRISMA-trained personnel. A possible solution to this problem is the selection and prioritisation of certain incident types, such as high-risk incidents, for PRISMA-analysis.⁷ Moreover, in the early stages of incident reporting caution is needed when interpreting results of PRISMA-analysis for the development of specialty-based preventive strategies, as interrater reliability tests pointed out that agreement among patient safety committees depends on the level of classification. Although substantial agreement was reached at the main level of ECM classification, and agreement among the committees within the SRK-level was acceptable, discrimination between rule-based errors (qualification, coordination, verification, intervention and monitoring) was more difficult to assess. Given the fact that some rater couples scored better than others (Table 3: rater 2, 5 and 7), agreement probably also depends on other factors such as level of training or time working in the unit.

Our study is the first extensive, multi-centre reliability study that is based on both major and minor incidents. Few other studies have investigated reliability of PRISMA-Medical. A study among three Health-Care Inspectors and a PRISMA-expert in The Netherlands found acceptable reliability scores at all levels of ECM classification, with increased reliability after repeated education (M. Habraken, 2005: Better care for incidents in health care, MSc thesis, Eindhoven University of Technology). However, this study was based only on severe, mostly fatal medication incidents, which is quite different from our study design.²⁵

The value of PRISMA-Medical for specialty-based incident analysis has been described previously. A USA study in transfusion medicine reported that PRISMA-Medical is a very useful tool in sense making of individual and specialty-based, aggregate causal data.^{7,12,17} The percentage of human causes in our study (64%) was rather high compared with their study (46%), and compared with a study on incidents

reported to The Netherlands Health Care Inspectorate (42%).^{17,25} Although this may reflect the actual situation in the NICU, it may also be due to remaining ‘person-oriented’ biases during the root cause analysis. If the latter is the case, one should expect to observe a decrease in the percentage of human failures after repeated education. Moreover, although several causes were identified for each incident, we did not test interrater reliability of the Causal Tree description method. Therefore, future research should develop methods to test the reliability of the Causal Tree description method as well. A pilot study has already shown the importance of this issue (H. Wijers 2007: Reporting of deviations, MSc thesis (in Dutch), Eindhoven University of Technology).

This study has some limitations. First, only 55% of all eligible reported incidents were analysed. Although our sample size is quite high, and we tried to minimise selection bias by analyzing every third report in case of time-management problems, this may have affected the final profile of root causes. Second, valid comparisons between the first and second reliability test were difficult to assess because of natural selection of analyst couples, as a reflection of the clinical setting. Third, the two-day PRISMA-course was a shortened version of the original three-day course. This may have affected the reliability of classification negatively, especially during the time before the PRISMA refresher course (Table 3).

In conclusion, with some restraints, PRISMA-Medical proves to be both feasible and acceptably reliable for specialty-based identification and classification of multiple causes of incidents in the NICU. The system approach states that preventive actions should primarily aim at the structural (latent) system failures in order to prevent inevitable human failures, which are usually at the end of the incident cascade, leading to patient injury.¹ In this light, preventive strategies in the NICU should be aimed at the technical and organisational weaknesses first, rather than at human failure. Future research should study the predictive validity of PRISMA by investigating the effect of system-based interventions on the prevalence of these system failures, as well as on patient harm. Moreover, future studies should also examine the role of the PRISMA Error Recovery Factors in the prevention of incidents.

REFERENCES

1. Reason J. Human error: models and management. *BMJ* 2000;320:768–70.
2. Barach P, Small SD. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. *BMJ* 2000;320:759–63.
3. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 1991;324:370–6.
4. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med* 1991;324:377–84.
5. Kohn LT, Corrigan JM, Donaldson MS. To Err is Human. Building a Safer Health System. Washington DC: National Academy Press, 2000.
6. Aspden PH, Corrigan JM, Wolcott J. Patient safety: achieving a new standard for care. Washington DC: National Academy Press, 2004.
7. Kaplan HS, Rabin Fastman B. Organization of event reporting data for sense making and system improvement. *Qual Saf Health Care* 2003;12:ii68–72.
8. Van der Schaaf TW. Near miss reporting in the chemical process industry. Eindhoven, The Netherlands: Eindhoven University of Technology, 1992, Thesis.
9. Leape LL. Reporting of adverse events. *N Engl J Med* 2002;347:1633–8.
10. Snijders C, van Lingen RA, Molendijk A, Fetter WPF. Incidents and errors in neonatal intensive care: A review of the literature. *Arch Dis Child Fetal Neonatal Ed* 2007;92:391–8.
11. Suresh G, Horbar JD, Plsek P, Gray J, Edwards WH, Shiono PH et al. Voluntary anonymous reporting of medical errors for neonatal intensive care. *Pediatrics* 2004;113:1609–18.
12. Battles JB, Kaplan HS, Van der Schaaf TW, Shea CE. The attributes of medical event-reporting systems. Experience with a prototype medical event-reporting system for transfusion medicine. *Arch Pathol Lab Med* 1998;122:231–8.
13. Van Vuuren W. Organisational failure: an exploratory study in the steel industry and the medical domain. Eindhoven, The Netherlands: Eindhoven University of Technology, 1998, Thesis.
14. Wright L, Van der Schaaf TW. Accident versus near miss causation: a critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors. *J Hazard Mater* 2004;111:105–10.
15. Vincent CA. Analysis of clinical incidents: a window on the system not a search for root causes. *Qual Saf Health Care* 2004;13:242–3.
16. Van der Schaaf TW, Habraken MMP. PRISMA-Medical: a brief description. Eindhoven: Eindhoven University of Technology, 2005. Available at: <http://www.who-icps.org/resources/PRISMA-Medical.pdf> (accessed 25 February 2008).
17. Kaplan HS, Battles JB, Van der Schaaf TW, Shea CE, Mercer SQ. Identification and classification of the causes of events in transfusion medicine. *Transfusion* 1998;38:1071–81.
18. Rasmussen J. The definition of human error and a taxonomy for technical systems design. In: Rasmussen J, Duncan K, Leplat J, eds. New technology and human error. London: Wiley, 1987.

19. Beckmann U, Baldwin I, Hart GK, Runciman WB. The Australian Incident Monitoring Study in Intensive Care: AIMS-ICU. An analysis of the first year of reporting. *Anaesth Intensive Care* 1996;24:320–9.
20. Frey B, Kehrer B, Losa M, Braun H, Berweger L, Micallef J et al. Comprehensive critical incident monitoring in a neonatal-pediatric intensive care unit: experience with the system approach. *Intensive Care Med* 2000;26:69–74.
21. Kaushal R, Bates DW, Landrigan C, McKenna KJ, Clapp MD, Federico F et al. Medication errors and adverse drug events in pediatric inpatients. *JAMA* 2001;285:2114–20.
22. Shoukri MM. Measures of interobserver agreement. Boca Raton: Chapman & Hall/CRC, 2004.
23. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378–82.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
25. Habraken MMP. [Better analysis of incidents: PRISMA-method improves Inspectorate's knowledge of medical mishaps]. *Medisch Contact* 2005;22:940-3 (Dutch).