

DEVELOPING RESEARCH AND PRACTICE

Development and testing of an assessment instrument for the formative peer review of significant event analyses

J McKay, D J Murphy, P Bowie, M-L Schmuck, M Lough, K W Eva

Qual Saf Health Care 2007;16:150–153. doi: 10.1136/qshc.2006.020750



Supplementary Appendix available online at <http://qshc.bmj.com/supplemental>

See end of article for authors' affiliations

Correspondence to:
Dr J McKay, NHS Education for Scotland, Postgraduate General Practice Education, 2 Central Quay, 89 Hydepark Street, Glasgow G3 8BW, UK; john.mckay@nes.scot.nhs.uk

Accepted 5 December 2006

Aim: To establish the content validity and specific aspects of reliability for an assessment instrument designed to provide formative feedback to general practitioners (GPs) on the quality of their written analysis of a significant event.

Methods: Content validity was quantified by application of a content validity index. Reliability testing involved a nested design, with 5 cells, each containing 4 assessors, rating 20 unique significant event analysis (SEA) reports (10 each from experienced GPs and GPs in training) using the assessment instrument. The variance attributable to each identified variable in the study was established by analysis of variance. Generalisability theory was then used to investigate the instrument's ability to discriminate among SEA reports.

Results: Content validity was demonstrated with at least 8 of 10 experts endorsing all 10 items of the assessment instrument. The overall G coefficient for the instrument was moderate to good ($G > 0.70$), indicating that the instrument can provide consistent information on the standard achieved by the SEA report. There was moderate inter-rater reliability ($G > 0.60$) when four raters were used to judge the quality of the SEA.

Conclusions: This study provides the first steps towards validating an instrument that can provide educational feedback to GPs on their analysis of significant events. The key area identified to improve instrument reliability is variation among peer assessors in their assessment of SEA reports. Further validity and reliability testing should be carried out to provide GPs, their appraisers and contractual bodies with a validated feedback instrument on this aspect of the general practice quality agenda.

Significant event analysis (SEA) is a method of reflective learning that is strongly promoted as a mechanism for improving patient safety and healthcare risk in the UK.¹ It typically involves an attempt to review in-depth an event identified as "significant" by any member of the healthcare team. Given the complexity and uncertainty in general medical practice, SEA may offer both an understanding of where care processes can fail patients and the means to implement systemic change in relatively non-bureaucratic organisations.² The National Patient Safety Agency—a special health authority created to co-ordinate learning from patient safety incidents in the NHS—has recently recommended that primary care teams should analyse significant events as part of their safety culture (box 1).

Evidence of the ability of general practitioners (GPs) and others to verifiably undertake SEA effectively is limited.^{5–8} This is highly important because superficial or informal discussion of an event is unlikely to lead to understanding, learning and the implementation of necessary change.^{3,9}

One method of informing on the quality of SEA is through external peer review. Peer review can be described as the critical evaluation of a specific aspect of a practitioner's performance by professional colleagues, preferably achieved through use of a reliable and structured instrument.^{10,11} However, few peer assessment instruments have been evaluated sufficiently with regard to validity and reliability to justify their widespread use.¹²

In the west of Scotland region, a voluntary educational model for the external peer review of SEA reports has been available to all GPs as part of their continuing professional development since 1998.^{5–8,13} This involves a submitted written report being sent to two trained GP assessors, chosen from a group of 20, who independently review it using a structured assessment instrument and provide educational feedback.¹³

Given the perceived importance of the SEA technique to the patient safety agenda,^{4,14} the development of a valid and reliable assessment instrument with which to facilitate the educational peer review of SEA would be highly desirable. In this way, a professional judgement could be made on the quality of the event analysis in question, and a formative feedback provided for consideration. Raising the standard of event analyses undertaken by GPs and their teams creates a clear potential to further enhance learning and the quality of patient care.

This study was undertaken to establish the content validity of a new peer assessment instrument, elucidate aspects of its reliability and investigate possible subsample differences, which would be relevant for generalising to a wider population of GPs.

METHODS

Content validity

The developmental stage to assimilate the proposed items for the instrument was carried out independently by three of the authors (JM, PB, DJM). This work was informed by previous focus group interviews with the west of Scotland Audit Development Group.¹⁵ These discussions used Marinker's six essential steps in formulating an enquiry into a significant event (REPOSE) to identify a set of items and domains that could be applied to a selection of events considered "significant" by the group.¹⁶ Agreement was reached on four criteria considered "essential" for assessment of a significant event analysis.¹⁵ Together with previous research,^{1,9} these criteria were developed to generate relevant domains and items. These were discussed by the three authors until consensus was

Abbreviations: CVI, content validity index; GP, general practitioner; SEA, significant event analysis

Box 1: Significant event analysis and the link with patient safety

- Significant event analysis (SEA) is a retrospective, qualitative clinical audit technique based on a synthesis of traditional case discussion and the principles underlying the critical incident technique.³
- A significant event is defined as “any event thought by anyone in the team to be significant in the care of patients or the conduct of the practice”. This normally involves suboptimal practice, but could also be an example of excellent care.¹
- A typical event analysis involves a non-threatening, structured investigation (normally team-based) to establish why an event happened, to learn from it and to introduce change where necessary.
- SEA has been recommended by the National Patient Safety Agency for the analysis of patient safety incidents in primary care, which have resulted in a “near miss” or low to moderate patient harm.⁴
- SEA facilitates identification of reportable safety incidents to local health organisations or national reporting systems to enable learning and sharing among health-care teams.⁴
- SEA is arguably more acceptable and feasible as an investigation technique in general practice than more established methods such as root cause analysis, which require more extensive training, time commitment and expense.

achieved on the items to be included in a content validity exercise.

The proposed instrument consisted of 10 items each rated on a 7-point adjectival scale, with anchor points ranging from absent to excellent (see supplementary appendix, available at <http://qshc.bmj.com/supplemental>). This was sent to 10 GP experts, identified as being well informed in SEA because they were experienced peer assessors or had published on SEA in peer-reviewed journals.

The relevance and appropriateness of each item was then assessed by asking the experts to rate each item and the instrument as a whole using a 4-point scale to create a content validity index (CVI). In all, 8 out of 10 experts were required to endorse each item by assigning a rating of at least 3 out of 4, to establish content validity beyond the 0.05 level of significance.¹⁷ This was determined to provide sufficient evidence for inclusion of each item as part of the final instrument. Experts were also asked to identify any missing items that they deemed important for inclusion when considering the quality of a SEA report.

Reliability testing

Participants and assessment exercise

The proposed instrument was introduced on a training day to the west of Scotland Audit Development Group from which all the peer assessors are drawn (box 2). The role of the assessors and any clarification points around using the instrument were discussed. Further issues raised by assessors were to be emailed to the authors as they arose, or discussed at three-monthly follow-up meetings.

All 20 assessors took part in a reliability marking exercise. A nested design consisting of five cells, each with four raters, was used. Members of each cell marked 20 separate SEA reports,

unique to that cell, using the proposed new assessment instrument. The exercise was repeated after 1 month, with the raters in each cell marking the same unique 20 SEA reports. The 20 SEA reports for each cell consisted of 10 submitted by GP principals (experienced doctors) and 10 from GP registrars (doctors-in-training).

Data analysis

A repeated-measures analysis of variance was undertaken using BMDP software, and analysed to establish the variance attributable to each study variable (SEA reports, $n = 100$, 20 per cell; raters, $n = 20$, 4 per cell; time, $n = 2$; items, $n = 10$). Generalisability theory (G theory), a statistical technique for determining the extent to which ratings consistently discriminate between subjects of measurement (ie, determines the reliability of observations), was used to investigate the instrument's ability to differentiate the quality of SEA reports.¹⁸ The internal consistency (a measure of item homogeneity), intra-rater reliability (agreement within rater across occasions) and inter-rater reliability (agreement among raters) were all calculated. These statistics range from 0 to 1, with 1 indicating perfect reliability.

To avoid the potential of artificially inflating the heterogeneity of the sample (and hence the reliability), we report separate analyses on the SEA reports provided by the GP principals and GP registrars.

RESULTS

Content validity

At least 8 out of 10 experts endorsed all 10 items listed in supplementary appendix (available online at <http://qshc.bmj.com/supplemental>) and the overall instrument, indicating a statistically significant proportion of agreement regarding the content validity of the assessment instrument ($p < 0.05$). No additional items were identified for inclusion.

Reliability

The G coefficients obtained for the overall test reliability, internal consistency and inter/intra-rater reliability values for the instrument when used to assess SEA reports are shown in table 1 for GP principals and in table 2 for GP registrars.

The internal consistency of the instrument was high when averaged over all items for both GP principals ($G = 0.94$) and GP registrars ($G = 0.89$). This indicates that the items included in the instrument are correlated with one another to a sufficient extent. Item reliability of a single item is low, however, indicating that no one item should be deemed a reliable indicator of SEA quality.

Box 2: Characteristics of west of Scotland Audit Development Group

- 20 principals in general practice with a minimum of 8 years experience, trained in peer review.
- All have a minimum of 5 years experience as peer reviewers of criterion audit and significant event analysis reports for continuing professional development and summative assessment.
- 18 (90%) are members or fellows of the Royal College of general practitioners (GPs).
- 2 (10%) are GP appraisers.
- 10 (50%) are GP registrar trainers.
- A further 3 (15%) have other general practice educational roles (eg, associate adviser, undergraduate tutor).

Table 1 Calculated reliability coefficients for general practitioner principals' significant event analysis reports marked using the peer review instrument (expressed with 95% CI)

	Overall	Internal consistency	Intra-rater	Inter-rater
Single item	0.25 (0.17 to 0.33)	0.62 (0.55 to 0.68)	0.64 (0.57 to 0.70)	0.31 (0.23 to 0.39)
Total score (ie, average of average of all nine items)	0.73 (0.68 to 0.73)	0.94 (0.93 to 0.95)	0.78 (0.73 to 0.82)	0.64 (0.57 to 0.70)
Global score (item 10)	0.80 (0.76 to 0.84)	N/A	0.70 (0.64 to 0.75)	0.43 (0.35 to 0.51)

N/A, not applicable.

The high intra-rater coefficients for SEA reports undertaken by GP principals (0.78) and GP registrars (0.71) suggest that individual assessors' opinions regarding the quality of each SEA report are reasonably stable over time.

The moderate G coefficients for inter-rater reliability, assessed using the average of scores provided by all four raters, for both GP principals (0.64) and GP registrars (0.6), indicate that there may be room for future calibration of assessors to ensure that consistent feedback is provided. Decision study analyses suggest that 10 raters are required for the average score to achieve an inter-rater reliability of $G > 0.8$.

The correlation between the global rating scale and the sum of the nine specific items was strong ($r = 0.87$ and 0.90 for GPs and registrars, respectively). A comparison of the mean scores between GP principals' and GP registrars' SEA reports is shown in table 3 and demonstrates no difference between the two groups.

DISCUSSION

This study demonstrates that the content validity and reliability of the assessment instrument are adequate, providing the first steps towards validating an instrument for providing educational feedback to GPs on the quality of their written SEA reports. The findings highlight specific areas that could improve instrument reliability, with the key area being variation among peer assessors in their assessment of SEA reports. Consistent with previous research,⁸ no difference was found in the quality ratings assigned to SEA reports completed by GP principals or GP registrars.

Limitations of the study

Validity testing

This instrument has been developed by GPs and so is doctor-centred, despite the frequent team involvement in significant events and their analyses.¹ Our "expert" raters were simply well-informed individuals as the number of individuals with sufficient knowledge and experience to be deemed truly an expert is limited (and it must be acknowledged, poorly defined).^{19, 20} The CVI exercise was adequate, but a different approach such as the Delphi technique may have added more depth to the process.

Reliability

The significant events chosen for peer review were self-selected. The finding that most SEA reports were rated as having a global score of ≥ 4 may indicate a bias towards submission of reports with which the submitting doctor feels comfortable.¹³ The impact of this limitation, however, should have worked against the observation of sufficient reliability.

In addition, it should be noted that the raters were individuals with extensive experience with SEA who had considerable opportunity to discuss how to interpret the rating task. Further study is required to determine whether or not similar findings would be achieved with less experienced raters. In addition, although the instrument is designed to provide written as well as numerical feedback, we analysed only numerical data. For a formative instrument, written feedback may be at least as important to the submitting doctor. This aspect of the instrument therefore requires its own separate evaluation.

SEA reports

Finally, we recognise that the SEA report content is merely a proxy indicator for what actually happened or was decided in practice. Personal and recall bias in addition to problems of understanding, interpretation and judgement may influence what is reported. An individual's ability to articulate the event analysis in writing may also be a factor.

Context/implications

There is no universally agreed method for the analysis of significant events. Our instrument mirrors previously suggested approaches,^{1, 4, 13} but is unique in providing written feedback by peers. A strength of this instrument is that it is for use in the workplace, and has been tested using events taking place as a result of actual experience. Systems to improve patient safety have been difficult to implement in primary care. Using an instrument that is based on educational theory and research methods—as opposed to simply applying one based on intuition—provides an element of scientific rigour when applied in this patient safety context. This should add to the potential attractiveness and relevance of the instrument and, therefore, to its impact.

The study demonstrated content validity, but further work is required to confirm the overall instrument validity. The high G

Table 2 Calculated reliability coefficients for general practitioner registrars' significant event analysis reports marked using the peer review instrument (expressed with 95% CI)

	Overall	Internal consistency	Intra-rater	Inter-rater
Single item	0.18 (0.11 to 0.26)	0.48 (0.4 to 0.55)	0.55 (0.48 to 0.62)	0.27 (0.19 to 0.35)
Total score (ie, average of all nine items)	0.71 (0.65 to 0.76)	0.89 (0.87 to 0.91)	0.71 (0.65 to 0.76)	0.6 (0.53 to 0.66)
Global score item 10	0.83 (0.80 to 0.86)	N/A	0.58 (0.51 to 0.64)	0.42 (0.34 to 0.50)

N/A, not applicable.

Table 3 Comparison of the mean scores between general practitioner principals' and general practitioner registrars' significant event analysis reports

	Principals	GP registrars	F value	p Value
Mean score over items 1–9 (1SD)	4.81 (0.79)	4.88 (0.63)	0.29	0.59
Mean global score of SEA reports (1SD)	4.85 (0.80)	4.94 (0.99)	0.3	0.58

GP, general practitioner; SEA, significant event analysis.

coefficients observed indicate that the domains and items are inter-related, and the CVI indicates that our judges considered the questions to be relevant, providing the first steps towards enhancing the assessment of significant event analyses.

Context specificity was not considered, so the instrument cannot currently be claimed to be useful for assessing a GP's proficiency in applying the SEA technique. The purpose of this instrument is to facilitate educational feedback on the merits and drawbacks of individual SEA reports. There is increasing recognition that professional self-regulation should not rely on unguided self-assessments for the improvement of practice.^{21–22} It is hoped that GPs would find feedback provided by external assessors using this form helpful in highlighting particular issues that could further improve their analysis, thus enhancing the quality or standard of future event analyses and, in turn, the safety of the GPs' patients.

The largest degree of instrument error when providing feedback is the variation among peer assessors. This is a common difficulty for assessment instruments.^{23–24} The moderately large G coefficients for intra-rater reliability imply a reasonable degree of instrument stability when used by individual peer reviewers to assess reports at different points of time. The lower inter-rater reliability is more likely, therefore, to be related to calibration issues among the assessors rather than to the robustness of the instrument. Further training of assessors or the continued use of multiple assessors when evaluating each SEA is necessary. This is particularly important if the instrument is to be used by other professional colleagues in different clinical settings.

An ideal educational tool would be "supportive and individualised, yet uniformly applied".²⁵ This is especially relevant, given the role of SEA in patient safety. A successful formative instrument should, therefore, give information via interpretable numerical scores and written comments, and should be used in conjunction with facilitated feedback.²⁶ Our model fits with both concepts because it promotes self-directed (and team-directed) reflective learning and provides written peer feedback.

SEA is part of GP appraisal in NHS Scotland,²⁷ the GMS contract in the UK,²⁸ and has been proposed as a component of revalidation.²⁹ However, uniform guidance on how it should be applied and monitored is lacking. Participation in our SEA model may demonstrate to patients, appraisers and healthcare organisations the willingness of the GPs to submit aspects of their own work for external review as part of an educational process.¹⁴ This would confirm that the GP is verifiably reflecting on how patient care can be improved as part of the clinical governance agenda.

Future work

The study findings justify further development of the instrument, particularly to widen validity testing, calibrate assessors and investigate the educational impact on patient safety.

ACKNOWLEDGEMENTS

We thank Dr J Stead, Exeter, Professor M Pringle, Nottingham, Professor G Elwyn, Swansea, Professor C Bradley, Cork, and Members of the west of Scotland Audit Development Group for their input into the development of the content of the peer review instrument. We also thank the west of Scotland Audit Development Group for their work on the reliability testing of the instrument.

Authors' affiliations

J McKay, Department of Postgraduate Medicine, University of Glasgow, Glasgow, UK

D J Murphy, P Bowie, M Lough, NHS Education for Scotland, Glasgow, UK

M-L Schmuck, K W Eva, McMaster University, Hamilton, Ontario, Canada

Funding: NHS Education for Scotland.

Competing interests: None.

REFERENCES

- 1 **Pringle M, Bradley CP, Carmichael, et al.** *Significant event auditing. A study of the feasibility and potential of case-based auditing in primary medical care*, Occasional Paper No 70. Exeter: Royal College of General Practitioners, 1995.
- 2 **Wilson T, Pringle M, Sheikh A.** Promoting patient safety in primary care. *BMJ* 2001;**323**:582–3.
- 3 **Bradley CP.** Turning anecdotes into data: the critical incident technique. *Fam Pract* 1992;**9**:98–103.
- 4 **National Patient Safety Agency.** *Seven steps to patient safety for primary care*. London: National Patient Safety Agency, 2005.
- 5 **McKay J, Bowie P, Lough M.** Evaluating significant event analyses: implementing change is a measure of success. *Educ Prim Care* 2003;**14**:34–8.
- 6 **Bowie P, McKay J, Lough M.** Peer assessment of significant event analyses: being a trainer confers an advantage. *Educ Prim Care* 2003;**14**:338–44.
- 7 **Bowie P, McKay J, Norrie J, et al.** Awareness and analysis of a significant event by general practitioners: a cross sectional survey. *Qual Saf Health Care* 2004;**13**:102–7.
- 8 **McKay J, Bowie P, Lough JRM.** Variation in the ability of general medical practitioners to apply two methods of clinical audit: a 5-year study of assessment by peer review. *J Eval Clin Pract* 2006;**12**:622–9.
- 9 **Westcott R, Sweeney G, Stead J.** Significant event audit in practice: a preliminary study. *Fam Pract* 2000;**17**:173–9.
- 10 **Norcini JJ.** Peer assessment of competence. *Med Educ* 2003;**37**:539–43.
- 11 **Grol R.** Quality improvement by peer review in primary care: a practical guide. *Qual Saf Health Care* 1994;**3**:147–52.
- 12 **Evans R, Elwyn G, Edwards A.** A review of instruments for peer assessment of physicians. *BMJ* 2004;**328**:1240–3.
- 13 **Bowie P, McKay J, Dalgetty E, et al.** A qualitative study of why general practitioners may participate in significant event analysis and educational peer assessment. *Qual Saf Health Care* 2005;**14**:185–9.
- 14 **Donaldson L.** *Good doctors, safer patients. A report by the Chief Medical Officer*. London: The Stationary Office, 2006:198.
- 15 **Lough JRM.** *The development of integrated audit for the training of general practitioners*, [MD thesis]. Glasgow, UK: University of Glasgow, 2003.
- 16 **Marinker M.** *Standards in: medical audit in general practice*, London: The MSD Foundation, 1990:12–13.
- 17 **Lynn MR.** Determination and quantification of content validity. *Nurs Res* 1986;**35**:383–5.
- 18 **Cronbach LJ, Gleser GC, Nanda H, et al.** *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: John Wiley, 1972.
- 19 **Waltz CW, Bausell RB.** *Nursing research: design statistics and computer analysis*. Philadelphia: FA Davis, 1981:71.
- 20 **Fink A, Kosecoff J, Chassin M, et al.** Consensus methods, characteristics and guidelines for use. *Am J Public Health*, 1984;**74**, 729–34.
- 21 **Eva KW, Regehr G.** Self-assessment in the health professions: a reformulation and research agenda. *Acad Med* 2005;**80**(Suppl):S46–54.
- 22 **Regehr G, Eva K.** Self-assessment, self-direction, and the self-regulating professional. *Clin Orthop Relat Res* 2006;**449**:34–8.
- 23 **Roberts C, Cromarty I, Crossley J, et al.** The reliability and validity of a matrix to assess completed reflective personal development plans of general practitioners. *Med Educ* 2006;**40**:363–70.
- 24 **Ramsay PG, Wenrich MD.** Peer ratings: an assessment tool whose time has come. *J Gen Intern Med* 1999;**14**:581–2.
- 25 **Norman GR, Shannon S, Marrin ML.** The need for needs assessment in continuing medical education. *BMJ* 2004;**328**:999–1001.
- 26 **Sargeant J, Mann K, Ferrier S.** Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;**39**:497–504.
- 27 **Scottish Executive, NHS Education for Scotland, RCGP (Scotland) & BMA (Scotland).** *GP appraisal: a brief guide*. Edinburgh: Scottish Executive, 2003.
- 28 **Department of Health.** *New GMS contract 2006/7*. London: Stationery Office, 2006.
- 29 **The Shipman Inquiry.** *Dame Janet Smith Fifth Report Volume 3. Safeguarding patients: lessons from the past-proposals for the future*. London: The Stationary Office, 2004.