Significant events in general practice: issues involved in grading, reporting, analyses and peer review.

JOHN MCKAY BSc (Hons), MB ChB, FRCGP

Thesis submitted for the degree of M.D.

Division of Community Based Sciences
Faculty of Medicine
UNIVERSITY OF GLASGOW

May 2009

ACKNOWLEDGEMENTS

My sincere thanks go to my research supervisor, Dr Murray Lough MD FRCGP, for his expert advice, unfailing support and encouragement throughout this thesis.

I was also fortunate to be able to call on the experience and counsel of my research advisor, Professor Stuart Murray MD PhD FRCGP.

I wish also to thank and acknowledge the under-noted individuals for their assistance with specific aspects of this thesis:

- Dr Paul Bowie, Associate Adviser and Co-Researcher, for his helpful comments on the design and analysis of the postal questionnaires used in the two surveys of GPs in Greater Glasgow and Clyde described in Chapters 6 and 7. He also assisted by jointly analysing for content the SEA reports studied in Chapter 8, the focus group data reported in Chapter 11 and the semi-structured interview data reported in Chapter 12.

- Dr Lindsey Pope, Higher Professional Trainee, for jointly analysing the semi-structured interview data reported in Chapter 12.

- Dr Annabel Shepherd, Higher Professional Trainee, for jointly analysing the focus group data reported in Chapter 11.

- Dr Lilian Murray, University of Glasgow, for advice about handling the data in Chapter 6 and for statistical advice and undertaking data analysis on aspects of the findings reported in Chapters 7 and 12, respectively.

- Dr Kevin Eva and Ms Mary-Lou Schmuck, McMaster University, Ontario, Canada, for advice on the design of the studies reported in Chapters 9 and 10 and for undertaking the subsequent data analysis.

I would like to thank members of the West of Scotland Audit Development Group who carried out the review of SEA reports described in Chapters 8-10 and Chapter 13 and attended the focus group discussions reported in Chapter 11.

The data generated for this thesis were made possible by the contributions of many general practitioners in the west of Scotland, who submitted SEA reports for assessment by peer review, completed survey questionnaires and agreed to take part in semi-structured interviews. I extend my grateful thanks to all of them for their assistance.

Finally, I wish to thank and acknowledge the contribution of my employer, NHS Education for Scotland, for contributing to my course fees.

DECLARATION

I declare that the contents of this thesis are my own work except where previously acknowledged on Page 1.

All text, tables, figures, data collection proforma and statistical analysis have been prepared the author using Microsoft Word, Microsoft Excel, Epi-Info Version 6.01, and MINITAB Release 13.20.

ETHICAL APPROVAL

As research data for the survey data reported in Chapter 6 had been collected or were in the process of being collected prior to the University of Glasgow research governance requirement of October 2002, which states that students must seek ethical approval for all research studies undertaken as part of a higher degree, it was agreed that no ethical approval for this study was required.

The study reported in Chapter 7 was approved by the NHS Greater Glasgow Primary Care Ethics Committee, REC Reference: 05/S0701/48

The study reported in Chapter 8 was approved by the NHS Greater Glasgow Primary Care Ethics Committee, REC Reference: 04/SO701/71

The studies reported in Chapters 9-13 were assessed by the Multi-Centre Research Ethics Committee for Scotland as not requiring ethical review, REC Reference: 05/MRE00/59 and letter from Mr Walter Hunter, Committee co-ordinator, 26 January 2006.

.

TABLE OF CONTENTS

LIST OF BOXES

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADG	Audit Development Group
ANOVA	Analysis of Variance
CHP	Community Health Partnership
CI	Confidence Interval
CIT	Critical Incident Technique
CVI	Content Validity Index
CPD	Continuing Professional Development
DoH	Department of Health (England and Wales)
FRCGP	Fellow of the Royal College of General Practitioners
JCPTGP	Joint Committee of Postgraduate Training for General Practice
GMC	General Medical Council
nGMS contract	new General Medical Services contract
GP	General Practitioner
GGPCT	Greater Glasgow Primary Care Trust
GG&CHB	Greater Glasgow and Clyde Health Board
LHCC	Local Health Care Co-operative
MRCGP	Member of the Royal College of General Practitioners
NES	NHS Education for Scotland
NHS	National Health Service
NHS QIS	National Health Service Quality Improvement Scotland
NPSA	National Patient Safety agency
PCO	Primary Care Organisation
PCT	Primary Care Trust
PGEA	Post Graduate Education Allowance
PMETB	Post Graduate Medical Education and Training Board
QOF	Quality and Outcomes Framework
RCGP	Royal College of General Practitioners
RCA	Root Cause Analysis
SEA	Significant Event Analysis/Audit
UK	United Kingdom
WHO	World Health Organisation
woS	west of Scotland

LIST OF PUBLICATIONS

The following publications are based on data reported in this thesis:

PEER REVIEWED ORIGINAL RESEARCH PAPERS:

1. McKay, J., Bradley, N., Lough, M &. Bowie, P. (2009)
   A review of significant events analysed in general practice: implications for the quality and safety of patient care,
   BMC Family Practice, 10:61

2. McKay, J., Pope, L., Bowie, and P. & Lough, J.R.M. (2009)
   External feedback in general practice: A focus group study of trained peer reviewers of significant event analyses,
   Journal of Evaluation in Clinical Practice, **15**, pp.142-147.

3. McKay, J., Shepherd, A., Bowie, P. & Lough, M. (2008)
   Acceptability and educational impact of a peer feedback model for significant event analysis,
   Medical Education **42**, pp.1210-1217.

4. McKay, J., Bowie, P., Murray, L. & Lough, M. (2008)
   Levels of agreement on the grading, analysis and reporting of significant events by general practitioners: a cross sectional study,
   Quality and Safety in Health Care **17**, pp.339-345.

5. Bowie, P., McKay, J., Murray, L. & Lough M. (2008)
   Judging the quality of clinical audit by general practitioners: A pilot study comparing the educational assessments of medical peers and NHS audit specialists,
   Journal of Evaluation in Clinical Practice, **14**, pp.1038-1043.

6. McKay, J., Murphy, D., Bowie, P., Schmuck, M-L., Lough, M. & Eva, K.W. (2007)
   Development and testing of an assessment instrument for the formative peer review of significant event analyses,
   Quality and Safety in Health Care **16**, pp.150-3.

7. McKay, J., Bowie P, Murray, L. & Lough, M. (2004)
   Attitudes to the identification and reporting of significant events in general practice,
   Clinical Governance: An international Journal, **9**, pp.96-100

SUMMARY

General practitioners (GPs) and their teams in the United Kingdom (UK) are encouraged to identify and analyse significant health care events. Additionally, there is an expectation that specific significant events should be notified to reporting and learning systems where these exist. Policy initiatives – such as clinical governance, GP appraisal and the new General Medical Services (nGMS) contract - attempt to ensure that significant event analysis (SEA) is a frequent educational activity for GP teams. The presumption from policymakers and healthcare authorities is that GP teams are demonstrating a commitment to reflect on, learn from and resolve issues which impact on the quality and safety of patient care. However, there is minimal evidence to support these assumptions while there is no uniform mechanism to ensure consistency in the quality assurance of SEA reports.

One potential method of enhancing both the learning from and the quality of SEA is through peer review. In the west of Scotland an educational model to facilitate the peer review of SEA reports has existed since 1998. However, knowledge and understanding of the role and impact of this process are limited. With the potential of peer review of SEA to contribute to GP appraisal and the nGMS contract, there was a need to develop a more evidence-based approach to the peer review of SEA.

The main aims of this thesis therefore are:

- To identify and explore the issues involved if the identification, analysis and reporting of significant events are to be associated with quality improvement in general practice.

- To investigate whether a peer feedback model can enhance the value of SEA so that its potential as a reflective learning technique can be maximised within the current educational and contractual requirements for GPs.

To achieve these aims a series of mixed-methods research studies was undertaken:

To examine attitudes to the identification and reporting of significant events a postal questionnaire survey of 617 GP principals in NHS Greater Glasgow was undertaken. Of the 466 (76%) individuals who responded, 81 (18%) agreed that the reporting of such events should be mandatory while 317 (73%) indicated that they would be selective in what they notified to a potential reporting system. Any system was likely to be limited by a difficulty for many GPs (41%) in determining when an event was 'significant.'

To examine levels of agreement on the grading, analysis and reporting of standardised significant events scenarios between different west of Scotland GP groups (e.g. GP appraisers, GP registrar trainers, SEA peer reviewers) a further postal questionnaire survey was conducted. 122 GPs (77%) responded. No difference was found between the groups in the grading severity of significant events scenarios (range of p values = 0.30-0.79). Increased grading severity was linked to the willingness of each group to analyse and report that event ($p<0.05$). The strong levels of agreement suggest that GPs can prioritise relevant significant events for formal analysis and reporting.

To identify the range of patient safety issues addressed, learning needs raised and actions taken by GP teams, a sample of 191 SEA reports submitted to the west of Scotland peer review model were subjected to content analysis. 48 (25%) described incidents in which patients were harmed. A further 109 reports (57%) outlined

circumstances which had the potential to cause patient harm. Learning opportunities were identified in 182 reports (95%) but were often non-specific professional issues such as general diagnosis and management of patients or communication issues within the practice team. 154 (80%) described actions taken to improve practice systems or professional behaviour. Overall, the study provided some proxy evidence of the potential of SEA to improve healthcare quality and safety.

To improve the quality of SEA peer review a more detailed instrument was developed and tested for aspects of its validity and reliability. Content validity was quantified by application of a content validity index and was demonstrated, with at least 8 out of 10 experts endorsing all 10 items of the proposed instrument. Reliability testing involved numerical marking exercises of 20 SEA reports by 20 trained SEA peer reviewers. Generalisability (G) theory was used to investigate the ability of the instrument to discriminate among SEA reports. The overall instrument G co-efficient was moderate to good (G=0.73), indicating that it can provide consistent information on the standard achieved by individual reports. There was moderate inter-rater reliability (G=0.64) when four raters were used to judge SEA quality. After further training of reviewers, inter-rater reliability improved to G>0.8, with a decision study indicating that two reviewers analysing the same report would give the model sufficient reliability for the purposes of formative assessment.

In a pilot study to examine the potential of NHS clinical audit specialists to give feedback on SEA reports using the updated review instrument, a comparison of the numerical grading given to reports by this group and established peer reviewers was undertaken. Both groups gave similar feedback scores when judging the reports (p=0.14), implying that audit specialists could potentially support this system.

To investigate the acceptability and educational impact associated with a peer reviewed SEA report, semi-structured interviews were undertaken with nine GPs who had participated in the model. The findings suggested that external peer feedback is acceptable to participants and enhanced the appraisal process. This feedback resulted in the imparting of technical knowledge on how to analyse significant events. Suggestions to enhance the educational gain from the process were given, such as prompting reviewers to offer advice on how they would address the specific significant event described. There was disagreement over whether this type of feedback could or should be used as supporting evidence of the quality of doctors' work to educational and regulatory authorities.

In a focus group study to explore the experiences of GP peer reviewers it was found that acting as a reviewer was perceived to be an important professional duty. Consensus on the value of feedback in improving SEA attempts by colleagues was apparent but there was disagreement and discomfort about making a "satisfactory" or an "unsatisfactory" judgement. Some concern was expressed about professional and legal obligations to colleagues and to patients seriously harmed as a result of significant events. Regular training of peer reviewers was thought to be integral to maintaining their skills.

The findings presented contribute to the limited evidence on the analysis and reporting of significant events in UK general practice. Additionally, aspects of the utility of the peer review model outlined were investigated and support its potential to enhance the application of SEA. The issues identified and the interpretation of findings could inform GPs, professional bodies and healthcare organisations of some of the strengths and limitations of SEA and the aligned educational peer review model.

CHAPTER 1

Defining and classifying significant events and their potential role(s) in incident reporting.

1.1 Defining a significant event

Defining a significant healthcare event can be a difficult task. Within the patient safety and risk management literature a range of terms such as patient safety incidents, adverse events, errors, critical incidents, near misses, and significant events are often used arbitrarily and interchangeably by healthcare professionals and quality improvement specialist staff. This is further complicated by there being no agreed international definitions of each term that can be applied uniformly to each phrase, although some attempt at agreeing commonly used terms have been proposed (World Health Organisation, 2005). Indeed, the identification and analysis of 'significant events' is mainly a UK general practice activity with the terminology rarely reported in international research literature or healthcare settings (Pringle, 2007; Bowie *et al*., 2008).

In the United Kingdom (UK) a 'significant event' is effectively an umbrella term for the various descriptions of quality of care issues including patient safety events that may occur. A possible relationship between these terms is described in Figure 1 and examples of definitions that are applied to the various terms are given in Box 1 and Box 2. It should also be recognised that unlike much of the terminology applied in the patient safety literature, significant events can also relate to aspects of healthcare where good practice has been identified (Pringle *et al*., 1995; Sweeney *et al*., 2000). In addition, the term is also used to describe difficult or interesting events that can involve complex patient care issues. These events are not necessarily classified as involving patient safety issues but often have emotional or educational resonance for the healthcare

provider(s) (Bowie *et al*., 2005a). However, from an analysis and incident reporting viewpoint it is likely that it is errors, adverse events and near misses which are potentially of most interest, particularly to GP teams as well as local and national NHS authorities.

Given the various terminologies, it is not surprising that it is problematic to determine what constitutes a 'significant event'. It is obvious that what one member of a health care team considers 'significant' in terms of patient care in its widest sense may not always match that of a colleague in the same practice or elsewhere. Understanding of 'significance' may vary both in the breadth of issues thought to be relevant to health care and in the threshold at which an occurrence is deemed severe or interesting enough to be considered 'significant'. Given the above, it would be extremely difficult for healthcare professionals to reach agreement on what is 'significant' in every single circumstance.

Despite this, it would be helpful for primary care teams to at least have guidance on what constitutes a significant event which may be worthy of a formal analysis and/or officially reporting to a relevant healthcare organisation. This could contribute to the body of knowledge about error and harm in primary care so that others can learn from these experiences.

Pringle *et al*. (1995), when advising on what events should be considered for analysis, suggested that, "*any event thought by anyone in the team to be significant in the care of patients or the conduct of the practice should be open for consideration*". This definition gives the health care team a broad range of possibilities to consider as significant enough to be worthy of analysis. The authors gave advice on a 'core list' of marker

events that included examples from preventative care, acute care, chronic disease management and organisational issues that would lend themselves to analysis. This served to demonstrate the breadth of topics potentially available for analysis, particularly when contrasted with the perception of a limited range of subjects which could be tackled by more conventional (quantitative) methods of audit. Such guidance aimed to prompt practices to examine their choice of subject so that common events with very similar outcomes did not dominate the choice that would go on to be discussed and analysed by the practice.

Harrison *et al*. (2002), in a report for the Department of Health (DoH) for England and Wales, examined the potential for significant events to inform incident reporting in UK general practice. It was highlighted that the term 'significant event' was used inconsistently in primary care, whereby it was often interpreted differently between those individuals within the general practice and those who were in local managerial or regional healthcare organisational roles. Jacobsen *et al*. (2003), in their discussion on issues arising for patient safety in primary care, recognised a potential role for significant event analysis but pointed out that, *"it is not clear how significance is bestowed, and by whom, on some events and not others*". The authors raised this issue to emphasise the difficulty in deciding which events would provide the best 'trade off' for the practice in addressing patient safety concerns, since reporting or analysing a significant event comes at an opportunity cost to other significant events and quality improvement activities .

This definition of a significant event was further updated by Pringle in 2007, when he described significant events as, "a *patient safety incident (including therefore adverse healthcare events and near misses), an expected adverse outcome, a patient defined*

outcome (such as a complaint), or a circumstance that can be used to reflect on the quality of care" (Pringle, 2007). This definition had taken into account the role for SEA in the new general medical services contract (nGMS) (DoH, 2004) and in the patient safety agenda being promoted by the National Patient Safety Agency (NPSA) - that an SEA should be performed if a patient safety incident has resulted in a 'near miss' or low to moderate levels of patient harm (NPSA, 2005).

For the purpose of this thesis, the definition of a significant event has been taken as:

"…any event thought by any member of the practice team to be significant in the care of patients or the conduct of the practice" (Pringle *et al*., 1995).

This was chosen at the commencement of this thesis since the definition was the only one available in 2002 that had been applied uniformly in a study of significant events and their analyses in general practice. It also allowed for a broad range of patient care, practice educational requirements and practice governance issues to be addressed. Although in theory this could encourage 'frivolous' significant events, such as the kettle not being boiled for morning coffee (anon, personal communication), as well as worthy topics for analysis, this was thought to be acceptable given the limited knowledge on significant events and their analyses amongst GPs and their teams.

1.2 The scale and nature of significant events

When reviewing the literature on significant events it is necessary at the outset to acknowledge that very few studies examine 'significant events' per se, but focus on sub-categories such as errors, complaints or adverse events. Although the understanding of patient safety issues has been limited by the different definitions used by researchers (Wilson *et al*., 2001; Sanders and Esmail, 2003), the identification and classification of the various sub-categories of significant events could inform

improvement measures. This is important since in a review of medical error studies in primary care and medico-legal databases, Sanders and Esmail (2003) put a figure on primary healthcare error of up to 1 in 120 consultations. The authors highlighted the wide variation in error rates given in different studies with incident reports, studies relating to taxonomies and medical defence databases being difficult to amalgamate because of variations in definitions and methodologies of data collection. Of note was the estimation that between 63% and 80% of errors were thought to be preventable. The authors did not describe how their estimate of error rate was obtained, but if the figure was accurate, then given that an estimated 260 million consultations take place annually in the UK (NHS Information Centre, 2008), up to 2 million errors may be made. However, the number causing serious harm is likely to be considerably less.

An adverse event rate 3.7 per 100,000 consultations was estimated in a retrospective review of incident reports in a risk management database in the Midwestern US (Fisher *et al*., 1997). Although it was acknowledged that this figure was likely to be a significant underestimate, similar to the above study, the authors reported that 83% of events were thought to be preventable.

Hider *et al*. (2005), as part of the New Zealand Quality of Healthcare Study (NZQHS), used case note reviews of 575 records from 13 public hospitals. The authors found that 2.6% of admissions were associated with community-based adverse events. This accounted for 20% of all adverse events in the study. Although representing only the authors' opinions, incidents originating in the community were described as significantly more likely to be 'highly preventable' than in-patient events.

Very few studies present any form of taxonomy of significant events in general practice. Much of the published literature on significant events has related to aspects of their analyses rather than describing any classification of the events themselves. Other potential sources of classification of significant events include medico-legal databases. These contain information on primary care significant events but are limited in informing taxonomies of patient safety incidents because of their selective nature (Esmail *et al*., 2004**).**

An early study to examine the range of 'significant events' in general practice was Pringle *et al*.'s 1995 paper on the feasibility of SEA. The authors identified 177 'clinical' significant events that the participating practices went on to analyse and 109 'administrative' events. This was not an exhaustive list of significant events occurring in practice since some events were identified but not analysed. There were 13 different categories of clinically significant events identified, with the most common areas chosen for analysis relating to chronic diseases, sudden deaths and cerebro-vascular disease and cancers. Of the 22 administrative categories identified, patient complaints, comments and requests were most likely to lead to analysis.

A second UK study to classify significant events selected for analysis was a review of SEA reports submitted to a primary care trust (PCT) in England. These reports were submitted as part of the Quality and Outcomes Framework (QOF) of the nGMS contract. As well as describing the types of significant events analysed, the authors investigated the relationship between the events and their potential effect on patient safety. Practice-based events with an administrative focus were the most frequent types of significant event chosen for analysis (28.2%), while medicines management (13.6%) and new cancer diagnoses (9.5%) were the most common subject choice for analysis of

clinical problems (28.5%). Over a quarter of significant events were classified as patient safety incidents (26.7%) with a further quarter of these representing 'serious' or' life threatening' events (Cox and Holden, 2008).

Several of the early studies in general practice on 'patients safety incidents' were carried out in Australian general practice. As part of a study of incident monitoring carried out between 1993 and 1995 volunteer GP participants completed incident report forms. These gave details on the type of event, potential and immediate outcomes and contributing factors (Britt *et al*., 1997). In a subset study of the data on 'potentially harmful' events in general practice, Bhasale (1998) examined reporting forms from 219 GPs. Of these incidents, 76% were thought to be preventable. Using a mix of the quantitative and qualitative data provided within the reports, the authors categorised potential adverse events into four areas: drug management, non drug management, diagnosis and equipment. They also examined the reported cause(s) of events and were able to classify errors of clinical judgment as by far the largest contributing factor. Communication issues and the actions of others were the other major contributing factors.

Makeham *et al*. (2002) undertook a pilot study that aimed to develop an international taxonomy to describe errors in general practice: the Primary Care International Study of Medical Errors (PCISME). This was a study of participant self-recognised medical errors (defined as a threat to patient safety that should not happen again). The study involved participants from six countries: Australia, Canada, UK, Netherlands New Zealand and the US. Using a non random sample of GPs, errors identified by the participants were reported via an electronic questionnaire. Although the majority of incidents were reported by Australian GPs, the study results implied that errors were

similar in countries with similar healthcare systems. All categories of error that occurred in Australia occurred in one or more of the other countries. In contrast to Bhasale's study, the majority of errors could be split into a first level classification of 'process errors' (80%) and 'knowledge and skills errors (20%). The data for the Australian arm of the study were very similar to those found in the United States using the same classification system.

Makeham *et al*. (2008) proposed a three level taxonomy for patient safety events reported in general practice. This study was based on pilot work in the United States (Dovey *et al*., 2002) and involved participation of a representative sample of Australian GPs in the state of New South Wales. The first level involved errors relating to the process of care (69.5%) or knowledge and skills of health professionals (30.5%). A second level of seven themes such as communication and investigations was added while a third level comprised 35 descriptors for the various themes. The advantage of this classification was the relatively strong inter-coder agreement for classification (k=0.66). This was a substantial improvement on the levels from the earlier pilots of the study (k =0.37) and would imply that training is required to improve consistency in categorising safety events. If agreement on definitions amongst researchers is only moderate, then it is probably a reasonable assumption that advice will be required for GPs and their teams if they are to be involved in coding or grading the various forms of significant events for incident reporting systems.

Indeed, Kostopoulou (2006) has criticised primary care error classifications as lacking in consistency and theoretical construct. Most taxonomies were based on small samples of volunteers and grouped descriptive and causative factors into several different

domains or themes. Categories were neither mutually exclusive nor comprehensive. The author proposed that in common with systems in place in the airline, chemical and nuclear industries, categorisation should include psychological cognitive factors such as memory errors and perception failures. Although one such system exists for medical error, its use in general practice remains unreported (Zhang *et al*., 2004).

Despite the preliminary taxonomies on the type of patient safety incidents and descriptions of underlying reasons as to why incidents happen, what remains unknown is how these could be usefully disseminated to GPs and their staff. This is important if learning and subsequent improvements in patient care as a result of events are to be applied (Rubin *et al*., 2003).

1.3 The reporting of patient safety incidents in primary care.

1.3.1 *Background*

One of the overarching themes of patient safety is that of an organisation's 'culture.' Kirk *et al*. (2007) described it as "*the shared attitudes, beliefs, values and assumptions that underlie how people perceive and act on safety issues in their organisation.*" Safety cultures have been described as ranging from the pathological (where the organisation is dismissive of safety issues) to generative (where addressing safety issues is engrained within the system) (Hudson, 2003). This culture has been described as having four crucial components (DoH, 2000):

- *a reporting culture - where all staff are prepared to report patient safety incidents*
- *a 'just' culture - that engenders an atmosphere of trust amongst colleagues*

- *a flexible culture – where appropriate staff can make decisions at the front line*
- *a learning culture – where lessons in patient safety are identified and actioned.*

The importance of developing reporting and learning cultures in particular is highlighted by findings that individual health care professionals and their organisations can fail to notify one another when adverse or potentially adverse events occur. In addition, some fail to disseminate lessons learned as a result of incident investigations (WHO, 2005). As a result, failures in care continue to be repeated with resultant patient harm.

One potential method of reducing patient harm is therefore through encouraging reporting. Although healthcare cannot be directly compared with other safety-critical industries, such as aviation, nuclear power and petrochemicals, it has been seen as lagging behind these fields in establishing effective safety reporting systems (Billings 1998; Barach and Small, 2000). The aviation industry, for example, has methods in place to detect, analyse and implement changes from safety incidents and near misses. The system also proactively identifies issues that have the potential to be safety problems unless addressed. Near miss incident reporting in aviation is seen as complementary to the systematic data collection of harm since the risk of harm has already been established through databases of accidents (Vincent, 2007). Such reports by themselves have been described as informing only weakly the causes and prevention of aviation safety incidents (Cook *et al*., 1998). These systems have subsequently moved away from the analysis of low frequency, high consequence events to more frequent near miss events. Emphasis is placed on describing how operators cooperate to detect and mitigate potential or actual adverse event (Johnson and Holloway, 2007).

Most reporting and learning systems at both local and national levels in healthcare in the UK have been described as being at an early stage in both their development and their effectiveness in improving patient care (NPSA, 2005). However, since 1964 a national “yellow card” system has been in operation that allows all medical doctors, dentists and pharmacists in the UK public and private sectors to report suspected adverse drug reactions in patients. In addition, confidential inquiries into deaths originated in the 1950s, when medical professionals established a system to share and learn lessons from maternal deaths: the Confidential Enquiry into Maternal Deaths. In 1992 a similar mechanism was put in place to establish the Confidential Enquiry into Stillbirths and Deaths in Infancy. These two enquiries were combined as the Confidential Enquiry into Maternal and Child Health in 2003 (www.cemach.org.uk). Two other national confidential enquiries into patient Outcomes and Deaths (www.ncepod.org.uk) and the National Inquiry into Suicide and homicide (www.medicine.manchester.org) exist. These and the national system for incidents involving medical devices received praise from the DoH in England as being examples where incident reporting could lead to improvement in patient safety (DoH, 2000).

The NPSA described six potential benefits of having a national reporting system:

- *To identify areas for change and improvement in patient care and safety and thus to provide evidence to target resources*
- *Timely reporting increases responsiveness particularly when undertaking investigations*
- *In primary care, patient safety incidents originating in secondary care can be identified and thus benefit communication at the interface between the two.*
- *Organisations can prepare for potential complaints and litigation cases.*

- *Financial benefits arise from reduced severity of incidents e.g. reduced lengths of stay*
- *The potential for national solutions to national problems such as equipment manufacturers and suppliers.*

The WHO stated that incident reporting systems, whether at local, regional or national level in a patient safety context, should be inextricably linked to learning (WHO, 2005). This enables any lessons highlighted to be communicated to other healthcare providers and, where appropriate, systems can be put in place to reduce the chance of similar events recurring (DoH, 2000).

These potential benefits of incident reporting and learning systems may be difficult to deliver. Vincent (2004) commented on the international proliferation of reporting systems that are at the centre of many initiatives to improve healthcare and cautioned that incident reports themselves tell us little about causes and prevention. Billings (1998) was even more critical, describing counting incident reports as a waste of time; while Renshaw *et al.* (2008) wrote that "*there is an absence of empirical evidence that incident reporting systems in health care can improve patient safety.*" These three authors contend that the under-reporting of patient safety incidents and subsequent loss of learning opportunity transfers the emphasis from classification and counting to the 'real meaning of incidents' that lie in the narrative.

1.3.2 *Policy initiatives to develop and support reporting and learning systems*

Despite the misgivings on the value of incident reporting systems described above, over the past two decades there has been increasing pressure from the public, politicians and the healthcare professions to develop and implement such systems. The

publication of the United States (US) Institutes of Medicine's report 'To err is human' (Kohn *et al.*, 1999) received widespread publicity (Vincent, 2006). The public reaction to the declaration that between 44,000 and 98,000 people died in US hospitals annually as a result of medical errors ensured that both politicians and healthcare professionals felt compelled to act on the issue of patient safety and improve outcomes (Leape, 2000; Leape and Berwick, 2000). A series of recommendations were made that included the creation of a Centre for Patient Safety and also a nationwide reporting system in the US. This had international repercussions with professionally led incident monitoring and reporting systems being introduced or augmented in Australia, Canada and New Zealand (Beckmann *et al.*, 1996; WHO, 2005). In addition, governmental agencies in countries such as Denmark and Sweden established reporting systems while national reporting systems based on specialties, most often including anaesthesia, were also implemented (Wu, 2002).

In 1999 the UK government's response to increasing concern around patient safety was to commission an expert group, chaired by the chief medical officer for England, to prepare a report on learning from adverse events in the NHS. This report was titled 'An organisation with a memory' (DoH, 2000).

This report criticised the NHS for lacking a standardised national reporting system and indeed a standard definition of what should be reported. A similar reprimand was directed towards local NHS organisations, which were viewed as the 'bedrock' upon which reporting mechanisms would be built. Specifically, a cultural problem where there was a failure to identify and learn from near misses in the service was highlighted. Systems were described as 'almost non-existent.' In particular, the primary care sector was highlighted as being 'poorly developed' in relation to incident reporting. The report

recognised that healthcare was prone – in some cases at least – to similar pressures and human errors as the other high risk industries mentioned previously. The report placed a strong emphasis on how these organisations learned from error and harm as well as the requirement for a shift away from the traditional 'blame' culture.

A further DoH report - 'Building a safer NHS' (DoH, 2001a) - described the steps that would be required to develop a national system for learning from adverse events. It would result in the establishment of the previously described NPSA. More detail on the methods that would be applied by the new NPSA in collecting and analysing patient safety data were given in the report 'Doing Less Harm' (DoH, 2001b). The report emphasised a duty on independent contractors such as GPs to report adverse incidents to their local primary care trust (PCT).

1.3.3 *Engaging general practice and GPs in reporting and learning systems.*

In primary care in England and Wales, engagement with the NPSA reporting system is significantly less than would be expected for the number of patient contacts occurring each year (NPSA, 2008). This lack of engagement in incident reporting by general practitioners has also been criticised by a parliamentary committee (House of Commons, 2006). In addition, although GPs may be encouraged to undertake reporting of particular significant events as part of the clinical governance requirements for local healthcare organisations, there is little evidence as to the degree of participation.

The reasons for poor engagement with incident reporting systems may be more complex than simple 'inertia' by GPs. A major issue for the application of both national and local 'generic' reporting systems is highlighted in several papers that have discussed the calls to action in patient safety in primary care (Pringle, 2001; Wilson *et*

al. 2001; Jacobsen *et al*., 2003). The authors describe what they perceive to be the unique nature of general practice compared to other areas of health care. They imply that safety issues and approaches to the identification and understanding of error established in other health care sectors cannot necessarily be superimposed. General practice involves many decisions concerned with managing uncertainty and marginalising risk.

Additionally, primary care has been recognised as a particularly difficult field since adverse events and some forms of patient harm such as drug side-effects are so frequent as to be considered as unremarkable by the doctor and not thought of as reportable events (Elder *et al*., 2004). Furthermore, in a study of the error rate associated with prescribing in general practice, Neville *et al*. (1989) found that although 3.17% of prescriptions were found to contain errors, none were judged to have serious consequences. Although it was acknowledged that these events constituted 'nuisance' value, it was not thought likely that these would result in significant learning for others.

It is known that in secondary care contexts individuals are reticent to report errors that reflect badly either on themselves or their team (Parker and Lawton, 2003; Kroll *et al*., 2008). If this were to be confirmed in primary care, then mandatory reporting schemes may be one method of addressing this problem. They may be useful in holding healthcare providers accountable, particularly for errors that cause serious patient harm. The downside is that they are expensive, difficult to monitor and learning opportunities are missed where errors have not led to harm. They also shift the emphasis from learning to accountability (Vincent, 2004; Vincent, 2006). Voluntary systems may be more likely to bring patient safety incidents to light as they do not necessarily involve harm and include more information on 'latent' harm. Indeed, some

commentators argue that all systems are ultimately voluntary systems since people can find reasons for not reporting, such as lack of time and personal decisions on whether the incident falls within the reporting remit. They also depend on the attitude of the organisation towards reporting (Billings, 1998; Barach and Small, 2000; Vincent, 2006). Confidentiality is common to nearly all reporting systems and the easiest way of ensuring this is through reports being filed anonymously. However, anonymity can be a hindrance to learning as analysts cannot contact the reporters for further information and this may lead to inaccurate or invalid reports. In particular, reports that are limited to only factual detail may miss substantial information from the narrative. Thus, significant detail of the human factors involved in incidents and behavioural changes required to address causal issues may be omitted (Barach and Small, 2000).

1.4 Summary

With the emergence of SEA from a relatively infrequent activity for GPs at the end of the 20th century to one that had the potential to become embedded in practice in the first decade of the new millennium, the possibility of using SEA as a substrate for both local and national reporting and learning systems grew. If such systems are to be successful in general practice, then there is a need to understand further the process and cultural issues that lead to either participation or a reticence to engage with reporting agencies (Sheik and Hurwitz, 2001). This thesis therefore seeks to provide some insight into the issues that would be involved if a link were to be established between SEA and reporting systems.

Box 1: Definitions and examples of terms used in patient safety literature

Patient safety incident: any unintended or unexpected incident that could have led or did lead to harm for one or more persons receiving healthcare (NPSA, 2005)

Adverse Event: An event or omission arising during heath care and causing physical or psychological injury to a patient (DoH, 2000, WHO, 2005). Healthcare includes all aspects of care such as diagnosis and treatment and the systems and equipment used to deliver care.

Example
A patient who was known to be severely allergic to penicillin was given amoxicillin, and developed an anaphylactic reaction requiring hospital admission.

Health care near miss
Any patient safety incident that had the potential to cause harm but failed to do so whether or not as the result of compensating action (DoH, 2000, NPSA, 2005)

Example
A patient was prescribed atenolol for hypertension, but the GP failed to notice the diagnosis of a history of asthma in the patient's notes. The pharmacist noted a previous prescription for salbutamol inhaler on the pharmacy records, phoned the GP and the patient was given an alternative antihypertensive.

Error
The failure to complete a planned action as intended, or the use of an incorrect plan of action to achieve a given aim (DoH, 2000)

a) Execution error where a plan of action is appropriate but the intended action does not take place usually due to a slip or lapse (Reason, 2001)

Example
Two health visitors check with one another that a DTP vaccine about to be administered to a child is correct and in date. Both fail to notice that the vaccine is not a DTP vaccine but an MMR vaccine and administer this to the child.

b) Planning error: the actions go exactly as planned, but the plan is insufficiently robust to achieve the intended outcome (Reason, 2001).

Example
A GP practice instituted a DMARD blood monitoring protocol for patients on methotrexate. This worked well for those on a stable dose but did not include advice that patients who had their dose altered should have had weekly blood tests for the following month. As a result, one patient with a low neutrophil count was only identified when they attended the surgery unwell a fortnight after the dose increase.

Box 2: Definition of a significant event and examples of different types of event

Significant Event: "*Any event thought by anyone in the team to be significant in the care of patients or conduct of the practice*" (Pringle *et al*., 1995).

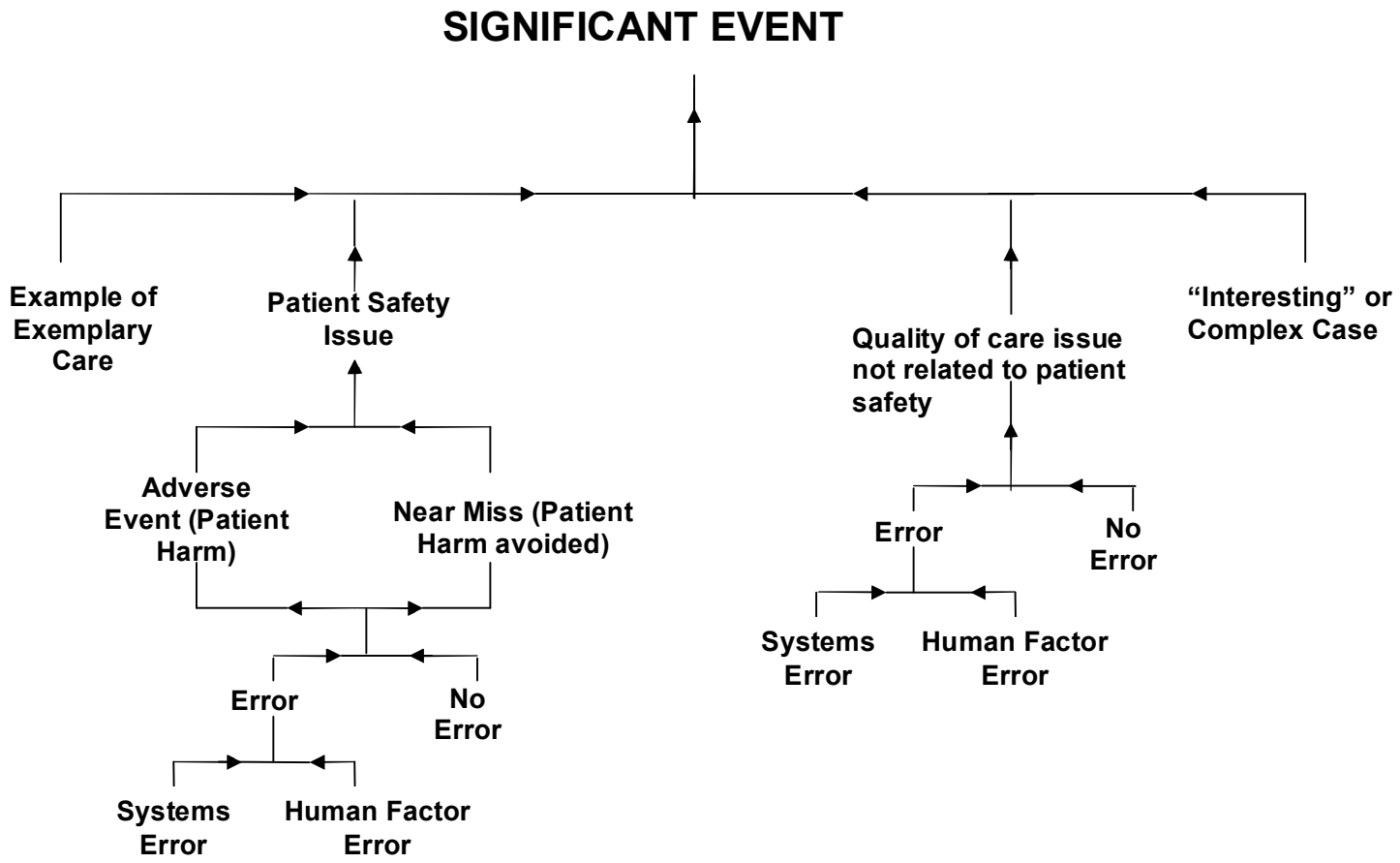
Example of exemplary care: A 75 year old patient with altered bowel habit and rectal bleeding was referred urgently for colonoscopy. This was performed the following week and the surgeon found a sigmoid tumour that was resectable. The patient underwent operation and had the tumour completely excised.

Example of a 'reflective' case: A patient with a terminal illness was being cared for at home by the GP. The patient was undecided as to the level of active healthcare intervention that they wished. This was complicated by different members of the immediate family asking the doctor and encouraging the patient to decide upon varying levels of intervention. The GP found the situation of providing the appropriate level of terminal care to the patient a difficult 'balancing act.'

Example of a patient safety incident: A GP prescribed the wrong dose of the antihypertensive drug doxazosin to an elderly patient. The starting dose was higher than recommended and the patient developed hypotension, fell and fractured their radius.

Example of an administrative incident: A patient booked a full life assurance medical at the GP surgery. The patient arrived for the appointment, however the medical forms had been misplaced and the patient was required to make a further appointment after the forms were eventually found four days later.

Figure 1: A possible relationship between significant events and terms in patient safety literature



CHAPTER 2

The analysis of significant events in general practice

2.1 The historical development of SEA

Pringle (2007) describes significant event analysis (or audit as it is commonly referred to in England and Wales) as "a *modern expression of an old idea*" – the review of single cases. He highlighted that the use of single case review has been practised in hospital medicine for several hundred years, for example, through grand ward rounds and post mortems. In general practice, established principals discussed single cases within small groups (Balint, 1975). Although mainly concerned with the interaction between a doctor and a patient, discussions allowed for lessons to be learned and potential improvements in care to be suggested. As part of the training of GPs who were not yet fully accredited to undertake independent general practice (GP registrars), established GPs would often review random or specific individual cases with the registrar. This was undertaken to identify educational issues for the doctor that could be applied in future practice.

During the 1980s and 1990s, the analysis of single cases or a series of similar cases was proposed as a method of audit to enhance and complement traditional cohort and criterion-based audit (Buckley, 1990; Donabedian, 1990; Irvine, 1983). These authors highlighted that using case-based discussions as a form of medical audit could shed light on areas of practice not illuminated by conventional criterion audit. One of the strengths of the analysis of individual cases (case-based audit) has been that it appeals to health professionals' human nature, being about people and situations rather than the impersonal quantitative data analysis of diseases and their management that was the subject of most criterion audits (Bradley, 1992).

Both Bradley (1992) and Buckley (1990) were explicit in stating that single case analysis as a form of audit was limited both in exploring and then generalising the issues raised to the whole population. A significant limitation described by Bradley was that doctors would often use single case "anecdotes" to support their own clinical decision-making but conversely would reject anecdotal evidence if it did not comply with their clinical belief. While recognising the potential value in the data contained in the analysis of single cases, he highlighted that this kind of evidence was unreliable since it depended on individuals' perceptions and interpretations of an event. This concern was expressed in more forceful terms by the Standing Committee on Postgraduate Medical Education (1989), who stated that case-based audit "*does not meet the requirements of medical audit."* In the committee's view, such 'audit' remained informal, unstructured and unproven as a method to drive learning and change.

To try and minimise the subjectivity applied to case-based discussion or review, Bradley suggested applying the principles of the critical incident technique (CIT) to these discussions. The CIT was developed from the work of the American psychologist JC Flannagan (1954). He described a factual approach to gathering data about the behaviour and experiences of individuals in defined situations in order to minimise the subjectivity of any study findings. The application of the CIT was reported in studies that involved the Aviation Psychology Program of the United States Army Air Forces in World War 2. Researchers set out to analyse *specific* reasons for recruits' failure in learning to fly. The technique went on to be employed in a variety of industries and trades, but maintained the same five core principles of analysis:

- *Specify the general aims*
- *Outline plans and specifications*
- *Collect data*

- *Analyse the data*
- *Interpret and report the data*

The application of CIT methodology to single cases or “anecdotes” that were “critical” to patient care (in terms of implications for clinical care) led Bradley to suggest that such “significant event analysis” may be a more successful way of changing behaviour than either single case analysis or conventional criterion audit.

2.2 The initial introduction of SEA into general practice.

The first influential study in UK general practice to explore the feasibility and acceptability of significant event analysis was carried out by Pringle *et al*. (1995). The authors recruited 20 multiple-partner practices in Lincolnshire and Manchester to represent both rural and urban practices. Single handed practices were excluded as the study aimed to examine aspects around the discussions and decisions over audit findings. The practices were randomly assigned to carry out audits over a one-year period using either the conventional quantitative cohort methodology or SEA. A significant finding was that the practices in the SEA arm of the study covered a larger and much broader range of topics with less time investment than those practices undertaking traditional audit. The authors concluded that SEA was both a feasible and acceptable option for UK general practice. They also introduced the concept of SEA as a quality improvement method that could enhance performance review. It was not a replacement for conventional audit but could complement it through examining areas of practice not amenable to cohort audit.

Although SEA was promoted as a voluntary, wide-ranging, less laborious technique than criterion audit, it remained the ‘junior partner’ in terms of audit uptake until at least

the late 1990s. In 1999, evidence given to the Select Committee on Health estimated that only 20% of GPs were participating in SEA, despite Pringle *et al*.'s early work in this area (RCGP, 1999). However, the evidence for this estimate was not described. This low figure for participation may have been partly because GPs were more familiar with traditional criterion audit and so participated in this rather than audits of single significant events.

However, in the decade following Pringle *et al*.'s feasibility paper, the requirement to participate in SEA increased. The Royal College of General Practitioners (RCGP) asked individual GPs to demonstrate participation in SEA as a requirement for membership by assessment of performance (MAP), and fellowship of the college if being undertaken via workplace assessment. A practice was expected to have a variety of SEAs available for external inspection if applying for either the Royal College Practice Accreditation Award or the Quality Practice Award (RCGP, 1983; RCGP, 2000; RCGP Scotland, 2002). GP training practices in the west of Scotland were required to demonstrate participation in SEA as part of their training accreditation. Established individual trainers were expected to have submitted SEA reports for external peer review (a judgement by trained GP assessors from outside the GP's own practice). Since 2003, all prospective GP trainers in NHS Scotland have had to demonstrate participation in SEA and again submit examples for peer review before they can apply for accreditation as a trainer (Shackles *et al*., 2007).

The above examples, although affecting a minority of GPs, were initial steps taken to encourage SEA as a 'mainstream' activity. However, the two major driving forces for the introduction of SEA into mainstream general practice over the past five years have

been the introduction of GP appraisal in Scotland in 2003 and the new UK general medical services (nGMS) contract, introduced in 2004.

2.3 The emerging evidence for the potential role(s) of SEA in general practice

It is not intended to detail all the studies that have reported on the evidence for the potential role(s) and effectiveness of SEA in general practice. For this the reader is directed to a review by Bowie *et al*., (2008). Although the authors reported some evidence for SEA in relation to improving team work, communication and team-based learning, discussed below, the evidence of its impact on healthcare was found to be severely limited. The reliability of the technique was also questioned because it lacked a consistent and structured approach.

Three areas of SEA that are examined in greater depth as part of this thesis are its potential as a team-based collective learning technique to improve the quality of health care, its possible contributory role as part of incident reporting systems and its application as a reflective learning technique for individual GPs,

2.3.1 *Improving the quality of patient care through improved team performance*

Pringle *et al*.'s feasibility study (1995) advanced SEA as "an additional tool for the enhancement of performance". Its perceived strength would be in identifying problems in the quality and delivery of healthcare. The finding of its effectiveness in team-based learning could extend audit from 'medical' to 'clinical' applications and also had the potential to encourage 'interface' audit between primary and secondary care.

Evidence for SEA in enhancing performance and teamwork started to increase from the mid 1990s. Bennett and Danczak (1994) demonstrated that SEA could be used as a

method to implement change in difficult palliative care cases. Evidence for its role as a multidisciplinary method to improve the general standards of care was reported in two studies (Robinson and Stacy, 1994; Robinson *et al*., 1995), while its strength as a means to enhance teamwork through increasing awareness of others' professional roles was confirmed by Robinson and Drinkwater (2000).

Sweeney *et al*. (2000) carried out a study with two of the aims being the identification of who raised significant events and the outcomes of the analyses. The study method adopted a combination of interviews with specific practice team members and observations of SEA meetings. The authors reported that the technique attracted a wide variety of primary care workers who all actively contributed, and that interpersonal relationships within the practice improved as a result of participation in the analysis. It was described as having a potent capacity to identify areas for quality improvement, with many analyses resulting in important (even if small) changes. They also reported that issues surrounding the selection of topics, leadership, confidentiality, and implementing and measuring outcomes were barriers to the success of the SEA technique.

One example of how a single SEA could motivate an individual general practice as an 'organisational unit' to improve the quality of care for larger groups of patients was described by Pringle (1998). He reported how the audit of a single significant event of a premature death from coronary heart disease, (that may have been previously accepted as an unfortunate but largely unavoidable death by practice team members), led to benefits for the wider practice population. Following the individual case review, the practice carried out a criterion audit on the recording of risk factors for IHD and the management of hypertension and lipid levels. With the exclusion of blood pressure

control all other indicators improved, although this was at a cost to the practice's commissioning and drug budget.

2.3.2 *SEA and incident reporting and learning systems*

The potential role that significant events could play in providing information to reporting systems has been discussed in Chapter 1. Two further studies in relation to SEA are also relevant. In 2002, Harrison *et al.*'s previously described report on the potential role of SEA in primary healthcare was commissioned in part by the government's intention to introduce a mandatory reporting system for adverse healthcare events and near misses and to amend terms of service to require GPs to monitor adverse or other significant events (DoH, 2001a). Their findings were based on a study that combined written and telephone surveys of NHS 'stakeholder' organisations, face-to-face interviews with members of primary healthcare teams and analysis of published or official documents on SEA. They proposed that SEA, as "*a systematic and flexible framework familiar to GPs*", could be aligned to team learning, individual professional development and play a role in national as well as local reporting and learning systems. A criticism of this study would be that the methods applied to collect and interpret the data were not explicit.

In a study to examine whether SEA could contribute to risk assessment through a reporting and learning system in a local health care co-operative (LHCC) in Lanarkshire, Murie and McGhee (2003) examined 56 written SEA reports from 10 GP practices. The authors used the clinical negligence and other risks (CNORIS)[1] classification to 'grade' significant events. This method was essentially the same as that

[1] CNORIS is a financial risk pooling and risk management system used by Scottish NHS boards to share good risk management practice and also to share costs of settlements made against individual health boards from compensation claims.

proposed at the time by the NPSA for the classification of patient safety incidents in England and Wales. An event was graded according to its consequences and the likelihood of it recurring. The study reported that 44% of SEA reports described moderate, major or catastrophic outcomes to the significant event. In over three quarters of events recurrence was thought 'possible', 'likely' or 'almost certain' unless the underlying cause of the event was addressed. The authors concluded that SEA was an appropriate method to support the incident reporting aspect of risk management but also highlighted discordance between the educational aspect of SEA where 'celebratory' reports were considered to be a 'false positive' within the CNORIS and NPSA's overall aim to identify incidents that could be used to improve patient care. Although very few matrixes for the severity grading of significant events were available at this time, the study was limited by adopting one which had been largely unused in primary care and lacked evidence of the consistency of its application.

The role of SEA in reporting was also likely to be compromised by the selectivity of events chosen for analysis (Bowie *et al*., 2005a) and the difficulty for some GPs in applying the technique with some degree of rigour (McKay *et al*., 2003).

2.3.3 *SEA as a reflective learning technique in general practice*.

For over eighty years reflective practice has been proposed as one method that professionals use to improve their practice (Dewey, 1933). In medical education, reflective practice has been conceived as the ability of doctors to think critically about their own reasoning and decisions (Mamede and Schmidt, 2004). Learning has been described as *"the process of transforming experiences into knowledge, skills, attitudes and values*" (Pitts, 2007).

Although participation in any form of reflective learning may not by itself lead to either change in professional behaviour or improved patient outcomes, undertaking an educational activity and demonstrating that learning has taken place are basic steps in addressing the quality of patient care (Kirkpatrick, 1967; Brigley *et al*., 1997).

SEA is commonly promoted as an ideal reflective learning technique both for individual practitioners and practice teams (Pringle *et al*., 2005; Henderson *et al*., 2002; Harrison *et al*., 2002; Sweeney *et al*., 2000). This is not surprising since models for reflective practice (Kolb, 1984; Pitts 2007) often ask the learner to apply very similar steps to those recommended for the analyses of significant events (Pringle *et al*., 1995; Lough, 2002). For Lough (2002), the interlinked processes of reflective practice and learning are core competencies in the application of SEA. The process of analysing a significant event should involve a demonstration by those undertaking the analysis of competence in their ability to learn from the event. Participants should be able to reflect on and demonstrate insight into the reasons for the event and suggest or implement change, if required, as a result of undertaking the analysis.

Given the wide variations in the ways in which medical practitioners approach and learn from problems, some doctors are more likely to engage in reflective learning than others (Grant, 2002). Doctors most often perform and practice in areas in which they are competent do so automatically (knowing in action). The reflective learning process, which can be either immediate (reflection in action) or delayed (reflection on action) is thought to occur most often in response to complex or new problems (Schon, 1983). In either case, a reflective practitioner has to relate their theoretical knowledge to a practical competence in order to demonstrate professional activity suitable for daily practice (Slotnek, 1996).

In a focus group study to explore why some GPs take part in SEA, Bowie *et al*. (2005a) presented evidence that analysing and writing up a significant event is an educational process for some GPs. These participants effectively described the writing up of an SEA report as 'reflection on action.' It involved key aspects of adult learning theory, such as self-direction, drawing on experience and being relevant to the participants' 'social role' (Knowles, 1980). The study also reported on the emotional aspects of undertaking SEA, which is an under-recognised but necessary process in effective reflective learning (Boud *et al*., 1985).

Participation in SEA by GP practices in one health care trust in England demonstrated learning across a broad range of clinical and administrative areas (Burroughs and Atkins, 2005). The learning points identified by a single practice were then shared throughout the trust, leading to further learning and change in a substantial number of practices (42%).

Learning is also associated with undertaking SEA in the areas of clinical care (Sweeney *et al*., 2000) and the roles of other members of the healthcare team (Robinson and Drinkwater, 2000; Bennett and Danczak, 1994). Team-learning was also thought to be enhanced by the immediacy of the SEA technique (Bowie *et al*., 2005). In a study by Cox and Holden (2008) in Merseyside, SEA undertaken for the nGMS contract demonstrated that key learning points were identified by over 80% of practices, although the individual learning points were not detailed.

In contrast to the above findings, learning opportunities may be missed through poor application of the SEA technique. In a study of 75 event analyses reports judged to be unsatisfactory by peer reviewers, Bowie *et al*. (2005b) reported a failure to identify or

describe appropriate learning issues in 12 (16%) of reports. This in turn led to a failure to implement change or resulted in the change described being judged as inadequate by the reviewers.

2.4 Policy initiatives promoting SEA in general practice.

2.4.1 *CPD, GP Appraisal and revalidation*

In the 1990s the public trust in the medical professions' ability to demonstrate effective regulation was severely damaged by several high profile media cases such as the Bristol Royal Infirmary hospital inquiry, the inquiries into Harold Shipman and the case of the gynaecologist Rodney Ledward. At this time the GMC was criticised for at least seeming to protect the interests of doctors ahead of those of patients. One of the changes proposed by the General Medical Council (GMC) to counter this was to introduce a regulatory measure - revalidation. Medical revalidation was described in an amendment to the 1983 medical act as '*the evaluation of a medical practitioner's fitness to practise*' (DoH, 2006).

The Chief Medical Officer for England's consultation paper on preventing, recognising and dealing with poor clinical performance of doctors, "*Supporting doctors, Protecting patients"* (DoH, 1999), proposed that appraisal should be made compulsory for those doctors who worked in the NHS. Of note was that the appraisal's primary aim was not seen as identifying poorly performing doctors but as a method to confirm and improve ongoing performance. It would also form a substantial core component of the evidence required by the GMC for revalidation. A thorough appraisal of a doctor's work by their 'employer' was also seen as a prerequisite for the responsible organisation to demonstrate its own statutory duty of quality through clinical governance (DoH, 1998; Scottish Office, 1998). This consultation paper also raised the prospect that the NHS

appraisal process would require input from an external peer review procedure both as part of the governance process for appraisal itself but also "*so that judgments can be made about how the local doctor's practice compares to his or her peers nationally as well as to best and excellent practice*". As part of the conclusions of the report, the CMO was of the opinion that "*the combination of clinical governance, revalidation and appraisal should result in many fewer cases of poor clinical performance*".

The Scottish GP appraisal scheme was instituted in 2003 (Scottish Executive, 2003) and became a contractual obligation for all UK practitioners with the introduction of nGMS in 2004 (DoH, 2004). In Scotland, amongst the principles underpinning appraisal were that it should be peer-based, supportive, formative and uniformly implemented throughout Scotland (Hunter *et al*., 2005). It was agreed between the various stakeholders involved in the development of the appraisal model that one of five core categories would be discussed in depth each year rather than having a superficial discussion of what were at that time the seven different areas of the GMC's publication "Good Medical Practice" (GMC 2001). The five categories were:

- Prescribing
- Audit
- Communication
- SEA
- Referrals and peer review.

In addition to these core categories, reflection on learning activities and complaints were thought over the five-year cycle to be consistent with Good Medical Practice. There was an expectation that undertaking this appraisal process would assist the

practitioner with revalidation, even if the final details of revalidation at this time remained unclear.

However, by 2004 the proposed role of appraisal within revalidation had been challenged. In her report into the enquiry into the case of the GP Harold Shipman, Dame Janet Smith criticised the proposed reliance on the annual appraisal of NHS doctors, judging it as not constituting a true evaluation of the full range of a doctor's performance and delivery of care and thus an ineffective method of detecting doctors who are "*incompetent, dysfunctional or delivering care to a poor standard*" (The Shipman Enquiry, 2004). By 2006 the CMO had agreed that revalidation as proposed by the GMC would fail to provide an objective evaluation if it was based largely on the existing model of NHS appraisal. The proposals for revalidation were effectively suspended (DoH, 2006a).

The report of the CMO for England's working group: *Medical revalidation – Principles and next steps* (DoH, 2008), while dealing primarily with revalidation, highlighted that a generic appraisal module was being developed by the GMC based on the framework of Good Medical Practice. This would include four key areas that could be appraised and objectively assessed:

- Knowledge, skills and performance
- Safety and quality
- Communication, partnership and teamwork
- Maintaining trust

For the present it is likely that SEA will remain a core category of the Scottish appraisal process (Cameron, personal communication, 2008) where its potential and application

in the patient safety, communication and teamwork categories of good medical practice can be assessed.

2.4.2 *SEA and the nGMS contract*

As stated previously, SEA was formally introduced as part of the Quality and Outcomes Framework (QOF) of the nGMS contract in 2004 (Scottish Executive, 2004). The decision to include SEA in the Education and Training section of the QOF highlighted its perceived role as an educational team-based exercise that demonstrated a practice's participation in audit. Practices were initially given financial reward for carrying out either six or 12 SEAs in the previous three years. In addition, those events appropriate for analysis were strongly encouraged to be from a 'prescribed' list that would include (if these had occurred) a patient complaint, one suicide, two new cancer diagnoses, two deaths where terminal care has taken place at home and one section under the Mental Health Act.

This requirement was altered slightly in the annual reviews of the QOF so that practices would still be rewarded for undertaking 12 SEAs in the previous three years but also for undertaking three SEAs in the past year rather than six SEAs in the previous three years (Scottish Executive, 2006). The 'prescribed' list of significant events for analysis has subsequently been amended so that the twelve SEA reports could include:

- Any death occurring in the practice premises
- New cancer diagnoses
- Deaths where terminal care has taken place at home
- Any suicides
- Admissions under the Mental Health Act

- Child protection cases
- Medication errors
- A significant event occurring when a patient may have been subjected to harm, had the circumstance/outcome been different (near miss).

Practices are encouraged to analyse their events in one of two structured ways (Pringle, 1995; Robinson *et al.*, 1995).

The only study so far to report on the value of undertaking SEA as part of the nGMS contract was that by Cox and Holden (2008), described earlier in this chapter, The authors carried out a retrospective review of 337 SEA reports undertaken as part of the nGMS contract by 35 GP practices in Merseyside during 2004/5. Using a categorisation recommended by the NPSA, over one quarter (26.7%) were classified as patient safety incidents. Given that the majority of practices (69%) were thought by the authors to have performed a successful analysis based on the two methods of analysis referenced above, there was an assumption that the SEA had addressed and reduced the risk to patients. However, the full value of SEA as part of the contractual duties of a GP practice remains to be fully explored.

2.4.3 *SEA and the NPSA*

As described in Chapter 1, the NPSA was established primarily to develop a national system for learning from adverse events in England and Wales. In 2005, the NPSA published *"Seven Steps to Patient Safety for Primary Care"* as a guide to good practice that acknowledged the different challenges that the primary care sector faced when compared to secondary care (NPSA, 2005). Step 6 of the guide related to the learning and sharing of safety lessons recommended that a SEA is undertaken for events or incidents which had the potential to cause harm but resulted in no harm to the patient

and those that resulted in low to moderate harm. This was thought to dovetail with practice frameworks for clinical governance and the requirements of the nGMS contract. The report described the technique as 'not usually involving major data collection' and hence feasible as a regular undertaking. Other methods of investigation, such as root cause analysis (RCA), were recommended as more suitable for cases of severe harm or death. This may reflect a viewpoint that SEA lacked the depth of understanding brought about by other techniques (NPSA, 2005), though whether this is indeed the case has not been researched.

2.3.4 *Clinical governance and SEA*

The term "clinical governance" first appeared in English and Scottish policy documents in 1998, which described reforms to the health service (DoH,1998; Scottish Office,1998). A frequently quoted definition of clinical governance is "*a framework through which NHS organisations are accountable for continually improving the quality of their services and safeguarding high standards of care by creating an environment in which excellence in clinical care will flourish*" (Scally and Donaldson, 1998). Elements associated with clinical governance were to include: clinical effectiveness, patient focus and risk management. Baker *et al*. (1999) recognised that many of the activities that could feed into the clinical governance agenda were already taking place in practice. Examples included: clinical audit, practice-based education programmes and complaints procedures.

While SEA predated the introduction of clinical governance, it was never explicitly confirmed as a component of clinical governance, although it was perhaps implicit through its role as a form of audit. In an editorial, Morrison (1999) described it as one method of audit that would contribute to the kind of evidence required for clinical

governance. Lough *et al*. (2002) advocated an integrated educational model for clinical governance that included SEA as a major component of the risk management domain. Stead *et al*. (2001) suggested that it could address what they proposed as the five cornerstones of clinical governance: system awareness; teamwork; communication; ownership and leadership. The authors cited their paper described in chapter 2.2.1 as evidence for this, while other studies also identified the technique as an ideal method of introducing the healthcare team to the concept of clinical governance (Robinson and Drinkwater, 2000). GPs themselves thought of it as one way of demonstrating participation in the clinical governance agenda through its perceived role as a risk management tool (Hillier, 2002). As described above, agencies such as the NPSA thought of it as a key component of clinical governance. The active engagement of general practice with clinical governance would, however, appear to be limited. Indeed, the parliamentary audit committee described primary care as paying only 'lip service' to clinical governance (Committee of Public Accounts, 2007). It may be appropriate to say that GPs essentially carry out their obligations to the clinical governance agenda through their contractual and Continuing Professional Development (CPD) duties.

2.4 Summary

The professional and policy initiatives reviewed in this chapter placed an expectation on SEA as a method by which GPs and their teams could demonstrate participation in different aspects of the quality and safety agenda for general practice. The relative paucity of the research on SEA meant that some of these potential roles were more of an assumption than an evidence-based application. Studies within this thesis therefore sought to supply evidence either for or against the SEA technique being an appropriate tool to fulfil the roles assigned to it within current professional and contractual duties.

CHAPTER 3

Peer review in general practice.

3.1 Background

Peer review in its broadest sense has been described as an evaluation of specific aspects of a person's performance by a peer group (Helfer, 1972). He described its modern origin in education as having developed from sociometry – a means of studying the attractions and repulsions of members of groups. This was developed further to allow *"the determination of perceptions of individual performance by members of a group*" (Cronbach, 1960).

Determining performance inevitably involves making some form of judgement. One of the first papers to describe the characteristics of a 'good judge' of another's performance was by Taft (1955), who described age, intelligence, social adjustment and interest in physical sciences as being beneficial characteristics, although these qualities have since been questioned (Helfer, 1972). However, Taft's assertion that the person "*must be motivated to make accurate judgments about his subject …and feel free to be objective*" is still appropriate.

Early studies on peer review in medical education involved students completing peer ratings on fellow students (Kubany, 1957; Liske *et al*., 1964). Although initially viewed as helpful only in research rather than assessment due to the limited correlation between student assessments and experienced assessors ratings, subsequent studies in the United States military identified that peers (internists) could reliably evaluate aspects of their colleagues' performance. This was particularly useful for assessing aspects of behaviour not readily accessible through existing instruments, such as inter-personal relationships (Caughey, 1967; Helfer, 1972).

Until recently, in general practice, peer review – if thought about at all - was arguably most commonly associated by general practitioners with the process by which manuscripts submitted to medical journals were assessed for suitability and scientific rigour for publication. Although Grol (1990) described peer review as "*undoubtedly a threatening prospect for most doctors*," peers are now asked to make judgments on many areas of their colleagues' medical practice (Norcini, 2003). This can range from overall judgements on a doctor's professional competence and performance to specific aspects of care in particular clinical areas (Southgate *et al*., 2001). Peer feedback is potentially well-placed to play a key role in informing educational processes in general practice and possibly also in governance standards.

3.2 Definitions of peer review, feedback and assessment.

Despite its increasing use in medicine, defining peer review from an educational perspective is problematic. The term 'peer' followed by 'review', 'assessment' and 'feedback' are used interchangeably in the literature. It is also complicated by the underlying intention of the review process and whether this is formal (e.g. awarding research grants) or informal (e.g. case conferences). Its purpose can also differ and range from a 'low-stakes' reflective discussion on a doctor's practice, to an external 'high stakes' evaluation of performance for re-licensure.

The lack of consistency in the use of terminology around peer review is perhaps not surprising. For example, there is no uniform definition of what constitutes a 'peer'. A dictionary definition of the term 'peer' is given as "*a person who is an equal in social standing, rank, age, etc.*" (Collins, 2000). In healthcare a 'peer' can refer to someone of equal or higher status (Gopee, 2001) or to "*a person of equal standing in terms of educational level and professional experience*" (Patterson, 1996). Callaham (2003)

describe peers as "*experts*" in the appropriate field, although the authors recognise that how 'expert' status is conferred can be inconsistent. Irvine and Irvine's (1991) definition of a peer as "*a person who is equal in any stated respect*" was adapted by Grol and Lawrence (1995), who refer to a 'peer' in medicine as "*a person working in the same branch of medicine who has comparable experience and training*."

When defining what constitutes a 'peer' in the general practice environment, they expanded this definition to include both established GPs and their trainees, or GPs and colleagues of different disciplines or professions working together in a practice**.** The term has also been used inconsistently in general practice studies where secondary care physicians (Kasje *et al*., 2004), pharmacists (Madridejos-Mora *et al*., 2004) and nurses (van den Hombergh *et al*.,1998) have been designated 'peers' in reviews of aspects of GPs' practice.

Bowie (2005a), when considering a specific model of peer feedback in general practice, described a peer within this context as "*an informed independent GP who has received appropriate training in the areas on which they are asked to review and provide formative constructive feedback to colleagues*." This definition recognised the expectation that a peer who is tasked with providing educational feedback to professional colleagues should at least have undergone related training in order that feedback is constructive and delivered within an acceptable framework (Pendleton, 1984). This reduces the risk of poorly delivered negative feedback being given to an individual.

Defining peer review is also complex because the concept and practice are applied differently depending on the healthcare and educational contexts. For example, peer

review has also been used interchangeably with the data collection feedback process that could be considered 'audit' by many GPs (Grol *et al*., 2005). For a feedback model available to GPs and their teams in the west of Scotland deanery, Bowie and Kelly (2007) defined peer review specifically as "t*he evaluation of one element of an individual's performance by trained colleagues using a validated review instrument to facilitate developmental feedback*." Although 'validation' is only one of the psychometric properties of the 'instrument' that requires to be considered in such a model (van der Vleuten, 1996), the expectation was that the assessment content would at least be relevant having been endorsed by 'experts'.

The term 'feedback' when applied in medicine is also used and interpreted in many different ways. There is little consensus amongst educationalists on its definition, which in turn can lead to misinterpretation and misunderstanding. In a medical education literature review which identified ten distinct definitions of the term 'feedback', van der Ridder *et al*., (2008) argue that none of these actually 'define', but merely assert the characteristics of feedback, such as its purpose and content. By merging these concepts and characteristics the authors developed the following definition of 'feedback' in relation to medical education as, *"specific information about the comparison between a trainee's observed performance and a standard, given with the intent to improve the trainee's performance."* The term 'trainee' in this definition relates to anyone in a learning situation. The feedback provider giving the 'specific information' is defined as "*someone who can envision a standard against which to compare the trainee's performance*".

A working definition of assessment is given as "*any purported and formal action to obtain information about the competence and performance of a candidate"* (Schuwirth

and van der Vleuten, 2006). From this definition it is not clear whether or how the information on the candidate is reported back to the individual being assessed. Some authors suggest that assessment should be undertaken for a specific purpose. This purpose can vary from purely formative (that is, it provides feedback to learners about their progress) to purely summative (whereby it measures the achievement of learning goals at the end of a programme of study) (Wood, 2007). However, others argue that feedback has to be thought of as a central component of assessment and cannot be distilled out (Rushton, 2005).

3.3 The purpose of peer review and related examples in UK general practice

Grol and Lawrence (1995) suggested that the purpose of peer feedback could be divided into two broad categories. Firstly, it could be directed at 'selection and control', for example, through evaluations to reduce inter-doctor or inter-practice variation in patient care or as a tool aimed at revalidation or reaccreditation. Secondly, it may be aimed at enhancing 'educational gain' by focusing on the learning and improvement process for GPs and their practices. The authors proposed that policy makers, funding bodies, NHS organisations and medical regulatory authorities were most likely to favour a peer review model where specific goals for attainment could be set to assure aspects of a practitioner's competence and performance. On the other hand, individual practitioners and their teams were more likely to emphasise the educational gain for themselves as care providers.

- '*Selection and control'*

'Selection and control' judgements can be considered as either 'high stakes' or 'lower stakes' decisions. For example, when the GMC is required to evaluate the performance of a doctor whose performance gives cause for concern, then this would be considered

'high stakes' since the impact of a negative assessment is clearly apparent for the individual practitioner. This involves the doctor undergoing a test of their competence and skills, as well as completing a portfolio to describe their training and experience. In addition, three trained reviewers - two of whom are 'specialty peers' - undertake a review of the practitioner's medical records, observe a number of consultations and discuss individual patient cases with the doctor (Southgate *et al*., 2001). The result of this evaluation by peers is then used to aid a decision on the doctor's fitness to practise.

When making a high stakes assessment it is necessary to understand the distinction between a doctor's competences (a demonstration of the ability to apply knowledge and skills in a specific situation) and their performance (what the doctor actually does in practice). Competence is a recognised prerequisite to satisfactory performance but it does not necessarily translate directly to patient care (Sanazaro, 1983). An example of this was reported by Rethans *et al*. (1991), who compared four standardised consultations during a normal working day (performance) with the same problem presented in a controlled experimental practice setting where time was not a limiting factor (competence). The number of 'appropriate' actions taken by the GPs (such as history, examination and treatment) was increased in the experimental setting. Issues such as efficiency (prioritising obligatory over additional actions) and consultation time were shown to affect performance.

In contrast, lower stakes judgements are often made by peers when verifying checklists or requirements for accreditation purposes. Although professionally important to the doctor, these are not likely to be career threatening. For example, since the 1980s members of the Royal College of General Practitioners (RCGP) who were applying to become fellows of the RCGP submitted a portfolio of evidence to three reviewers, two

of whom were themselves fellows by assessment (FBA). The reviewers then visited the practice to assess a cross-section of the candidate's work, such as a video of patient consultation and patient record review. The evidence from the portfolios and visits were checked against over 60 criteria to be met in 12 sections relating to the GMC's good Medical Practice for General Practice guidelines. If successful, the candidate was awarded the fellowship (RCGP, 1997; Holden and Kay, 2005). Since 2006, if applying for RCGP fellowship, candidates have to submit a portfolio of evidence that is judged by an adjudication panel (which includes GPs) and evaluated against published criteria (RCGP, 2008a). A similar model is in place for the interim membership by performance (iMAP) of the RCGP, where the candidate's portfolio is independently reviewed by two assessors (RCGP, 2008b).

The RCGP Quality Practice Award (QPA) was introduced in 1996 and is a process undertaken by GP practices to demonstrate that the quality of patient care contributed by all members of the practice team is of a high standard. The practice submits a portfolio of written evidence against 21 sets of criteria, such as accessibility to healthcare professionals, management of chronic diseases and team-working. When a practice's evidence has been accumulated and completed, an assessment visit is conducted by a panel of four visitors, which comprises a combination of GPs, a nurse, a practice manager and a lay assessor (RCGP, 2008c).

Other similar accreditation examples would include health authorities that use teams of trained external peer reviewers to monitor compliance with the nGMS contract in GP practices; and in GP training, where practices until 2007 underwent regular external peer feedback as part of gaining or retaining their accreditation to train GP registrars (Joint Committee on Postgraduate Training, 1992). Currently in Scotland only a

selection of established training practices and all new training practices will undergo assessment.

- '*Educational Gain'*

With the introduction of GP appraisal, nearly all GPs in Scotland have been exposed to peer review processes that aim to support and enhance learning. As described in Chapter 2, this involves GPs reviewing aspects of their own professional learning with a named peer in order to facilitate further learning and development. Although GPs are required to 'engage' with the appraisal process, there is no formal verification judgement made on the evidence presented for appraisal.

A further example of 'educational' review involves the Scottish Prospective Trainers Course (SPTC). This was developed in 2001 to standardise training for those General Practitioners (GPs) in Scotland who wish to become GP Trainers (Shackles *et al*., 2007). Participation and subsequent completion of the SPTC prepares and enables a GP to take up the role of a GP Trainer who can then supervise and mentor a doctor during their training year as a GP registrar. In order to enrol on and complete the SPTC, GPs must commit to submitting a video of six patient consultations, a completed criterion audit and a significant event analysis report. Submitted materials are then reviewed by trained GP peers and independent developmental feedback is provided.

3.4 Evidence for the effectiveness of peer review as a quality improvement method in improving healthcare in general practice

When considering what Grol described as 'educational' peer review, the evidence for its the impact is inconclusive. In the Netherlands, local peer review networks of GPs are well-established in the healthcare system (Grol, 1995; Kasje *et al*., 2005). Feedback from colleagues is encouraged within the groups, particularly on ways to overcome

barriers to implementing change within practice (Grol *et al*., 2005). In these small groups national practice guidelines are frequently used to initiate discussion within the group on how the practitioners could improve their care.

In one study where guidelines and comparative data were provided, there were improvements in the prescribing of oral steroids in acute asthma and in the prescribing of short courses of antibiotic treatment for urinary tract infections (UTI) (Veninga *et al*., 2000). Another study using this method demonstrated improved test-ordering behaviour by increasing knowledge about the clinical guidelines requesting laboratory investigations and by practising communication skills to reassure patients that no test was needed (Verstappen *et al*., 2003).

In the UK this approach with similar peer groups has led to reported improvements in interviewing skills and techniques in general practice consultations (Verby *et al*., 1979). Standard setting by peers improved prescribing and follow-up in conditions such as cough, wheeze and bed-wetting in a study of general practices in the north of England (North of England, 1992).

Peer processes that could be considered as primarily 'selection and control' procedures have also been shown to be effective. In the UK, up until 2007, the standards expected to be attained in order that a GP practice could achieve 'training practice' status were set by the body then responsible for GP training, the Joint Committee on Post Graduate Training for General Practice (JCPTGP). Verification that a practice had met the requirement was carried out by a peer review practice visit. This appeared to be a strong incentive to employing more staff, undertaking preventative health screening, being better equipped and having summarised case records (Baker, 1985), as well as

being more innovative in areas such as staffing, premises and audit (Baker and Thompson, 1995).

Peer review practice visits undertaken by members of a primary care group in Yorkshire to make baseline assessments for the clinical governance process were reported to have benefits for the practices in areas of chronic disease management, risk management and education and training (Vautrey and Neal, 2000).

Holden and Kay (2005) reported that care had improved for those patients whose GPs were successful in achieving the RCGP award of fellowship by assessment. Part of this success was attributed to a culture of continuing education but no specific details on how care was improved were given.

When attempting to evaluate the success or otherwise of peer feedback, it is often difficult to tease out the difference between and impact of audit and peer review methods since these can be used interchangeably. For example, in a Cochrane review of the effects of these two techniques on professional practice and healthcare outcome, Jamtvedt *et al*. (2007) concluded that both can be effective in improving professional practice but the effects were generally small to moderate. The importance of the interventions could not be clearly distinguished. Similarly, in reviews of the effectiveness of medical education strategies and their effects on patient care, although audit and feedback were both effective to a 'moderate' extent, it was not apparent what the 'size effect' of the individual components was (Oxman *et al*., 1995; Davis *et al*., 1995).

Peer feedback, however, is not always successful. An intensive small-group education and peer review programme aimed at implementing national guidelines on asthma and COPD was not effective in changing aspects of care such as pharmacological and non-pharmacological treatments. Nor was it effective in changing patients' health status (Smeele *et al*., 1999).

In Veninga *et al*.'s study on the effect of peer discussion on national guidelines for prescribing and comparative prescribing data there were improvements in some areas of prescribing. Conversely, feedback had little success in improving the prescribing of inhaled steroids for chronic asthma or in altering first line prescribing choice for urinary tract infections (Veninga *et al*., 2000).

Kasje *et al*. (2005) examined an educational programme for peer review groups to improve the treatment of chronic heart failure and diabetes in general practice. The authors examined an intervention in the form of guidelines discussions within the groups to highlight best practice in the prescribing of Angiotensin Converting Enzyme inhibitors (ACEI) in patients with heart failure and the prescribing of ACEI in hypertensive diabetic patients. Despite discussion on barriers for practitioners in implementing best practice and suggestions from colleagues on how to overcome these difficulties, there was no difference in the prescribing pattern of GPs within the intervention group. The authors concluded that an interactive programme for peer review groups focusing on education, individual feedback, identification of barriers and social influence was not successful in changing the treatment of 'chronic' patients.

A postal survey study to assess the impact of the Scottish appraisal system was reported by Colthart *et al*. (2008). The findings indicated over 80% of GPs self-reported

either no change or only minimal change in their clinical care of patients as a result of appraisal, while over 60% of respondents thought the appraisal process either did not alter or changed in only a small way their approach to maintaining good medical practice. This led the authors to conclude that many GPs perceive little or no benefit for either themselves or patient care from the process.

3.5 Recruiting and training peer reviewers to deliver effective feedback

One important aspect of the feasibility of any peer review model is the ability to recruit individuals who are prepared to participate as professional reviewers of aspects of their colleagues' performance and provide appropriate feedback. This will depend on factors such as whether reviewers need to be local, national or international, the degree of 'expertise' required, and the acceptability of the process to the proposed reviewers (Callaham, 2003). Where difficulties in recruitment have been experienced, suggested methods of recruiting reviewers include inviting professionals who are known to organisers or administrators of the model, requesting participants to suggest reviewers or contacting existing reviewers to nominate other professional colleagues (Smith, 2003).

For the learner there is an expectation that reviewers giving educational feedback will be professionally credible and acceptable. Grol (1995), in a summary of the literature on this topic, reported that peer review would appear to be particularly effective if, amongst other factors, it is provided by colleagues who are respected by the learner. The credibility of the reviewer is also known to be important to individual practitioners in their perception and acceptance of feedback (Sargeant *et al*., 2005; Bing-You *et al.,* 1997).

It may also be possible that reviewers are credible but not acceptable to participants. For example, anecdotal evidence suggests that while GPs in one region of Scotland may be prepared to submit video consultations of their consulting skills to a peer reviewer in a different health board region (who they would perceive as unlikely to recognise them), they would be reticent to submit the same material to a local reviewer. This is because it raises the prospect of professional embarrassment if the assessment highlighted aspects of consultation technique which were poor or management decisions that were incorrect.

Within the context of medical education, it is desirable that anyone who is involved in giving feedback should be skilled in the process. Despite this, reviewers have been described as learning their trade like apprentices (Rennie, 1999). Schroter and Groves (2004) outlined four aims for any process used in the training of peer reviewers for medical journals. Three of these could be applied to most models of medical peer review: to make clear what constitutes a fair, specific and constructive review for the learner; to understand what matters to those commissioning the review and to produce reviews that reflect the qualities desired from training.

It is highly important that reviewers are trained in the principles of effective feedback. These were described for clinical medical education by Ende (1983) and the elements of good feedback were summarised by Wood (2007) as:

- *Helping clarify what is good performance*
- *Facilitating the development of reflection in learning*
- *Delivering high-quality information to the participant about their learning*
- *Encouraging reviewer and peer dialogue around learning*
- *Encouraging positive motivational beliefs and self-esteem*

- *Provides opportunities to close the gap between current and desired performance*
- *Peer reviewers need to be chosen appropriately and trained in delivering appropriate feedback*

The importance of each of these elements of feedback will vary depending on the type and purpose of the feedback model, but it would seem essential that reviewers are not only trained but updated in the elements of good feedback practice. This may not always be achieved. In their review of peer assessment instruments used for judging professional performance, Evans *et al*. (2004) found that none of the studies described rater training when developing the instrument. Guidance, if given, consisted only of short written instructions.

The format of training may also be important. In a randomised controlled trial, Schroter *et al*. (2004) studied the effects of training on the quality of journal peer reviewers' feedback. The authors reported that short training packages resulted in a non-significant increase in the 'quality score' given to a review in the short term but the benefit of training was no longer apparent after six months. However, in a peer review programme aimed at assessing the professional competence of primary care physicians, the authors held annual workshops with assessors to review potential problems and were able to introduce changes in the process of the evaluation and training of the reviewers that increased the reliability of the assessment (Norman *et al*., 1993).

3.6 The personal impact of peer feedback

Grol and Lawrence (1995) reported that in their experience, doctors who take part in peer review undergo an educational process that allows them to adopt a self-critical

stance with regard to their professionalism. Effectively, they need to undertake 'critical thinking,' described by Greco and Eisenberg (1993) as a disposition to carefully consider problems that present in practice and then be able to act on feedback given.

However, the ability of doctors to be self critical with feedback varies. Sheldon (1982), when assessing feedback given to doctors who participated in audit, described two situations that can arise as a result of feedback. Firstly, when their performance equates with good performance, then this is accepted as normal and to be expected, or boosts self confidence thus encouraging ongoing participation in peer review. Secondly, if the doctor's performance does not meet expected or defined standards, then, if the doctor is aware of the discrepancy, they may accept the feedback and tackle any deficiency. Alternatively, their self-esteem may be lowered, resulting in a 'cognitive dissonance' that leads to defensiveness on the part of the doctor.

An example of GPs disagreeing with their feedback was given by Sargeant *et al*. (2005). In a study exploring GPs' reactions to peer and non peer feedback in Canada, the authors found that participants who received what they interpreted as negative feedback did not agree with their feedback, nor were they inclined to use it for practice improvements. Issues around the objectivity of the assessment, credibility of reviewers and quality of feedback were more likely to be raised by this group as perceived reasons for the 'poor' feedback.

These findings would be similar to studies in industry where managers who had undergone performance review and who had received negative feedback often become discouraged and poorly motivated (Pearce and Porter, 1986). In addition, assessments

that compared peer rankings led to lower self esteem for those not ranked in the upper quartile of distribution, even for those individuals ranked as 'satisfactory'.

One possible mechanism to improve self-esteem issues arising from feedback was suggested by Sargeant *et al*. (2005), who proposed that a mentoring service should be made available to the learner to support reflection, the implementation of change and increase the acceptance of feedback. This proposal would need to be considered within the overall feasibility of any peer model.

3.7 SEA and the educational model of peer review in the NES west of Scotland region.

At the commencement of this thesis in 2003, a voluntary educational model for peer review of SEA reports had been available to all GPs in the west region of NHS Education for Scotland (NES) for five years, as part of the arrangements for continuing professional development. The model was initially set up as a means of promoting and enhancing educational understanding of the SEA technique. A key motivation was the judgement and verification of the quality of an educational process that had possible implications for patient safety.

This model existed concurrently with the post-graduate education allowance (PGEA) system that formed part of the General Medical Services Contract until 2004. This system made a financial payment to GPs for attending 50 educational sessions over five years. Each session was the equivalent of 2.5 hours' educational work. This peer review model was applied uniquely in the west of Scotland by offering one PGEA session for each SEA report submitted.

GPs who submitted an SEA report were asked to submit this in a standardised format that encouraged a structured analysis of the event. This consisted of four sections that prompted the report's author to describe 'What Happened?', 'Why did it happen?', 'What have you learned?' and to 'Describe any change implemented'. Participants are asked to anonymise the SEA report before submission so that healthcare practitioners' details, patient details and places of work remained confidential. Where this was not applied it was carried out by the peer review co-ordinator (a GP, this author, who has specific responsibility in NES west of Scotland region for implementing and ensuring the ongoing function of the peer review model). Each report was sent to two trained GP reviewers chosen from a group of twenty who form the west of Scotland ADG. The professional characteristics of these 20 reviewers at the commencement of this thesis are detailed in Box 3. Each SEA report was at this time assessed using a four criteria instrument (Figure 2). Reviewers were asked to provide formative feedback on the quality of the report and to make a dichotomous judgement as to whether the SEA report was of a satisfactory quality or not.

GPs who had submitted a report considered satisfactory by both reviewers were sent a letter at this time confirming this. If either of the two reviewers considered the SEA report 'unsatisfactory', the report underwent a 'second level' independent assessment by this author and a trained colleague, who is based in the deanery. After discussion between these two reviewers, a final decision was made on whether the report was satisfactory or not and agreement reached on the level of formative feedback to be provided to the participating GP. A summary of this process is detailed in Figure 3.

3.8 Summary

Various methods of peer review have the potential to improve the different aspects of the quality of a GP's work, although not all are successful. With the increased participation of GPs and their teams in SEA, particularly through nGMS and appraisal there is a gap in the evidence as to whether external peer review has a place in enhancing the educational value of participation in SEA for GPs. The woS peer review system for SEA described above offered a starting point to investigate whether such a model could be updated to improve the learning from and the quality of GPs' SEA reports.

Box 3: 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 five 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 GPs
- 2 (10%) are GP appraisers
- 10 (50%) are GP registrar trainers.
- A further three (15%) have other general practice educational roles e.g. associate adviser, undergraduate tutor

Figure 2: Peer review instrument for a significant event analysis report - 2002

SIGNIFICANT EVENT ANALYSIS

Peer Review Instrument **Project no……**

1. What happened?	Has personal impact Important to individual or organisation Causes reflection	☐
2. Why did it happen?	Clear reason sought	☐
3. Was insight demonstrated?	Aware of previous suboptimal care Decision – making process altered Assessment of "risk" demonstrated Level of personal responsibility linked to circumstances	☐
4. Was change implemented?	• Yes – describes implementation of relevant change • No - risk of similar significant event unlikely • No. Unable to influence change but suggestions for change given	☐

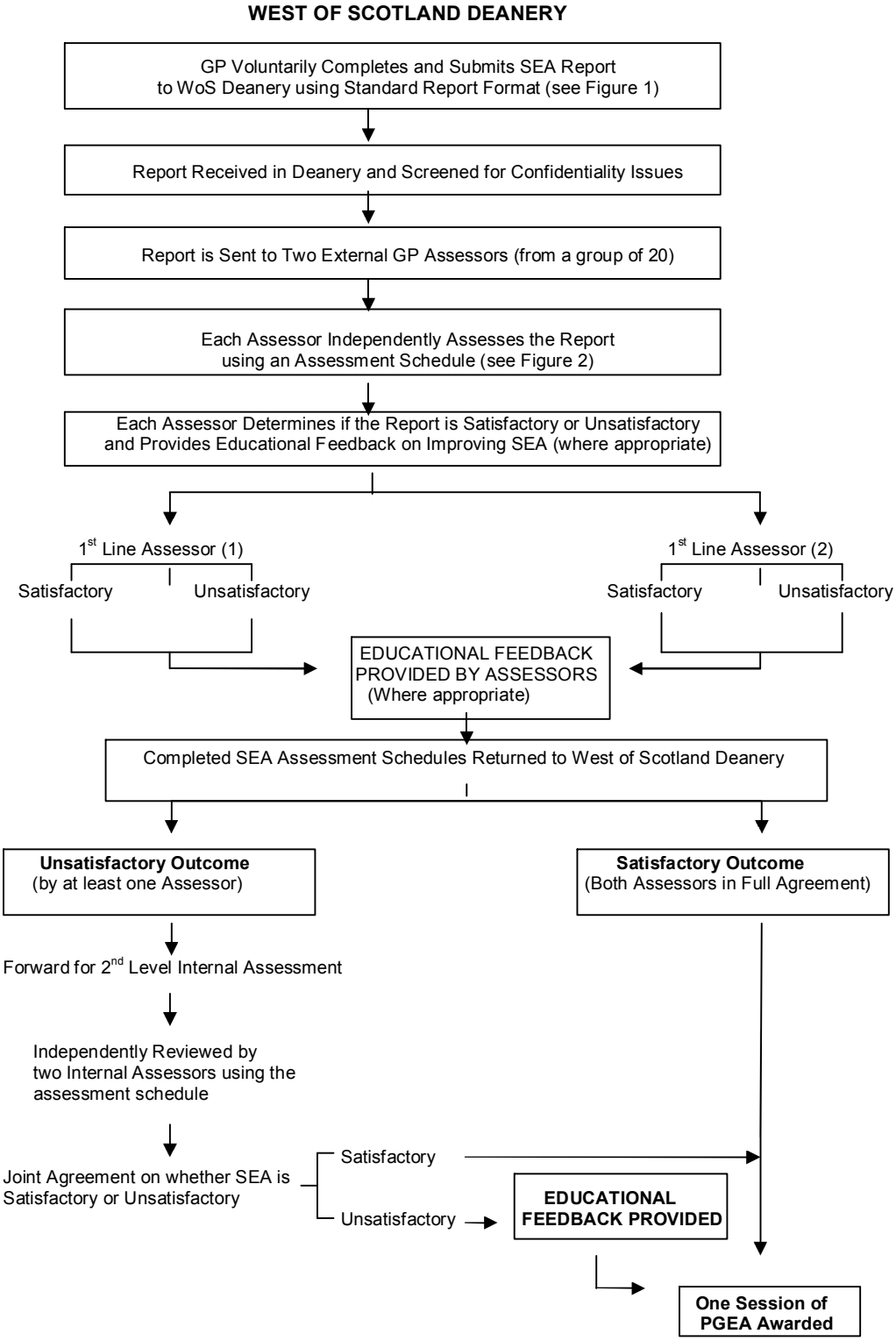
Satisfactory analysis of significant event Yes ☐ No ☐

Comments for feedback (continue overleaf)

Assessor signature ……..

Capitals …… **Date……..**

Figure 3: System for the submission of a significant event analysis report for peer review by GPs in the west of Scotland (Bowie, 2005c)



CHAPTER 4

Aims of the thesis

4.1 Background

At the start of this thesis the evidence base relating to significant events, their analyses and potential use in reporting and learning systems was limited. However, over the six-year period of this study, the previously discussed policy changes had taken SEA from a minority, voluntary and non-embedded activity in UK general practice to one that was expected to be undertaken by the vast majority of practitioners and GP practice teams as part of contractual (and in Scotland appraisal) requirements.

Under the 'umbrella' of clinical governance, GPs and their teams were expected to identify, analyse and, where appropriate mechanisms existed, report relevant significant events. The associated implication was that GPs and their teams were demonstrating a commitment to addressing issues related to patient safety as well as undertaking reflective learning. However, it was unclear if this assumption accurately reflected the reality in general practice

In addition, the role of educational peer review of SEA in general practice was largely unknown and untested at the outset of this thesis. The existing woS peer review model had been set up in the context of the 'old' general medical services contract. With the introduction of the new professional appraisal and contractual requirements in Scotland, a potential need existed to develop and subsequently evaluate the SEA peer review model.

Against this backdrop of limited evidence and policy rhetoric, there was therefore an opportunity to investigate issues that affected GPs in relation to the identification, grading, analysis and reporting of significant events, and to establish whether peer review of SEA was an acceptable, feasible and worthwhile educational undertaking.

4.2 Consideration of hypothesis generation and testing

Many theses, particularly those of the 'experimental' tradition of medical science, are hypothesis-based.

Although a research hypothesis can help clarify the aim of a research project, there may be occasions when it is not appropriate, particularly in the context of medical education and assessment (McIntyre and Popper, 1983). For example, Norman and Eva (2007) state that the development of an evaluation instrument is not served by having a null hypothesis such as "*the reliability of the new reflective practice instrument will be less than 0.5*." They contend that effort should be directed towards appropriate research questions, ignoring the 'stylistic' effort of hypothesis generation, and instead deal with, for instance, the 'psychometric' properties of the subject under investigation. As an overall null hypothesis for this thesis would be difficult to construct and could make little sense, one was not put forward.

4.3 Research Questions.

The research supporting this thesis aimed to explore two broad areas:

1. To identify and explore the issues involved if the identification, analysis and reporting of significant events are to be associated with quality improvement in general practice.

The specific objectives of studies supporting this aim were:

- to assess the attitudes of GPs towards the identification and reporting of significant events,
- to examine the levels of agreement on the grading, analysis and reporting of selected significant event scenarios by GPs,
- to identify the range of quality and safety related issues addressed by SEA reports submitted to the woS peer review model, and
- to establish the reported types of learning needs raised and actioned by GP teams using the SEA technique.

2. Can a peer feedback model enhance the value of SEA so that its potential as a reflective learning technique can be maximised within the current educational and contractual requirements for GPs?

The specific objectives of studies supporting this aim were:

- to develop and test aspects of the validity and reliability of a formative assessment instrument to facilitate the review of SEA reports by trained peers,
- to explore the views and experiences of trained GP peer reviewers of SEA reports in the woS model,
- to examine the acceptability and educational impact of the woS peer review model for participants; and,
- to investigate any difference in the outcome of assessments between clinical audit specialists and experienced medical peer assessors using the woS SEA peer review instrument.

4.4 Scope and limitations of the thesis

As described above, the work in this thesis took place during a time when there were major policy, contractual and professional educational initiatives for GPs. As a result, some of the objectives evolved over time and were not necessarily immediately apparent at the beginning of the research. For example, the introduction of the national Scottish appraisal system had placed a major emphasis on individual practitioners demonstrating participation in SEA that was not envisaged in 2002. In contrast, although a national reporting system was introduced in England and Wales, perhaps against expectations a similar system has not been introduced in Scotland.

The scope for research on significant events is wide because of the relative paucity of existing studies. For instance, much of the early literature on SEA emphasised its role as a team-based activity. The potential role of different team members in the identification, analysis and potential reporting of events and the associated inter-personal dynamics of this could have been subject to investigation. However, the research was focused on GPs rather than other healthcare staff. The justification for this is that GPs are frequently viewed as the 'leaders' of the team, have professional responsibility for governance and safety matters, and are most likely to take part in the SEA peer review model.

The main focus of the thesis therefore relates to either GPs in one health board area in Scotland – NHS Greater Glasgow and Clyde - or to a self-selected, motivated group of GPs from the west of Scotland region (which covers five NHS boards including greater Glasgow and Clyde) who submit or review SEA reports.

The limitations associated with this approach include the role that significant events and their analyses play according to local governance requirements. The SEA peer review model has not been used by the majority of GPs in the deanery and by necessity involves reports that are selective, are ‘owned’ and interpreted by the author who may not be the main protagonist, and can serve only as a ‘proxy’ for the event analysis. No external verification of the details of the report takes place.

Within the context of the scope and limitations described, this thesis set out to test the aims and objectives stated above.

CHAPTER 5

Methods

5.1 Introduction

The basis of scientific investigation is the application of rigorous methods of analysis to the particular research issue or question. These methods are often described as either quantitative or qualitative. This is recognised as a rather artificial dichotomy. As a general description, *quantitative* research (such as randomised controlled trials and observational studies) aims to be objective, deductable and generalisable. On the other hand, *qualitative* methodologies (such as grounded theory and phenomenology) are more likely to be to be subjective, inductive and non-generalisable (Kuper *et al*., 2008a). They investigate how people behave in a particular social setting and explore what meanings people actually wish to convey when describing personal experiences (Cresswell, 1998a). A combination of quantitative and qualitative methods has been applied in the studies undertaken for this thesis.

Although these two approaches have previously been construed as incompatible, there is increasing acknowledgement that in health care they can be complementary (Dixon-woods *et al*., 2001). For example, taking a qualitative approach to understand why a patient is poorly compliant with medication for hypertension (perhaps due to a side-effect or a failure to understand the need for life-long compliance) can complement existing quantitative data – that reducing blood pressure can reduce cerebro-vascular disease. Individual interviews and focus groups have previously been used to generate hypotheses and questions for questionnaire studies (McLeod *et al*., 2000; O'Brien, 1993).

5.2 Quantitative methods and data analysis

5.2.1 *Background.*

Advances in natural, physical and medical sciences have traditionally been based on the 'scientific method' of observation, hypothesis generation, prediction and experimentation that adds to the accumulation of knowledge (McIntyre and Popper, 1983; Illing, 2007). A central pillar of these studies has been quantitative research methods, with their focus on answering the "what?" and "how much?" questions in research (Kuper *et al*., 2008a).

The 'gold standard' approach in clinical medicine is often taken as the 'experimental' randomised controlled trial (RCT). Other quantitative studies are described as 'observational' studies. Three types of observational study are commonly described in healthcare: cohort studies, case control studies and cross-sectional studies. Cross–sectional studies, such as those undertaken in Chapters 6 and 7 of this thesis, provide a 'snap-shot' of the issue under study at that particular point in time (Lowe, 1993).

5.2.2 *Postal questionnaire surveys*

Postal questionnaire surveys have been used extensively in medical education research. They are particularly useful in identifying the attitudes, perceptions, values and experiences of large groups of individuals (Boynton and Greenhalgh, 2004a). The processed data may suggest information that are generalisable to a particular group beyond those who have chosen to participate in the study. Surveys can also generate large volumes of standardised data to be collected. In addition, these data can be given anonymously by potential participants who are thus encouraged to be open and honest in their replies (Oppenheim, 1966). The usefulness of survey questionnaire data is constrained by the limited ability of people to accurately assess their own behaviour

(Drewnowski, 2001). Data can be affected by the characteristics of the respondents: some may not take the survey seriously while others may not accurately report their beliefs and attitudes (Boynton and Greenhalgh, 2004a; Boynton and Greenhalgh, 2004b). In addition, it is difficult to know the characteristics of non respondents and hence whether the survey sample is representative of the population under study (Robson, 2002).

As described previously, the two postal questionnaire surveys featured in this thesis (Chapters 6 and 7) are designated as cross-sectional studies since they attempt to collect data on significant events from subjects (GPs) within a short time frame. They highlight the advantages of this type of study, in that a large volume of data (in this case on significant events) can be collected in a short time scale in comparison to cohort or observational studies.

Four principal considerations have been described in designing a questionnaire (Fallowfield, 1995):

- *Does an appropriate questionnaire already exist?* There is clearly no point in developing a questionnaire if an acceptable and useful one already exists.

- *Who will complete the questionnaire?* This is particularly important in areas such as old age psychiatry and paediatrics where intended respondents may be incapable of reply. Additionally if a questionnaire contains ‘open’ questions that require textual replies the researcher needs to be aware of whether respondents are ‘educationally’ able to do so. The use of ‘language’ particularly ‘jargon’ also needs to be considered in relation to the target group. Many GPs for instance

would be unfamiliar with a questionnaire related to significant events which does not outline and define what a significant event is.

- *Are questions brief, relevant and unambiguous?* Items need to be clear and unambiguous. Co-operation is likely to decline if the questionnaire has too many items and these are not seen as relevant to the respondent. Questions can create ambiguity if they ask about more than one issue within the same phrase. For example, 'how do you identify and analyse significant events?'. Methods used by healthcare staff to identify events will not be the same as those used to analyse such events. Questions such as, 'have you analysed a significant event recently?' do not identify the time frame for the respondent, for example, whether within the last day, week or month. Leading questions such as, 'do you agree that significant event analysis is a worthwhile patient safety technique?' should be avoided.

- *What response format will be used?* Some surveys may prioritise categorical scales that are easy to score with answers such as 'yes/no' or 'true/false'. These are 'closed' questions that facilitate a quick response and make data easy to analyse. They do, however, restrict the respondent in their reply as the questionnaire will effectively gather data that may reflect the researcher's agenda. To avoid this, many questionnaires will include a free text section so that additional or alternative answers and issues can be highlighted. Adjectival (where descriptors are placed along a continuum varying from, for example, 'very poor' to 'very good' with a middle descriptor of 'fair'), or Likert scales (with bipolar descriptors that range, for example, from 'strongly agree' to 'strongly disagree' with a 'neutral' descriptor in or around the middle of the scale) can be

considered as categorical data, although there is debate as to whether the resulting data are either 'ordinal' or 'interval' (Carifio and Perla, 2008).

Categorical data collection should be avoided if replies of a 'continuous' nature are expected. Continuous judgments can be given using a visual analogue scale where a simple line is drawn with two descriptors. Alternatively, data could be obtained by using 'open' questions, although consideration needs to be given as to whether participants are willing to engage with a longer, more taxing questionnaire.

In addition, the design of a questionnaire should aim to maximise its validity and reliability in order to demonstrate a degree of scientific rigour. In the context of survey design, validity is the extent to which the questionnaire actually measures what it is supposed to measure rather than some other concept (Carmines and Zeller, 1979). A questionnaire survey analysing attitudes to SEA will not be valid if it explores attitudes to 'audit' in general. Reliability has been defined as the extent to which the same measure gives the same results on repeated occasions (Carmines and Zeller, 1979). For example, we would expect the use of a questionnaire survey to collect consistent responses from participants if they were asked to state whether they were male or female, but not to factors that can vary from day to day, such as emotional well-being.

The entire survey design should aim to minimise bias. This can be described as producing results that are actually different from the true population value. It may be introduced through aspects such as:

- sampling bias (for example excluding locum GPs in a survey aimed at establishing the views of all GPs)

- respondent bias (where those who choose to take part in the survey can have different characteristics, opinions and motivations to those who decide not to participate)
- recall bias (where the passage of time can affect the accuracy of recall in relation to an item).

Bias is often detected by comparing data with that obtained from previous research in the field of interest (Campanelli, 2008).

Many of the potential problem issues discussed above can be avoided by pre-testing (pilot testing) the questionnaire on a group of individuals who are similar to the intended target group for the study. For instance, the questionnaire used in Chapter 6 of this thesis was pre-tested amongst GP educator colleagues. It was then field-tested amongst a larger group of GP principals in one local health care co-operative in Greater Glasgow NHS board. Feedback to the researchers on issues such as question relevance, participants' understanding of questions and the format of the questionnaire allowed for adjustments to be made prior to the main data collection exercise.

One method of enhancing the quality of data is to maximise the survey response rate. This minimises the risk of non responder bias (McColl and Thomas, 2000). Non-material factors that can secure a good response rate have been categorised into three main headings (Robson, 2002):

- The design and layout of the questionnaire: The appearance of the questionnaire is vitally important and its contents must be arranged to maximise response. The design and wording of questions should be optimised, as previously described. Coloured paper can also enhance return rates (Beebe *et al*., 2007). Initial questions should be 'easy' and non-

threatening with more difficult questions occurring towards the middle. If the last questions are more 'interesting', then this is liable to increase the chances of a 'return'. It is helpful to sub-sectionalise groups of questions (e.g. 1a, 1b, 1c.) so that specific topic issues are dealt with prior to moving onto a different section. For example, a section on gathering respondents' experiences of how significant events are identified would precede a section on the perceived attitudes to reporting of identified significant events to appropriate health care bodies. A short sentence at the end to thank participants for their help and an offer to send information on study findings can also be useful.

- Covering letters: A personalised letter explaining the purpose of the study on organisational headed notepaper should be included. The potential participant should be assured of confidentiality and anonymity (where appropriate). Information on, and inclusion of, a pre-paid reply envelope should be given.

- Follow up: Respondent follow up is the most effective means of increasing response rate. A follow up cover letter, questionnaire and prepaid reply envelope may be sent to non respondents. The cover letter should stress the importance of the study and why the participants support would be valued and appreciated. There are no set timescales for follow up, although two to three weeks is often used as a guideline. The number of follow up requests will be limited by time and financial constraints but will not generally exceed two (Boynton, 2004).

From a healthcare perspective there is no 'gold standard' for an acceptable questionnaire response rate. McColl and Thomas (2000) suggest that a response rate equal or greater than 85% is excellent, a rate between 70% and 84% can be considered very good, with a rate of 60% to 69% being acceptable. A barely acceptable response rate would be below 60% with a rate that falls below 50% being considered unacceptable. However, this is a theoretical approach since the reality is that survey research with less than a 60% response rate is regularly published, either because the actual sample size is substantial and/or because the topic is important. An example of this would be a recent survey that was the first to describe a national sample of GPs' attitudes to the Scottish appraisal system (Colthart *et al*., 2008a) although the effect of respondent bias may limit the findings.

5.2.3 *The psychometric evaluation of assessment instruments.*

Psychometric methods are designed to ensure that data collected as part of related research are sufficiently robust to allow meaningful interpretation. Van der Vleuten (1996) suggested that in the development of any measurement instrument, there are five characteristics of 'good' measurement: validity, reliability, feasibility, acceptability and educational impact. These factors when taken together are referred to as the 'utility' of the instrument. This in turn has to be viewed in the context of the purpose of the instrument and the population to whom it is applied or on whom it is tested (van der Vleuten and Schuwirth, 2005). For example, the purpose of the proposed assessment instrument developed as part of this thesis is to provide formative educational feedback to a general practitioner on the analysis of a chosen significant event by trained GP colleagues. It would not be appropriate to apply the instrument to a population of first year nursing students because they are unlikely to have experience either of SEA or general practice.

Assessments must also balance rigour (validity and reliability) with practicality (feasibility, acceptability) and educational impact (consequential validity) in achieving their overall utility. The balance will depend on the purpose of the assessment, although the various components have been described as multiplicative, meaning that if one aspect is untested then the overall utility is zero (McKinley *et al*., 2001).

The importance of each attribute will depend on the purpose of the instrument used. Most instruments will have to 'trade-off' one or more aspects of utility depending on the purpose and stakes of the assessment. A high stakes examination such as the nMRCGP that must be passed by prospective GPs to allow them to undertake independent general practice requires a high level of validity and reliability. A formative peer feedback instrument for consultation or audit skills requires an increased emphasis on educational impact but a decreased requirement for high reliability.

In an evaluation by Evans *et al*. (2004) the authors examined papers describing instruments for peer or colleague review. They reported that only three studies met their inclusion criteria of having sufficient psychometric data about either their development or their validity and reliability to make a decision on their appropriate use. No instrument was rated as having met all established standards of instrument development. This finding mirrors that of a review of the utility of assessments used in postgraduate medical certification. Hutchinson *et al*. (2002) found that of 55 papers examined, only two evaluated consequential validity and 'a handful' construct validity.

Validity Testing

As described previously, validity is the extent to which an instrument actually measures what it is supposed to measure rather than some other concept. Validity studies have focused on several aspects of individual forms of validity testing (Downing, 2003). The most common are face validity, content validity, criterion validity and construct validity. All of these forms of validity can be applied to establish a basis for an assessment to be valid.

Face validity refers to the belief or perception by individuals using or being assessed by an instrument that "on the face of it" the instrument measures what it is purporting to measure. For example, a provisional assessment instrument for SEA may be sent to a group of GPs with an interest in health care risk and safety to determine if, in their opinion, the instrument appears to measure the necessary factors commonly associated with analysing a significant event. If there is majority agreement that the instrument does indeed cover all of the important factors that would be associated with the assessment of an event analysis, then it can be said to have ‘high’ face validity. Conversely, if most respondents feel it does not, then the instrument would be considered to have "low" face validity.

Content validity is the determination of the content and the representativeness of the items ‘contained’ in an instrument. It is generally considered to be a more rigorous test than face validity. It can be demonstrated by a two-stage process: content development and content judgement (Carmines and Zeller, 1979; Waltz and Bussell, 1981; Lynn, 1986)

The content development stage involves two phases: firstly, the identification and generation of items relevant to the various 'domains' of the instrument; and secondly, the assimilation of the items into a usable instrument.

- The identification and generation of instrument items.

In devising an instrument to measure the effectiveness of GPs in analysing a significant event, it is necessary to first of all devise items which are appropriate for that scale. Theoretically it may be desirable to start with a blank sheet of paper, however, an accepted initial step is to look at what has been carried out previously (Goldberg, 1971). There are several reasons for using earlier work. Firstly, it may save time generating new items; secondly, previous items may have already gone through psychometric assessment and been shown to be appropriate; thirdly, there may be a limited number of ways to tap into a specific question; and finally, there may already exist a source of knowledge on the assessment of event analyses (Lough, 2002).

Five possible sources for generating new items were suggested by Streiner and Norman (2003): subjects themselves, expert opinion, research, clinical observation and theory. No single source is sufficient for item generation and most instruments will use a mix of these sources. The researchers involved in the development of the SEA assessment tool in this thesis (Chapter 9) had three of these sources readily available: previous focus group data from subjects, "expert" opinion and research data to inform them on item generation (Bowie *et al.*, 2005a; Pringle *et al.*, 1995; Bradley, 1992). Clinical observation, while not directly relevant to this instrument, could be taken as analogous to observation of previous SEA reports submitted for peer review.

There are however three suggested limitations with this methodology. Firstly, the items may be outdated (i.e. the subject under review may have been moved on in terms of both language and scientific content); secondly, previous scales may have had limited validity and reliability testing in the specific field to which the new instrument applies; and finally, the researchers' own belief that the existing instruments are insufficient for their own purpose.

- Content Judgement

The judgement phase of content validity can be carried out using a Content Validity Index (CVI), as described by Waltz and Bausell (1981) and quantified by Lynn (1986). A content validity index involves the judgement of a small (usually less than 10) number of experts that the items are "content valid." In other words, that they are appropriate items to include in the assessment instrument and that the instrument overall is appropriate.

At this stage the experts can comment on other items which they think should be included. Each item and the overall instrument is rated on a four-point scale, where 1 denotes an item as "not relevant" and 4 denotes an item as "very relevant and succinct". The total proportion of experts who score 3 or above for each item can then be calculated to establish the content validity of the instrument. The experts are also asked to detail any domains or items which they feel have not been covered by the instrument.

Limitations of the CVI as a method of assessing content validity have been described by several authors. These include the possibility that:

- the scale is used inconsistently
- there is a chance agreement by the small number of experts on the CVI (Waltz and Bausell, 1981),
- there is debate over what constitutes an "expert" (Fink *et al*., 1984)

Construct validity traditionally refers to the extent to which the measurements obtained by an instrument correlate with expectations (often unobservable qualities) and the demonstration that other factors are not confounders (Hutchinson *et al*., 2002). For instance, the ability to analyse a significant event (the construct) might be expected to improve with training. Thus, an instrument to measure the ability of a doctor to analyse a significant event is more likely to be construct valid if it shows better results in those who have had training in this process. However, some researchers now view validity testing as a unitary concept best represented by construct validity (Messick, 1980; Downing, 2003). Evidence for construct validity can be gathered from five distinct sources:

- *Content* (e.g. item content and test representativeness)
- *Response process* (e.g. understandable interpretation of scores for subjects)
- *Internal structure* (e.g. standard errors of measurement such as psychometric testing)
- *Relationship to other variables* (e.g. convergent or divergent correlation)
- *Consequences* (e.g. impact of scores on participants).

The type and purpose of the assessment will determine whether some or all the validity sources are required. For instance, the representativeness of the items being measured in an assessment instrument for SEA would intuitively seem important, as would the implications for the participant's future learning. However, for the purpose of formative assessment, the relationship of the proposed instrument with other instruments (variables) to establish convergence or divergence may be less important. For example, the relationship between doctors' ability to undertake SEA and their ability to carry out patient-centred consultations would seem less of a priority.

Reliability testing

In the psychometric tradition, reliability does not relate only to the agreement, variability or consistency of scores achieved by a measurement instrument, though these are all important. Instead it relates to the ability of the assessment tool to consistently differentiate between the subjects of interest (Cronbach *et al*., 1972; Shavelson and Webb, 1991; Streiner and Norman, 2003)

For the purpose of the SEA peer review assessment instrument we would wish that for any given report assessed using this instrument, the same "score" will be given, independent of the date the report is assessed by the same rater (intra-rater reliability) and of the reviewer who assessed the report (inter-rater reliability). However, the information gained from using what we would traditionally call a "reliable" instrument for the assessment of reports can be limited. For example, we may find that 10 GPs who each submitted a single report gained a "satisfactory" assessment from all five reviewers. It is also possible that the reviewers assessed the same 10 submissions on a further occasion and gave exactly the same assessment. This is an example of "perfect" reliability, in that all assessors agreed with each other (perfect inter-rater

reliability), and each assessor awarded the same score on two different occasions (perfect intra-rater reliability). It does not give any information on whether there is a difference in the quality of the reports.

Most assessment will not have perfect 'reliability' because of error in measurement. Two broad areas of error are: firstly, non-random error, where a result is systematically different from the true score - similar to having a thermometer that gives temperature readings 1 degree Celsius higher than the true temperature; and secondly, random measurement error. For example, even though a sphygmanometer has been calibrated, it is unlikely that a group of doctors will give exactly the same measurement of that person's blood pressure and the 'true' value of patient's blood pressure will never be precisely known.

To be able to give useful information on the error within an instrument we need to compare data on the measurement error with the total variation in the achievement level of the individuals undergoing the assessment. Therefore, to achieve high reliability for an SEA assessment instrument, when reporting back the reviewers' assessment, it is desirable that the total error in that given score is small in relation to the total range of scores.

A formula to express the reliability of an instrument, which takes account of the total variability between subjects (in this case SEA reports) and measurement error, has been constructed (Cronbach *et al.*, 1972). It is expressed using the equation:

Reliability = subject variability/ (subject variability + measurement error)

A "perfect" reliability ratio would therefore equal 1 as there would be no measurement error and all variability is due to true differences between the subjects (in this case SEA reports). Conversely, where all the reports were of equal merit, any difference in scores would be due to measurement error and the reliability would be zero.

The feasibility and acceptability of assessment instruments and models.

The feasibility and acceptability of measurement instruments would seem relatively straightforward aspects of an instrument's utility. The feasibility of any measurement instrument will be affected by issues such as the time and cost involved for participants, reviewers and administrators. Factors include the training of reviewers and, for example in simulated surgery assessments, the recruitment and training of actors. Technical issues, such as the setting up of video equipment and the availability of accommodation will all influence the feasibility of an assessment.

It could clearly be argued that the existing peer review model for SEA in the west of Scotland deanery had established its feasibility since it had been running successfully, on a voluntary basis since 1998 and had increasing levels of participation by GPs within the region. Issues around the financial and staffing components of the model had previously been addressed by the director of postgraduate general practice. However, if the model were to be compulsory, then feasibility issues around recruitment and training of assessors, the volume of administrative staff required and the ability to deliver feedback within an appropriate timescale would require re-evaluation.

The acceptability of assessment instruments will be tied in partly with these feasibility issues, but also with the purpose of the assessment and the outcome from the assessment. The acceptability of peer review instruments to both the participants and

those peers completing the instrument can vary and is difficult to generalise (Norcini, 2003). Published measures of the acceptability of an instrument can vary. Some studies report as measures of acceptability the response rates for those asked to complete an assessment questionnaire and the levels of missing data in those replies (Campbell *et al*., 2008). Other studies use individual interviews, focus groups or questionnaires to seek participants' or reviewers' opinions on the acceptability of assessment instruments and peer review models (Bowie 2005a, Dowling *et al*., 2007). It could be argued that those participants in the woS peer review model who submit material on several different occasions are by implication providing evidence of its acceptability, since they would be unlikely to repeat the voluntary process were this not the case.

The educational impact of assessment instruments.

The educational impact of an assessment, which has also been described as the 'consequential validity' (Messick, 1995), reflects the extent to which an instrument has an effect on learning and/or behaviour (i.e. whether an educational gain has been made). It is known that assessment is a stimulus to learning (Newble and Jaeger, 1983; Crossly *et al*., 2002a) but how much learning activity is stimulated by individual assessment instruments is largely unknown. The educational impact of formative assessment is likely to be affected by issues on the quality of feedback, discussed in Chapter 3, who gives the feedback and the 'stage' of the adult learning cycle that the learner is currently at.

Issues to be considered in the relationship between assessment and learning were highlighted by Ringsted *et al*. (2005). Among them were how to:

- *achieve congruence between educational objectives and assessment*
- *increase feedback from assessment*

- *sustain formative feedback*
- *balance formative and summative assessment*.

In a study aimed at assessing the educational impact of a tool for the formative assessment of GP registrars' consultation skills in south east Ireland, researchers found that participants regarded the educational value as 'positive' (Dowling *et al*., 2007). They learned about their own consultation style and reported improving their performance through knowledge of their strengths and weakness. These findings were similar to the experience of registrars in Leicester who found the formative assessment of their consultation skills educationally beneficial (McKinley *et al*., 2000).

'High stakes' assessments have also been described as having a positive educational impact (Lough *et al*., 1995). GP registrars who had undertaken a criterion audit as part of their GP summative assessment portfolio reported increased confidence in undertaking and implementing change as a result of undertaking the exercise within their training practice. Feedback of the assessment results for substandard projects demonstrated an example of how learning could be derived from a 'summative' process since the revised projects all passed a subsequent assessment. Bowie *et al*. (2002) aimed to assess the impact on GP non-principals of the summative assessment audit project, which at the time was one requirement for GPs to become independent practitioners. The authors found that those GPs who had successfully undertaken an audit project had a significantly better knowledge of audit method than those GPs who had not been required to undertake a criterion audit as part of their training.

The relevance of the assessment to the participant can also contribute to its perceived value. In a study to evaluate an in-training assessment programme in anaesthesia,

Ringsted (2004) described that trainees valued assessments they learned from even if they did not pass these assessments. In contrast, feedback from patients on their communication skills were found to be meaningless because they were universally positive and so had no effect on learning.

5.3 Statistical analysis and interpretation

5.3.1 *Introduction*

In general terms, quantitative research methods use statistical analysis to interpret and present data in order to inform the association between variables. Statistical approaches can be described as descriptive or inferential. Descriptive statistics refers to *information* on the spread of study data, such as the mean, median, standard deviation and inter-quartile range; while inferential statistics allow 'inferences' or deductions to be made from data. This is usually to test hypotheses or relate findings to the population beyond those who formed the study sample (University of the West of England, 2007).

The following statistical tests and analyses are discussed and explained because they were applied to the data analysed in Chapters 6, 7 and 12.

5.3.2 *Significance testing.*

While the acceptance or rejection of a hypothesis can enhance scientific theories and understanding, it is necessary to consider that although data analysis may strongly infer a relation or effect between variables, it is still possible that a difference (or the finding of no difference) has arisen by chance (Cook *et al*., 2004). For example, a study could be set up to investigate whether GPs trained in criterion audit method were more likely to complete an audit cycle. The null hypothesis would be that there is no association

between the training and the completion of an audit cycle. At the conclusion of the study, the authors may reject the null hypothesis and report that there is an association between these two variables. The *probability* of observing this finding when it is false (a type 1 error), is expressed by a 'P' value. A P value of <0.05 implies a less than one in twenty chance that the association is incorrect.

Alternatively, an association may be rejected when it is in fact true. The probability of not committing a type 2 error is described as the power of a hypothesis test. The sample sizes of study populations need to be large enough to detect any difference that the researcher is enquiring into. The number of participants required will be dependent on predictions of the magnitude of the difference to be detected and the variability within the sample (Norman and Eva, 2007). This information may be based on information from a pilot study, from previous research or may simply be estimation.

In the study described in Chapter 7, the numbers of participants required for the study population was based on a pilot project that informed the variability within the 'expert' group. The magnitude of the difference thought to be important between groups in grading the severity of a significant event was decided by the research team as an educated 'estimate'. The smaller the difference to be detected, the larger the population required. The power of a study is expressed numerically as 1 minus the probability of a type 2 error. The most commonly accepted chance of a type 2 error is 0.2 (20%), giving a power of 0.8 (80%) (Cook *et al*., 2004).

Although a study may be adequately powered, it is important to recognise that the magnitude of the effect needs to be considered (Streiner and Norman, 2003). A study of an intervention - for example a new antihypertensive drug treatment - may be

adequately powered but the clinical significance of a drop in systolic pressure of four mmHg is at best marginal.

5.3.3 *Chi squared test.*

A common test that compares two proportions is the chi-squared (χ^2) test. It is suitable for drawing comparisons between two groups of categorical data. The test determines whether there is an association between the data from one of two groups (e.g. patients who have had a myocardial infarct (MI)) and those who have a particular characteristic (e.g. smoking status). Information on the probability of any association having occurred by chance (the P value) will be given. It is a significance test only and so will not give information on size effect.

5.3.4 *Confidence Intervals*

The confidence intervals give a more precise and accurate estimate of a difference between two data proportions than the simple acceptance or rejection of a null hypothesis. Conventionally, such a difference is given as a numerical value within a specified range - the '95%' confidence intervals. This means that if the 'experiment' were repeated 100 times, then we can be confident that the true value of the finding will lie within the range given on 95 occasions.

5.3.5 *Analysis of Variance (ANOVA)*

Subjects are often grouped together according to a particular characteristic, for example in Chapter 7, by their GP educator status (e.g. GP trainer or GP appraiser). We may wish to examine if there is a difference between these two groups, for instance in their mean score in grading the importance of a significant event scenario. This could be performed by applying a two sample t test and the result expressed as a probability. If

there was a difference in mean scores between the two groups, this may be considered as 'significant' at a probability of less than five percent ($p<0.05$). If we wish to compare the mean score in the grading of scenarios of several groups (e.g. trainers, appraisers, audit experts and GPs who are not trainers), then this could be examined using a succession of t tests. This, however, presents a problem. If a significance level is set at 5%, one in 20 comparisons will be expected to be 'significant' by chance. As we carry out multiple tests, the chance of a false significant result increases with each comparison (Peacock and Kerry, 2007). ANOVA allows us to make multiple comparisons by using statistical corrections to allow for the multiple testing. The null hypothesis in the above example would indicate that there is no difference in the mean grading score between the various groups. If results indicated that the null hypothesis should be rejected, then further investigative statistical techniques (multiple comparison tests) or descriptive statistics are required since the investigator will only know that the samples with the highest and lowest means are different (University of Leicester, 2009).

5.3.6 *Generalisability co-efficient and D studies*.

As described previously, the 'classical' approach to measuring the reliability of assessment instruments has involved examining areas where 'errors' could threaten reliability, for example, intra-observer variation (good day, bad day), and inter-observer variation (hawks and doves). These tests require to be carried out one-by-one and do not allow comparison of the contribution of each of these sources to the overall reliability of the assessment measure.

To overcome this time-consuming task, Cronbach *et al*. (1972) proposed the use of generalisability theory. It was developed to provide a flexible framework for examining the dependability of behavioural measurements (Shavelson and Webb, 1991). It

described the extent to which assessment scores assigned to individual subjects (in this thesis, the ‘score’ given to an SEA report) could be generalised to the score assigned in a different context (by another reviewer or at a different time).

Underlying this theory is the assumption that any ‘score’ can be split into two components: a ‘true score’ and an ‘error’ score. Failure to identify the value of a ‘true–score’ can be due to a variety of reasons, such as intra- and inter-observer variation, case specificity, different items (internal consistency), and interactions between the learner and the observer. While it is possible to investigate all the different sources of error that may lead to deviation from the true score by embarking on multiple studies, ’G’ studies allow for the various possible sources of error to be combined into a single research design (Crossley *et al*., 2002b; Streiner and Norman, 2003).

Generalisability is a statistical method within the family of regression analysis - techniques that model and quantify relationships between variables to allow predictions to be made (Cronbach, 1972; Crossley *et al*., 2007). The theory uses ANOVA procedures and variance component estimation to enable a coefficient - the ‘G coefficient’ - to be constructed for each of the variables (called facets) under study. In addition, an overall ‘G’ coefficient for the measurement instrument will be estimated. This coefficient is a measure of the error variance that can be attributed to each facet and will be quantified as 0 (poor reliability) and 1 (perfect reliability). By identifying the value attributable to each component in adding error to the measurement, strategies can be developed to improve those aspects that compromise reliability (Downing, 2004).

5.4 Qualitative methods

5.4.1 *Introduction*

'Qualitative' methodology is an umbrella term applied to a group of research designs. Different qualitative methods are applied dependent on the theoretical and practical considerations of the study (Kuper *et al*., 2008a). Qualitative research has its origins in social sciences and humanities (for example, in psychology and sociology), where the theoretical approach used by researchers aimed to provide conceptual understandings of social interactions, societal 'norms' and relational phenomena (Reeves *et al.,* 2008).

The approach to undertaking qualitative research can vary at different levels including the 'systems of enquiry' available for data collection methods. These are usually applied through observations, interviews (either individual or focus group), or analyses of documents and other written materials (Denzin and Lincoln, 2005). Although observational data give researchers access to what people do rather than simply what they say they do, none of the studies in this thesis involved this method. Three of the studies did, however, involve either interviews or document analyses.

5.4.2 *Individual in-depth interview method.*

Individual in-depth interviews are widely used in general practice research to examine GPs' perceptions and experiences of healthcare delivery. Interviews are usually described as structured, semi-structured or unstructured

Structured interviews tend to be undertaken where there is an expectation of quantitative data by the researcher. Unstructured accounts are most commonly described in ethnographic studies where data involve not only conversational data but also observational studies (Cresswell, 1998a).

Semi-structured interviews are a widely-used format in qualitative research. The discussion is guided by a set of 'open' predetermined questions. The primary research question will always form a part of the discussion with further questions usually raised to enquire into various aspects of the research topic. These questions are not fixed, in order that both the interviewer and participant(s) can pursue additional relevant issues as they emerge (Dicicco-Bloom and Crabtree, 2006). The standard interview technique is the 'depth interview' that aims to contribute abundant information on the research topic. This will generally last between 45 minutes and several hours.

Qualitative interviews are usually audio-taped with consent and then transcribed as soon as practicable. This is because a common technique in data gathering is the iterative process, where initial data analysis is carried out prior to the next interview or set of discussions. This allows the interview schedules or topic guides being used by the researcher to be adapted. Subsequent interviews can then explore new issues raised by the previous participants. Equally, questions that do not elicit a useful or informative response can be adapted or dropped.

Data collection can also be undertaken using several different methods or perspectives within the same study. For instance, a study may use different methods such as observations, one-to-one interviews and focus groups to increase the level of understanding and insight into a research question. Alternatively, a study may use different contexts to examine a phenomenon; for example, interviewing some patients who have suffered a myocardial infarct in the hospital wards and others in their home (Mays and Pope, 2000). These processes are known as 'triangulation' and can be used as a form of validation in qualitative research.

For any qualitative study a crucial concern is whom to sample. The purpose of in-depth interviews is to explore and highlight either shared understandings or differences within a particular group or between groups. Therefore, those chosen to participate are required to have some significant similarities related to the research question (Britten, 1995). Qualitative studies do not have predetermined sample sizes. Nor is it usually appropriate to have a sampling strategy based on random sampling as this is unlikely to produce the diverse range of opinions and experiences necessary for qualitative research (Barbour, 2005). It is more suitable to use 'purposive' sampling (also called 'theoretical' sampling) as this allows the researcher to consider what groups of potential participants are likely to yield the most appropriate range of views, either from existing research or a theoretical construct.

The role of the interviewer in qualitative studies is also vitally important. It is advantageous for the interviewer in a one-to-one situation or the moderator in a group to develop a positive relationship during the course of the meeting. Establishing rapport is described as an essential component of an interview (Dicicco-Bloom and Crabtree, 2006). Rapport involves gaining the "*trust and respect for the interviewee and the information he or she shares*."

As discussed previously, data analysis in qualitative research will often occur concurrently with data collection to inform both sampling and the research questions being asked of participants. Data analysis can be conducted individually or as part of a team that analyses data as a group (Silverman, 2005). Where individuals analyse data separately, they would normally meet to compare and discuss results of individuals' work. This iterative process of analysis eventually leads to a situation where no new

themes or categories emerge in relation to the research question. This is referred to as ‘saturation’ and implies that data collection is complete (Cresswell, 1998b).

Respondent validation is frequently described as a requirement for qualitative research (Barbour, 2001). This is also referred to as ‘member checking’ and involves the researcher asking participants if some or all of the study findings are in are in accord with their experience (Kuper *et al*., 2008b). This process can also form part of the validation of the research process.

The interview process is limited because the information given is filtered through the memory of the participant, and is influenced by the social context of the interview (Reeves *et al.,* 2006). Both data collection and analysis can be prejudiced by ‘Hawthorne effect’ (Holden, 2001), where the presence of the interviewer and/or observer may influence participants’ behaviour or responses. In educational settings, hierarchical structures between researchers (who may also be teachers or examiners) and participants (who may be undergraduates or trainees) may influence participants’ motives and actions. As a result, ‘insights obtained’ may create a false impression of a normative situation rather than the ‘sometimes unseemly realities of “behaviour”‘ (Reeves *et al*., 2006)

5.4.3 *Focus groups*

Focus groups have their origin in four separate traditions; social science, organisational research, community development and market research (Barbour, 2005). In medical education research, focus groups have been commonly used to develop items for inclusion in questionnaires or in hypothesis generation (Mcleod *et al*., 2000; Britten, 1995; Patton, 2002). The definition of what constitutes a focus group varies between

reported studies. One definition is “a group discussion exploring a specific set of issues” (Kitzinger and Barbour, 1999).

Focus groups can be attractive as a research method as they provide access to a wide variety of experiences. They are “*useful when it comes to investigating what people think, but they excel at uncovering why participants think as they do*” (Morgan,1987). They provide a forum to explore different opinions, reflection on common practice and participants’ assumptions (Elwyn *et al.*, 1999).

As a research method, Sim (1998) outlined that focus groups can have four broad advantages:

- they provide information on the 'dynamics' of attitudes and opinions in the context of the interaction that occurs between participants.

- they may encourage a greater degree of spontaneity in the expression of views than alternative methods of data collection.

- they can provide a 'safe' forum for the expression of views.

- Participants may feel supported and empowered by a sense of group membership and cohesiveness.

As focus groups are a form of group interview, many of the issues applying to one-to-one interviews also apply to the sampling, data collection, analysis, theory and ethics of these groups. Researchers using this method must decide whether the topic would benefit from the discussion, interaction and comparison between groups. Some

subjects, such as deeply personal issues, may be better explored in a safer and more productive way through one-to-one interviews. Ideally, sampling should aim for enough heterogeneity within the group to stimulate discussion but sufficient homogeneity to facilitate comparison between groups (Barbour, 2005). There is no defined number of subjects who have to present to form a focus group, although numbers of between eight and 12 have been suggested as ideal although studies involving between four and six participants are also acceptable (Sim, 1998). The group is co-ordinated by a moderator or facilitator, who may be assisted by a fellow researcher acting as an observer. This assistant's role will most often be to take notes during discussions to give added information and meaning to the group interaction, non verbal cues and discussion (Kruguer, 1994).

As with one-to-one interviews, focus group studies need to consider venues when planning meetings. Meetings need to be confidential and undertaken in a venue with appropriate facilities to ensure adequate recording of data. Although there is "*no such thing as a neutral or ideal location*", the venue(s) chosen should aim to maximise participation (Kitzinger and Barbour, 1999).

Analysis of data should recognise that focus groups can overemphasise consensus and can be dominated by either influential or opinionated group members (Lingard and Kennedy, 2007). In addition, analyses should include the dynamic interactions within the group. This is particularly important as it is problematic to generalise from a focus group. This is partly because participants, as previously described, are usually selected through a process of non-random sampling and, in addition, there may be a tendency for more self-confident and articulate individuals to agree to take part in a focus group (Morgan, 1995). Moreover, as focus groups are contextualised, it cannot be assumed

that the information given by a subject in that group is a predictor of what they may say or do in a different social situation (Sim, 1998).

5.4.4 *Content analysis.*

Content analysis has been defined as a research technique for making replicable and valid inferences from texts (Krippendorff, 2004). It was developed in the twentieth century through its use in the social sciences, whereby researchers analysed and interpreted the press and mass media coverage of economic and social issues of the early 1920s and 1930s. The technique was also employed in efforts to extract information from propaganda in World War 2 (Krippendorff, 2004).

In medical education, content analysis is a data interrogation technique that involves the systematic examination of text by identifying and grouping codes, classifying and developing categories and themes (Pope *et al*., 1995). It can be viewed as a subset of thematic analysis, where data are organised according to topics, ideas or concepts (known as themes). The content can be analysed on two levels; firstly, a descriptive account of the data with nothing read into or assumed from the data (manifest analysis); and secondly, interpretive analysis, which concerns what was meant by a response or what was inferred (latent analysis) (Hancock, 2002).

The procedure for content analysis involves a series of steps (Hancock, 2002):

1. *The transcript is read through and areas of interest or relevant information are 'coded'.*

2. *As further related examples of interest (codes) are identified, the data is sorted and organised into categories that represent similar trends.*

3. *Identify whether the categories can be linked in some way. If so, these can be listed as major categories and the original smaller categories as minor categories*.

4. *Compare and contrast the various categories again to see if adjustments are required or whether a code fits into several categories.*

5. *Repeat steps 1 to 4 with subsequent transcripts, identifying new categories of information until saturation is achieved.*

6. *Re-examine the data to ensure that it is in the appropriate minor or major category.*

7. *Examine the categories to see if two or more fit together or are similar in concept. If so, these will form a theme from the research.*

8. *Interrogate the original text once more to consider whether any previously excluded data is relevant and should be included in the results.*

5.5 Conclusion

A series of eight mixed method studies are reported in this thesis. They were conducted over the five-year period 2002-2007. The justification for each of the methods used is discussed in the appropriate chapters.

Two quantitative studies that examined issues in the identification, analysis and reporting of significant events, using cross-sectional postal questionnaire surveys were carried out. Three further quantitative studies to examine aspects of the validity and reliability of a peer review instrument are described.

Two studies using interview methods are reported, one using a focus group method and the other using semi-structured in-depth interviews. This research examined factors affecting the feasibility, acceptability and educational impact of the west of Scotland model of SEA peer review. A final study involved the content analysis of SEA reports

submitted to the west of Scotland peer review model. This study aimed to identify the types and causes of significant events chosen for peer review and how the learning and change reported through the analysis of these events reportedly impacted on the quality and safety of patient care.

CHAPTER 6

Attitudes to the identification and reporting of significant events by general practice principals.

6.1 Introduction

At the time of this study in 2002/3, the topical and inter-linked issues of patient safety and adverse incident reporting were becoming increasingly recognised as potentially valuable components of the clinical governance agenda (DoH, 2000; Pringle, 2001). However, it was clear that research was lacking in these important areas, especially in general practice, where little was known about the commonest forms of significant or adverse events, the exact nature of their underlying causes, and how these are investigated and resolved by healthcare professionals (Chapter 1). As previously discussed, the profusion of terminology associated with quality and safety-related incidents caused confusion and misunderstanding around significant events at this time.

The link between the routine use of significant event analysis (SEA) and the potential reporting of serious significant incidents by healthcare professionals had been described Harrison *et al.*, (2002). However, it was clear that much has still to be learned about how these types of events might improve the safety and the care of patients. In particular, how effective systems could be implemented to successfully identify, analyse and share this information between GPs and their teams was unknown.

The formation of the NPSA in England and Wales was one attempt to address this by putting in place a national system to co-ordinate the confidential and mandatory reporting of adverse incidents by healthcare professionals (Mayor, 2001). However, the

previously described barriers potentially preventing NHS staff from fully participating in such a reporting system presented a challenge in engaging GPs and their practices with reporting systems. Additionally, there was a perception amongst some healthcare practitioners that linked personal involvement in an adverse incident with the professional competence (or incompetence) of that practitioner (Cohen, 2000). The personal attitudes of healthcare staff and independent contractors towards reporting appropriate incidents would therefore be a key factor in overcoming these barriers. Moreover, an understanding of these attitudes could be used to help develop a reporting and learning system and the creation of a safety culture in the NHS.

As part of a wider study into significant events and their analysis, a survey of principals in general practice was undertaken to determine their attitudes towards identifying and dealing with significant events and their subsequent preparedness to participate in external reporting of adverse incidents. In particular, the study aimed to explore any differences in attitudes between principals from training and non-training general practices as it was thought that training practice GPs were more likely to have been involved in the identification and analysis of SEA (Bowie *et al.*, 2003)

The main hypothesis was that principals from training practices would express more positive attitudes than non-training colleagues because of their required involvement in significant event analysis as part of a long-established regional audit programme in the west of Scotland deanery. At this time, neither the nGMS contract nor GP appraisal had been introduced, and so SEA was not a 'mainstream' activity in general medical practice. The study also aimed to determine if length of service as a principal in general practice (as a proxy for experience) affected attitudes towards identifying significant events and the reporting of adverse incidents.

6.2 Methods

6.2.1 *Study design and participants*

The study was restricted to principals in general practice, rather than a mix of general practice and primary care based staff, because of the availability and easy access to an accurate and up-to-date database with postal details of these doctors.

A cross-sectional survey of all principals in general practice in the Greater Glasgow area was completed in 2002. Principals were identified from the deanery's database, which was developed to track the Postgraduate Educational Allowance (PGEA) activities of relevant general practitioners (GPs) in the west of Scotland. They were sent a self-completion postal questionnaire; the parts of which that relate to this study are given in Appendix 1; a covering letter outlining the aim of the study is given in Appendix 2. Non-respondents to the initial survey were mailed two subsequent reminders at two-weekly intervals.

The questionnaire was devised by the authors and pre-tested amongst five departmental colleagues who are part-time general practitioners. A pilot survey involving 12 principals in general practice based in the Eastern Glasgow Local Health Care Co-operative (LHCC) was undertaken. Minor amendments were made to clarify the wording of some of the questions based on the feedback received from participants in the pilot.

6.2.2 *Definition of a significant event*

The definition of what constitutes a significant event (Box 4) was outlined on the front page of the questionnaire to assist those participants unfamiliar with the term. As described in Chapter 1 the breadth of this definition provided a 'catch-all' as it clearly

covered the numerous and often confusing terms associated with risk management issues and incident reporting systems at this time.

Box 4: Definition of a significant event

> "..*any event thought by any member of the practice team to be significant in the care of patients or the conduct of the practice*.." (Pringle *et al*., 1995)

6.2.3 *Attitudinal statements*

Respondents were asked whether they agreed or disagreed with five statements about the reporting of adverse incidents and identifying and acting on significant events. The statements used were based on a mix of frequently recurring comments made by general practitioners at previous educational medical meetings on significant event analysis. A four-point Likert scale was used to indicate the strength of agreement with four of the different statements (from strongly agree to strongly disagree). A 'don't know' option was added to the Likert scale as the nature of one question allowed for a 'don't know' response.

6.2.4 *Data analysis and statistical tests*

Data were analysed using Minitab Version 13.2 software. Data are presented to indicate strength of agreement with each statement. This was achieved by combining the 'agree' and 'strongly agree' options on the response scale. Confidence intervals (CI) were calculated to quantify the differences in proportions agreeing to each statement by training practice and non-training practice participants. Relationships between attitude and length of time as a principal were examined using a Chi Squared Test (χ^2).

6.3 Results

6.3.1 *Response rate*

A total of 617 principals in general practice were surveyed and 466 responses (76%) were received, with 162 responses (35%) from training practice principals and 304 (65%) from non-training practice principals. A disproportionately greater number of responses were received from principals attached to training practices (162/186, 87%) than those from non-training practices (304/431, 71%). The main survey results are outlined in Tables 1 and 2.

6.3.2 *Type of practice*

Significantly more non-training practice principals agreed that the reporting of adverse incidents should be mandatory (21% v 13%, p=0.04). However, there was no evidence of any difference in attitudes between principals in training and non-training practices in agreeing that they would be selective in reporting incidents if a mandatory system was introduced, nor in their willingness to take part in a local anonymised system of reporting.

More principals in non-training practices than training practices agreed that it can be difficult to determine when an event is significant (p=0.03) and also that significant events are often not acted upon in their practices (p=0.01).

6.3.3 *Length of service*

Individuals were classified into three groups according to length of time as a principal – those who had worked five years or less, those working between six and 15 years, and those with more than 15 years of service. There was no evidence that length of time as a principal was associated with a difference in attitude to the mandatory reporting of

adverse incidents, nor with being selective in reporting incidents in a mandatory system. Similarly, length of time as a principal was not a factor in agreeing to take part in a local anonymised system of reporting.

A large minority of all principals (41%) agreed that determining when an event is significant can be difficult and also that significant events are often not acted upon in their practice (30%). There was, however, an association between length of time as a principal and agreement that determining when an event is significant can be difficult ($p=0.01$), with those of five years or less experience more likely to have difficulty in determining when an event is significant. Those principals with 15 years or more experience were much less likely than expected to agree that significant events identified in practice were not acted upon, compared with the other age groups ($p<0.001$).

6.4 Discussion

6.4.1 *Main findings*

The survey elicited a very satisfactory response rate (McColl and Thomas, 2000) from all principals in general practice in Greater Glasgow, although there was a disproportionately greater return from principals in training practices, which may introduce an element of bias to the reported results. It can only be speculated that this may have been as a result of their experiences of significant event analysis (SEA) gained as part of the regional audit programme for training practices in place at the time of the study. In common with most survey data, the researchers relied on participants to self-report and have no means of verifying responses. This can potentially limit the validity of the data.

As the identification and reporting of significant events is not intended to be confined to GP principals, it is possible that different attitudes would have been expressed either by GPs who are not principals or by members of the associated primary healthcare team. However, at the time of the survey, GP principals were thought likely to be strongly involved in the decisions on which events should be reported. In addition, if reporting was to become a contractual obligation for a practice, this would be the responsibility of the individual GP principals.

6.4.2 *Findings in Context*

Mandatory reporting

Only a minority of principals agreed that the reporting of adverse incidents should be mandatory. Moreover, many of these respondents also agreed that if mandatory incident reporting were to be introduced, then they would be selective in the types of events they reported. This could lead to under-reporting and would have implications for the validity of the NPSA system proposed at that time in terms of the quantity and quality of events reported. Importantly, it may also have a negative effect on the potential for sharing and learning from adverse incidents on a local or regional basis.

It is possible that the mandatory nature of any reporting system is too threatening for many general practitioners from both training and non-training environments. Although the preferred mechanism for reporting incidents was not investigated, around three-quarters of respondents indicated that they would be agreeable to participating in an anonymised incident reporting system. However, there is still the possibility that many doctors may be sceptical about certain reportable issues remaining anonymous to others (Ness and Cordess, 2002), whilst assurances of complete confidentiality might

not be enough to convince them to participate fully in a reporting system (Runciman *et al*., 2001)

Identifying and acting on significant events

A substantial minority of respondents from all practices, but particularly those in non-training, agreed that it is difficult to determine when an event is 'significant'. This may indicate a lack of understanding, inexperience, or an educational issue in relation to the identification and analyses of significant events. Previous research had already identified an educational need in this particular area (Bowie *et al*., 2003; McKay *et al*., 2003). A separate sub-analysis of the questionnaire used in this study also found a small group of GPs who were unaware of a recent significant event (Bowie *et al*., 2004). In addition, this may also reflect a difference in the perception of 'significance' between individual practitioners.

It is noteworthy that over one-third of principals in training practices agreed that there was a difficulty with the issue of 'significance', despite these practices routinely identifying and analysing significant events as part of the long-running regional audit programme alluded to earlier. Participation in this audit programme at this time was a conditional requirement for training practice status and documentary evidence was periodically checked and verified at a practice visit. However, the study findings suggest that principals from training practices are more likely to act on significant events once identified than colleagues from the non-training environment. Experience of SEA gained through the regional audit programme or other educational initiatives may, therefore, have been a factor in this difference. It was also known that training practices were more likely to participate in conventional 'internal audit' (Baker, 1985; Baker and Thompson, 1995) and they may thus be more willing to reflect and, where appropriate,

act on a significant event. Other variables, for example, team dynamics, practice structure and the type of event identified may also have an influence.

Voluntary reporting

Voluntary and anonymous reporting are acknowledged as important elements of a successful adverse incident reporting system (Cohen, 2000; Runcimen *et al*.,2001). It is clear from the study that respondents may prefer this approach to the proposed confidential but mandatory system. As discussed in Chapter 1 a number of concerns have been made about mandatory reporting systems. Perhaps the most cogent argument from an independent contractor perspective is that it implies that the individual practitioner is ultimately responsible and must therefore report the event in question, regardless of level of involvement. This is despite the evidence that analyses of serious errors often highlight multiple systems failures and the involvement of many individuals. The implication is that failure to participate in a mandatory system will lead to some form of punitive action being exacted on the general practitioner, whether this is accurate or not.

Length of service and attitudes

The reasons why less experienced principals reported difficulty with determining when an event is significant can only be speculated upon. For example, they may have a higher or a lower threshold for identifying significant events compared with more experienced principals, based on their limited experience of general practice and knowledge of significant events.

The finding that significant events are more likely to be acted upon by experienced principals with 15 years or more of service may reflect their management ability and

confidence in implementing change, or may be related to the type of significant event identified with less complex issues being raised.

6.5 Conclusions

It was apparent that the then nascent NPSA and any similar systems being proposed for Scotland and Northern Ireland faced a major challenge in getting many principals in general practice to participate in a national reporting system if the attitudes highlighted in the study were generalisable throughout the UK. More detailed research was required into why principals had shown such negative attitudes towards the mandatory reporting of adverse incidents, given the importance of the latter in potentially reducing risk to patients and enhancing patient safety. Similarly, the related concept of 'significance' and why significant events are often not acted upon also required elucidation and is studied further in Chapter 7.

Table 1: Principals in general practice: level of agreement with attitudinal statements

Attitudinal Statements	All Principals	Training Practice Principals	Non-Training Practice Principals	Difference in Proportions (95% Confidence Intervals)	P value
	n (%)	n (%)	n (%)		
The reporting of adverse incidents that happen in general practice should be mandatory	81/444 (18)	21/156 (13)	60/288 (21)	7.4% (0.3 to 14.5)	0.04
If a mandatory incident reporting system was introduced, I would be selective in what I reported	317/433 (73)	117/152 (77)	200/281 (71)	5.8% (-2.7 to 14.3)	0.18
I would be willing to take part in a local anonymised reporting system	332/444 (75)	122/154 (79)	210/290 (72)	6.4% (-1.4 to 15.0)	0.10
Determining when an event is 'significant' can be difficult	189/461 (41)	56/162 (35)	133/299 (44)	9.9% (0.7 to 19.1)	0.03
Significant events are often not acted upon in my practice	135/456 (29)	35/161 (22)	99/295 (34)	11.9% (3.5 to 20.2)	0.01

Table 2: Principals in general practice: proportion in agreement with attitudinal statements and length of time as principal

Attitudinal Statements	Principals in Agreement	≤5 years Service	6 – 15 years Service	≥15 years Service	P value
	n (%)	n (%)	n (%)	n (%)	
The reporting of adverse incidents that happen in general practice should be mandatory	81 (18)	18/77 (23)	30/195 (15)	33/168 (20)	P=0.27
If a mandatory incident reporting system was introduced, I would be selective in what I reported	315 (73)	47/74 (64)	139/185 (75)	129/170 (76)	P=0.10
I would be willing to take part in a local anonymised reporting system	330 (75)	63/78 (81)	146/194 (75)	121/168 (72)	P=0.34
Determining when an event is 'significant' can be difficult	187 (41)	44/80 (55)	69/199 (35)	74/178 (42)	P=0.01
Significant events are often not acted upon in my practice	135 (30)	24/79 (30)	72/196 (37)	37/177 (21)	P<0.001

CHAPTER 7

Levels of agreement on the grading, analysis and reporting of significant events by general practitioners.

7.1 Introduction and Aims

The previous study and other associated research highlighted concerns with some of the barriers to the identification, analysis and reporting of significant events in general practice (Harrison *et al.*, 2002; McKay *et al.*, 2003; Bowie *et al.*, 2004). As described in Chapter 1 by 2005, as part of its overall strategy to improve learning from safety incidents across the NHS, the NPSA had recommended that primary care teams should identify, prioritise and analyse 'significant' events (NPSA, 2005). This recommendation stated that a significant event analysis should be undertaken by the appropriate primary care team if a 'safety incident' had caused 'minor' or 'moderate' harm to a patient or had the potential to do so.

To inform a consistent approach to both SEA and incident reporting, greater evidence was still needed on whether broad agreement could be achieved on what actually constituted a 'significant' event and how likely GPs were to prioritise specific events for formal, structured analysis. Despite increasing engagement in SEA across UK primary care, due to contractual and clinical governance specifications, knowledge of the willingness of GPs to formally report relevant significant events as part of reporting and learning systems was still limited. This is particularly important because general practice teams have yet to engage in the process to the same extent as other clinical specialties (NPSA, 2006). A level of consistency in these areas is desired so that learning opportunities from quality and safety-related incidents are maximised and those events notified to reporting systems will be sufficiently relevant (DoH, 2001b).

Against this background, this study aimed to compare agreement on the grading severity given to a range of significant event scenarios by different professional groups of GPs. The null hypothesis was that there is no difference between the different professional groups in their grading of significant events. The study also ascertained their willingness to formally analyse and confidentially report each event. Opinions on which members of the practice team GPs would involve in the event analyses and whether other GP teams would benefit from this knowledge were also sought. Finally, GPs were provided with a wide ranging list of UK health care bodies and asked to choose their preferences for formal reporting of each significant event scenario assuming that these bodies have the facility to collect reports.

7.2 Methods

7.2.1 *Study Participants*

All GP principals in Greater Glasgow and Clyde NHS Board area were identified from the previously described organisational database which is used to track continuing professional development activity. GPs were categorised into the following groupings: GP trainers (Trainers), GP principals based in training practices excluding trainers (TPNT), GP principals based in non-training practices (NTP), and GP appraisers (Appraisers).

A further group of 20 GPs was included in the study (the west of Scotland Audit Development Group [ADG]). They are trained and highly experienced in peer assessing SEA reports and providing developmental feedback to colleagues. The group was accorded "expert" status for the purpose of the study.

7.2.2 *Power calculation*

A pilot study using 10 "experts" from the ADG to grade seven significant event scenarios for 'perceived significance' using a 1-7 rating scale (1 = no significance, 7 = extremely significant) suggested that the standard deviation (SD) for grading each scenario is 1.0 (range 0.65 to 1.45). A pragmatic decision was taken that a difference of 2 on average could raise an educational issue while a difference of less than 1 would not be considered important. Assuming the SD is 1 and a significance level of 0.01 (to take account of multiple testing,) then 20 GPs in each group would have 67% power to detect a difference of 1 on average between groups and 85% power to detect a difference of 1.2. If the SD was as large as 1.5, 20 in each GP group would have a 92% power to detect a difference of 2 and 67% power to detect a difference of 1.5 at the 0.01 significance level.

7.2.3 *Sampling of GPs in NHS Greater Glasgow & Clyde*

A random sample of 20 doctors from each of the identified groups was recruited by letter of invitation. Non-responders were sent reminders after three and six weeks. The process was repeated until the study minimum of 20 GPs or more per group was recruited.

7.2.4 *Significant event scenarios*

Significant event scenarios were identified for use in the study by:

- Identifying and reviewing the literature on common 'significant' events and/or 'errors' in international family practice (Fischer *et al.*,1997; Bhasale, 1998; Makeham *et al.*, 2002; Dovey *et al.*, 2002; Rubin *et al.*, 2003)
- Merging this information with further evidence from 181 categorised significant events analysed by GPs collected as a part of (unpublished) research involving the

NES west regional educational peer review model to develop a preliminary taxonomy of significant events in general medical practice.

Ten significant event scenarios were originally identified by the authors as being representative of this preliminary taxonomy. Each was graded for 'significance' (using the previously described seven-point rating scale) by 10 GPs from the ADG. Seven diverse events were then chosen because their rating scores were indicative of a range of 'significance' severity. Following this, these events were given a severity grading by this author and a co-researcher using the NPSA grading categorisation (Box 5). Most event scenarios chosen impacted 'negatively' on care because GPs may favour analysing these over 'positive' events, such as early diagnoses of curable cancers or successful CPR in a cardiac arrest (Bowie *et al.*, 2005a)

7.2.5 *Data collection*

A postal questionnaire was developed and piloted initially with five associate advisor colleagues outwith Greater Glasgow and Clyde NHS Health Board area. Minor changes were undertaken to the structure of the questionnaire and the wording of questions. The questionnaire was then piloted with 10 randomly chosen GPs in Greater Glasgow. No modifications to the scenarios or questionnaire were necessary (Appendix 3). The scenarios and questionnaire were sent to the study population with a covering letter (Appendix 4) explaining the study purpose. Respondents were asked to grade the severity of event scenarios on the 7-point rating scale and to indicate if they would undertake a formal analysis of each event. Data were also collected on the face validity of each event scenario together with respondents' willingness to involve team members in an event analysis, formally report each event and, if so, to whom from a diverse list of NHS organisations provided.

7.2.6 Statistical analysis:

Differences in grading severity amongst groups were examined using analysis of variance (ANOVA) and calculated using Minitab software version 13. Differences in the proportions of respondents' reporting preferences for select health care bodies were calculated along with 95% confidence intervals.

7.3 Results

The overall response rate was 125/162 (77%). Response rate by GP group is outlined in Table 3.

7.3.1 *Face validity of scenarios*

The proportion of respondents who agreed that each event scenario could conceivably happen in general practice was as follows: Scenario A - 117 (94%), Scenario B - 122 (98%), Scenario C - 113 (90%), Scenario D - 120 (96%), Scenario E - 119 (95%), Scenario F - 122 (98%), and Scenario G - 114 (91%).

7.3.2 *The grading of event scenarios and willingness to analyse*

Summary data on the grading of significant events and GPs' willingness to analyse these are documented in Table 4. There was no statistical difference amongst the individual groups in their rating of severity of any of the significant event scenarios except for scenario G, where there was a difference between the ADG and Appraisers. Similarly, there was no evidence of difference between groups in their willingness to analyse the significant event scenarios. Overall, GP groups displayed a high level of consistency in grading events and indicating a willingness to analyse them.

7.3.3 *Opinions on sharing knowledge of significant event scenarios*

The levels of agreement indicating that sharing the event scenario with others would be of benefit and the respondents' preparedness to report each scenario as part of an anonymous and confidential system are outlined in Table 5**.** There was no evidence of a statistical difference amongst the groups on whether they would be prepared to report these significant events. Similarly, respondents were in agreement on whether they thought that others in general practice would learn from these events**.** A high degree of consistency was apparent in GPs' willingness to report significant events and the perception that others in general practice would also learn from having knowledge of these events. Scenario C that described a possible situation that did not result in patient harm was thought by all groups to be as worthy of dissemination and reporting as scenario D that did result in patient harm.

7.3.4 *Formal notification of significant events*

Where respondents indicated their preparedness to formally report a particular event scenario, they were then asked to choose their preferences from a range of NHS organisations provided (Table 6). There was no statistical difference between GP groups in their reporting preferences for all event scenarios. NHS Education for Scotland (NES), a special NHS board with responsibility for GP education, was the most frequently chosen body followed by the local medical committee (LMC), to whom all GPs would hypothetically report all of the significant events outlined. Even amongst the top preferences, a clear statistical difference was still apparent between NES and the LMC (Table 7). For the majority of other significant events there was minimal preference by all GPs for local or national reporting bodies.

7.3.5 *Wider practice team involvement in analysing event scenarios*

GPs' willingness to involve other members of the primary care team if an event analysis was undertaken for each significant event is reported in Table 8. The results imply that in the majority of cases, GPs would involve their immediate GP colleagues in the SEA. Those events rated as being most "significant" were most likely to be shared with GP colleagues. In the majority of cases, GPs were selective as to whom they would involve in the event analysis, depending on the nature of the event; i.e., events with a clinical or administrative focus being the major determinant of which other members of staff would be involved. Of particular note is the low level of willingness to involve patients in the analysis of patient-centred events.

7.4 Discussion

A high survey response rate was achieved. The main findings show no significant variation between different professional groups of GPs in grading the severity of each significant event scenario. The greater the 'significance' level that GPs' attributed to each scenario, the more likely they were to indicate that they would undertake a formal analysis and also involve the most appropriate members of the healthcare team in this process.

Where a majority of GPs agreed that others in general practice may learn from a given significant event scenario, they were also prepared to formally report this event as part of a confidential and anonymous NHS reporting and learning system. Paradoxically, when respondents were presented with a range of health-related bodies to which they could potentially choose to report events, neither a national reporting agency nor the local primary care organisation featured as the most popular choices. The clear first choice reporting option for most respondents for almost all event scenarios was the

national body responsible for postgraduate GP education, followed by the local medical committee in most cases. The NPSA reporting and learning system does not extend to NHS Scotland, but health authorities in Scotland look to this organisation for support and guidance. GPs are still expected to report significant events under local NHS reporting arrangements and undertake SEA like their counterparts in the rest of the UK.

7.4.1 *Limitations of the study*

With the exception of the ADG, study participants in each professional grouping were selected at random rather than through stratified sampling to take account of other factors such as college membership, duration of service and size of practice. It is clear that a different sampling strategy may have influenced the outcome of the study. A disproportionately greater number of GP appraisers and GP trainers responded compared with other GP groups, which may indicate respondent bias. GP trainers and appraisers would also be expected to possess some experience as SEA is a core educational activity and one of five categories in GP appraisal in Scotland. As with the previous study (Chapter 6), any cross-sectional questionnaire data collected are self-reported and therefore cannot be independently verified. The event scenarios selected by the authors were chosen based on their notion of 'representativeness', but the representativeness of the breadth of significant events in general practice can only be assumed.

7.4.2 *Context of findings*

<u>Grading of significance and willingness to analyse events</u>

The results demonstrate that study participants were consistent in their perceptions of the "significance" of a range of event scenarios, and in their willingness to analyse these events. This is encouraging, given the difficulty for a substantial minority of GPs

in identifying, prioritising and analysing significant events (Bowie *et al.*, 2004, McKay *et al.*, 2006). However, there is no uniformly agreed definition as to what constitutes a health care 'significant event'. Arguably, even by Pringle *et al.*'s (1995) previously described broad-based definition, one of the 'significant event' scenarios (G) is non-significant as it does not impact on patient care or the conduct of the practice. Equally, there is no defined 'severity level' for a significant event which should prompt the health care team to undertake a formal analysis. All of this is problematic for bodies like the NPSA and local primary care organisations in their attempts to encourage learning and the voluntary reporting of relevant significant events by GPs and their teams.

Although the nGMS contract in the UK requires specific events to be considered for a written analysis, neither the severity of the event nor its potential for learning need influence the ultimate decision to analyse. However, there is an opportunity-cost to focusing on designated events which may or may not have safety and/or educational implications. Given that GPs in this study linked more severe event scenarios with an increased potential for learning, policy-makers may wish to consider focusing attention on constructing a list of common significant events which can or do lead to patient harm in general practice. Potentially, this is one way of initiating and developing a consistent and effective approach to the investigation of common safety incidents in general practice.

At a practice level the findings would indicate that GPs can prioritise which events should be analysed. 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 organizations (Wilson *et al.*, 2001). However, we cannot assume that even if practitioners are

prepared to analyse an event, that they will be able to do so effectively and consistently (McKay *et al.*, 2006). Issues such as team dynamics, lack of education and training, selectivity of events and humanistic and emotional factors can act as barriers to the effective analysis of events (Sweeney *et al.*, 2000).

Educational benefits to others of sharing knowledge of significant events

A key finding was that the greater the severity rating for a significant event, the more likely it was that participants would perceive that other health care teams should be able to learn from it. This would appear to be intuitively appropriate if we relate the severity of an event outcome to the degree of potential learning. However, there is no evidence to suggest that the learning gained from highly significant events is greater than that from lesser significant events. Again, although participants identify that there is learning for others in significant events, what they would actually do in practice may be different. Nevertheless, this finding does indicate that GPs may be capable of discriminating between a wide range of significant event scenarios and prioritising the potential for learning from these events. This may offer some solace to NHS organisations and related agencies about GPs' knowledge of significant events, ability to rate severity and recognition of the wider educational benefits associated with them.

Formal reporting of significant events

It is already know that incident reporting systems in healthcare are subject to high levels of under-reporting (Stanhope *et al.*, 1999). As discussed in Chapter 1, various barriers and difficulties, such as the guarantee of anonymity, potential litigation, the ergonomics of reporting mechanisms, clarity around what should be reported and lack of feedback, have all been shown to impact negatively on formal reporting systems (Vincent *et al.*, 1999; Evans *et al.*, 2006), Equally, reporting systems could be deluged with reports of

significant events because of low thresholds for reporting or misunderstanding or disagreement over the severity of the events which should be reported into a system. GP participants in this study would be prepared to report the more "significant" events to an anonymous and confidential system and most would not report those considered to be less significant. This would imply that GPs, perhaps with the aid of additional guidance, would be prepared to identify and report those significant events which formal reporting systems were designed to attract.

Incident reporting is not a panacea and is only one part of the larger strategy for improving patient safety (Vincent, 2007). The finding that each GP group found the educational implications from an event that did not result in patient harm as salient as one that did highlights that there needs to be more to quality improvement than simple incident reporting within a narrow set of parameters. However, this information is still crucial if we are to gain greater understanding of the scale and nature of error in general practice, develop appropriate solutions and disseminate lessons learned.

The apparent willingness to report specific significant events scenarios should be seen in context because of the reluctance of many participants to report events to certain named organisations. If we temporarily exclude NHS Education for Scotland (NES) as an option, GPs still chose to notify significant events to a professional interest group (local medical committee) ahead of a range of NHS organisations which could potentially host a reporting system. Neither NES nor the local medical committee currently has a reporting system in place for adverse events as this is the responsibility of local health care organisations. The willingness to report significant events to NES may of course be due to bias as this is the body undertaking this study and the employer of the main investigators. Alternatively, we could interpret this finding within

the context of NES being the lead educational organisation in Scotland, and so it may be viewed as a less threatening option compared with the others. Culturally, there may be concerns about the consequences of reporting to bodies that have a clinical governance remit over practices.

Involvement of the wider general practice team in SEA meetings

The findings support previous research on the involvement of other members of the practice team in the discussion and analysis of significant events associated with their practice (Bowie, 2004). Practice managers and receptionists are more likely to be involved in the analysis of an "administrative" significant event (e.g. scenario B) than GP colleagues. Similarly, GP colleagues of respondents are more likely to be involved in analysis of "clinical" significant events (e.g. scenarios A and E).

It is interesting to note that only a small minority of respondents would opt to involve their patients in the analysis of a significant event, even for those scenarios which directly impact on the quality of the patient's clinical care (e.g. scenarios A, C, D, and F). Patient involvement is strongly promoted in all aspects of care and service delivery in the NHS (DoH, 1998). The reason why patients' potential participation in SEA meetings that directly affect them is not high on most GP respondents' agenda is an area for further study. It is known from GPs that although complaints themselves do not precipitate the majority of significant events or their analyses, the linking of SEA with patients' complaints is thought to be good practice. This may provide the patient with credible evidence that the practice is taking the complaint seriously (Bowie *et al*., 2005a).

7.5 Conclusion

The consensus of GPs' perceptions on grading, analysing and reporting significant events in this study may be a positive finding for healthcare authorities and educational bodies. This could provide an opportunity to develop specific guidance on the types of events that GPs should prioritise to maximise learning opportunities and optimise relevant reporting. However, there is still a difficulty to be overcome in engaging GPs in reporting and learning systems. Further exploration of the reasons why GPs appear to prefer an educational body as a potential reporting source or may not be willing to include patients in relevant event analyses is also required. These two issues are therefore explored further in Chapter 12.

Box 5: Significant event scenarios and associated agreed NPSA grading

CODE	Significant Event Scenario	Agreed NPSA Grading
A	A patient was on long-term warfarin therapy as anti-thrombotic prophylaxis (having had insertion of a prosthetic aortic valve). For several years the patient's INR was being monitored on a monthly basis in the surgery. When going through the routine hospital letters, the patient's doctor received notification that the patient had been discharged from hospital having suffered an embolic CVA. The doctor, on reviewing the patient's INRs, had noticed that the most recent INR reading was 2.2 and that the preceding month to this his INR had been 1.8. His target INR however was 3.5. No alteration had been made in the patient's dosage of warfarin to try and bring the INR near the target level.	**SEVERE**
B	A patient handed in a form for a Blue Badge "disabled parking" for completion by the doctor. The receptionist attached the forms to the patient's notes and put them in the doctor's tray for his attention. When the doctor checked his tray the patient's notes were present but there was no Blue Badge form. It had become detached from the paper clip. The doctor asked the receptionist why the case notes were in his tray. He was told the reason and it was surmised that the form had become detached. After half an hour of searching the receptionist finally retrieved the form in amongst doctor's other paperwork.	**NO HARM**
C	A parent arrived at reception to collect a prescription for her son. The prescription, although generated, had not been signed and the "duty doctor" was asked to do so by the reception staff. The duty GP noticed that the telephone advice slip which the practice used for all telephone calls, was still attached to the child's file. The details on the advice slip did not seem to readily match with the treatment prescribed and the duty doctor then checked the patient's case notes and computer entry. Unfortunately there was no entry regarding a consultation that morning either in person or over the telephone; however, a prescription for Penicillin V syrup had been printed in the child's name. The duty doctor therefore spoke to the child's mother and a history was obtained and recorded in the files. A prescription for paracetamol was done and the printed prescription for penicillin was destroyed. In addition there were three previous entries for penicillin V with no data regarding its indication in the case notes.	**NO HARM (HARM PREVENTED)**
D	A female patient attended the surgery with breakthrough bleeding and abdominal pain. Amongst other investigations a cervical swab was taken which returned as positive for chlamydia. She was prescribed azithromycin and advised that her partner should be treated (partner was not a patient at the practice). She re-consulted 4 months later with similar symptoms, which again gave a positive chlamydia result. On questioning, she was asked if her partner had been treated after the first attack. She replied that he had not, as he had no symptoms and thought it unnecessary. She was given further treatment and advised her that her partner must contact his GP or sexual health clinic for treatment.	**MODERATE**
E	A doctor noticed that a female patient had been making regular monthly appointments to see her. However during consultations the topic frequently changed to her ill husband and her concerns about him. The consultations were lasting 30 minutes resulting in the doctor running late. This was stressing the doctor and also causing frustration to the patients left waiting to be seen.	**NO HARM**
F	A 32 year old female patient attended the surgery complaining of anxiety symptoms. The doctor decided to prescribe some Inderal to help with physical symptoms. He did not notice in the patient's case notes that she had a previous history of asthma, although she was not currently taking any medication. The patient became dyspnoeic and wheezy overnight and called the emergency services. She was taken to hospital where a diagnosis of beta-blocker induced asthma was made. She was given appropriate treatment and was released the following day.	**MODERATE**
G	A GP received a phone call on his mobile during his lunch break from his wife. The phone call was to tell him that she could not pick up their children from the private nursery at 5pm as she was working late. The GP would therefore have to pick up their children after his surgery finished at 6pm. The nursery had been informed of this by his wife, and was aware that the children would be picked up by their father nearer 6.30pm despite the fact that the nursery officially closed at 6pm. This meant that one of the nursery staff had to wait behind with the two children for half an hour after closing time and they made it clear that the nursery was not pleased with the situation.	**NO HARM**

Table 3: Response rates of different professional and academic groups of GPs

Professional & Academic Group	**Response Rate** n, (%)
GP Audit Development Group (ADG)	20/20 (100)
GP Trainers (Trainers)	26/32 (81)
GP Principals in training practices (TP-NP)	25/38 (66)
GP principals not in training practices (NTP)	26/40 (65)
GP Appraisers (Appraisers)	28/32 (88)
Overall Response Rate	**122/162 (77)**

Table 4: Rating severity of significant event scenarios and subsequent preparedness to undertake a formal significant event analysis by GP Group

		SIGNIFICANT EVENT SCENARIO CODES Mean (SD)						
QUESTION	**GP GROUP**	**A**	**B**	**C**	**D**	**E**	**F**	**G**
1. Please rate the 'significance' of this scenario on a scale of 1-7, where 1=not at all and 7=extremely significant (n=125)	**ADG**	6.4 (0.7)	2.3(1.1)	5.6 (1.1)	4.7 (1.4)	4.0 (1.3)	6.7 (0.5)	1.5 (0.7)
	Appraisers	6.3 (0.9)	1.9 (1.0)	5.4 (1.3)	3.9 (1.5)	3.8 (1.4)	6.4 (1.2)	3.1 (2.1)
	NTP	6.9 (1.3)	1.9 (1.2)	5.0 (1.3)	4.3 (1.7)	3.5 (1.4)	6.4 (0.7)	2.0 (1.5)
	TP-NP	5.9 (1.0)	2.7 (1.6)	5.3 (1.9)	4.1 (1.8)	3.7 (1.7)	6.2 (1.4)	2.5 (1.8)
	Trainer	6.2 (1.1)	2.8 (1.7)	5.7 (1.1)	4.2 (1.6)	4.0 (1.4)	6.6 (0.6)	2.0 (1.4)
	ANOVA p-value	0.35	0.37	0.30	0.50	0.79	0.36	0.01
2. Please indicate if you would be prepared to undertake a formal significant event analysis on this scenario on a scale of 1-7, where 1=definitely not and 7=definitely yes (n=125)	**ADG**	6.0 (1.4)	2.5 (0.7)	5.6 (1.3)	4.1 (1.6)	2.8 (1.4)	6.5 (0.7)	1.1 (0.3)
	Appraisers	5.8 (1.2)	2.4 (1.1)	5.0 (1.6)	2.9 (1.5)	2.6 (1.4)	6.3 (0.9)	1.6 (1.4)
	NTP	5.3 (1.2)	2.4 (1.3)	4.7 (1.6)	3.4 (2.1)	2.2 (1.3)	5.9 (1.4)	1.5 (1.3)
	TP-NP	5.7 (1.3)	2.8 (1.2)	4.8 (1.9)	3.5 (1.7)	2.8 (1.6)	5.8 (1.7)	1.6 (1.0)
	Trainer	5.7 (1.3)	3.0 (1.2)	5.2 (1.5)	3.2 (1.7)	3.0 (1.5)	6.4 (0.8)	1.2 (0.7)
	ANOVA p-value	0.60	0.50	0.40	0.24	0.28	0.17	0.41

Table 5: Proportions of positive agreement ('yes' responses) on sharing the event scenarios and preparedness to report each scenario by GP group

		SIGNIFICANT EVENT SCENARIO CODES N (%)						
QUESTION	**GP GROUP**	**A**	**B**	**C**	**D**	**E**	**F**	**G**
1. Do you agree that others in general practice would benefit from knowing about this event scenario (n=125)?	**ADG**	19 (95)	2 (10)	17 (85)	13 (65)	8 (40)	18 (90)	1 (5)
	Appraisers	26 (93)	2 (7)	17 (81)	17 (61)	10 (36)	26 (93)	2 (7)
	NTP	24 (92)	4 (15)	18 (69)	12 (46)	9 (35)	24 (92)	2 (8)
	TP-NP	24 (96)	6 (24)	18 (72)	17 (68)	9 (36)	23 (92)	7 (28)
	Trainer	22 (85)	4 (15)	21 (81)	17 (65)	9 (35)	24 (92)	3 (12)
	χ^2 p-value	0.60	0.49	0.33	0.52	1.00	1.00	0.09
2. Would you be prepared to report this event scenario as part of an anonymous and confidential reporting system (n=125)?	**ADG**	18 (90)	9 (45)	15 (75)	13 (65)	10 (50)	16 (80)	3 (15)
	Appraisers	27 (96)	10 (36)	15 (68)	16 (57)	11 (39)	25 (89)	5 (18)
	NTP	24 (92)	10 (38)	17 (65)	19 (73)	14 (54)	24 (92)	5 (19)
	TP-NP	24 (96)	11 (44)	18 (72)	17 (68)	13 (52)	23 (92)	8 (32)
	Trainer	25 (96)	11 (42)	22 (85)	18 (69)	10 (38)	23 (92)	6 (23)
	χ^2 p-value	0.84	0.96	0.57	0.78	0.69	0.71	0.65

Table 6: GPs' willingness to formally notify each significant event scenario to a range of potential organisations (n = number prepared to report)

	Significant Event Scenario n (%)						
Organisation	**A (n=118)**	**B (n=51)**	**C (n=91)**	**D (n=91)**	**E (n=58)**	**F (n=111)**	**G (n=27)**
Local Medical Committee (LMC)	55 (47)	29 (57)	51 (56)	46 (55)	34 (59)	57 (48)	14 (52)
Local Primary Care Organisation	37 (31)	21 (47)	35 (38)	34 (41)	27 (47)	37 (31)	10 (37)
NHS Education for Scotland	87 (74)	45 (88)	70 (77)	67 (81)	49 (84)	83 (70)	24 (89)
Local Community Health & Social Partnership	53 (45)	26 (51)	42 (46)	43 (52)	32(55)	49 (42)	14 (52)
Regional NHS Board	23 (19)	18 (35)	23 (25)	27 (33)	20(34)	25 (21)	11 (41)
National Reporting System (NHS)	59 (50)	23 (45)	47 (52)	41 (49)	29 (50)	57 (48)	14 (52)
National Reporting System (Independent of NHS)	48 (41)	22 (43)	44 (48)	35 (42)	29 (50)	54 (46)	14 (52)
Other	7 (6)	7(14)	12 (13)	13 (16)	6 (10)	8 (7)	6 (22)

Table 7: Differences in proportions of all GPs' responses to top two reporting preferences by significant event scenario

Significant Event Scenario (n)	NHS Education For Scotland n (%)	Local Medical Committee n (%)	Difference, (95% CI), P value
A (118)	87 (74)	55 (47)	0.27, (0.15, 0.39), P<0.001
B (51)	45 (88)	29 (57)	0.3, (0.15, 0.47), P<0.001
C (91)	70 (77)	51 (56)	0.2, (0.07, 0.34), P<0.05
D (91)	67 (81)	46 (55)	0.23, (0.09, 0.36), P<0.05
E (58)	49 (84)	34 (59)	0.26, (0.1, 0.41), P<0.05
F (111)	83 (70)	57 (48)	0.23, (0.11, 0.35), P<0.001
G (27)	245 (89)	14 (52)	0.37, (0.14, 0.59), P<0.05

Table 8: GPs' preparedness to involve other members of the primary care team if each significant event scenario was to be formally analysed

	Significant Event Scenario n =125 n (%)						
Group	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Immediate GP colleagues	123 (98)	72 (58)	118 (94)	102 (82)	99 (79)	122 (98)	37 (30)
Practice Manager	86 (69)	86 (69)	91 (73)	44 (35)	55 (44)	75 (60)	25 (20)
Reception & Administration	48 (38)	91 (73)	81 (65)	12 (10)	42 (34)	33 (26)	8 (6)
Practice Nurse(s)	95 (76)	95 (76)	51 (41)	80 (64)	19 (15)	69 (55)	5 (4)
Health Visitor(s)	7 (6)	2 (2)	10 (8)	11 (9)	12 (10)	9 (7)	3 (2)
District Nurse(s)	23 (18)	3 (2)	6 (5)	5 (4)	13 (10)	10 (8)	3(2)
Pharmacist	4 (3)	0 (0)	2 (2)	1 (1)	0 (0)	13 (10)	0 (0)
Patient/Relative(s)	15 (12)	0 (0)	14 (11)	21 (17)	13 (10)	15 (12)	0 (0)
Allied Health Professional(s)	32 (26)	2 (2)	11 (9)	32 (26)	13 (10)	16 (13)	3 (2)
Other	4 (3)	0(0)	1 (1)	5 (4)	1(1)	0 (0)	4 (3)

CHAPTER 8

A review of significant events analyses: reported improvements in the quality and safety of patient care

8.1 Introduction

Of the factors identified in Chapters 1 and 2, two interrelated issues arguably hinder the progress of the safety agenda in general practice more than most**.** Firstly, there is very limited evidence for the effectiveness of SEA in terms of its value in facilitating learning and change which leads to improvements in health care quality (Chapter 2). Secondly, fully engaging the primary care team in the notification of appropriate patient safety incidents as part of formal reporting and learning mechanisms has proved to be extremely challenging (DoH, 2001b; NPSA, 2006).

Gaining some insight into how SEA attempts by GP teams may contribute to our understanding of important quality and safety issues is clearly desirable to help close the evidential gap. A substantial bank of reports has been submitted and retained in the past decade by the west of Scotland peer review model, which offers important potential for research and learning.

The aim of this study was therefore to review the contents of SEA reports submitted by GPs and in doing so, to identify the range of quality and safety issues analysed, the types of learning needs raised and the purported actions implemented by health care teams.

8.2 METHODS

8.2.1 *Study sample and timescale*

All SEA reports analysed in this study were submitted to the NES peer review model between July 2005 and February 2007 using a standardised SEA framework and report format (Box 6). These were sent to two trained GPs for independent review using the feedback instrument described in Chapter 10. For the purposes of this particular study, reviewers were asked to make a judgment on whether a GP's SEA report demonstrated a 'satisfactory' level of insight and analysis. Where there was disagreement between two reviewers, two further experienced reviewers would make a final decision. Reports considered 'unsatisfactory' after peer review were excluded from this study because it was highly likely that key information was missing (Bowie *et al.*, 2005b). For example, insights into why the event happened may be lacking, learning needs may not have been identified or appropriate action was not taken to reduce the risk of recurrence. The reports were from two GP groups: GP principals and GP registrars. All GPs consented to their anonymised reports being used in this study (Appendix 5).

8.2.2 *Content analysis of SEA reports*

<u>Coding of events, reasons for occurrence, learning needs and actions taken</u>

Each SEA report was analysed for content independently by this author and a fellow researcher. The researchers developed a preliminary coding framework by merging and adapting the main categories and subcategories of errors (Makeham *et al.*, 2002; Rubin *et al.*, 2003), adverse events (Fischer *et al*.,1997) and potentially harmful events (Britt *et al*.,1997) previously reported in published research from primary care. The codings were further developed and refined as the study progressed. The authors were unaware of similar frameworks for classifying learning needs and actions taken. These were developed on an iterative basis as each SEA report was reviewed. Reports often

described multiple events, reasons for occurrence, learning needs and actions taken. A decision was taken not to assign a single code for each of these factors because of the inherent difficulty in reaching agreement since events could be highly complex, information could be missing and improvements were often multi-factorial.

8.2.3 *Data validation*

Joint meetings between the researchers took place after a set of five SEA reports had been reviewed for the purpose of comparing the codings assigned to the reports by the individual researchers. Where there was disagreement between researchers, a consensus was reached on the codes assigned. To enhance validity, a third researcher independently analysed one-in-five reports and the associated coding before meeting with the other two researchers to triangulate final agreement on the data collected.

8.2.4 *Data collection and analysis*

The following data were collected using a pre-designed proforma (Appendix 6): type of significant event; reasons cited for event occurrence; contributory factors as defined by the researchers using the NPSA model (Box 7); involvement of external agencies/individuals; level of patient harm (NPSA grading system: death, severe, moderate, low and none); type of learning issues identified; and type of actions taken. Data were entered into a Microsoft Excel spreadsheet for descriptive statistical analysis. Differences in proportions between GP groups were calculated along with 95% confidence intervals.

8.3 Results

8.3.1 *SEA reports reviewed and multiple coding*

286 SEA reports were submitted during the study timescale. Of these, 95 (33%) were excluded because they were judged 'unsatisfactory' by peer review. A total of 191 SEA reports were therefore subject to review, of which 98 (51%) were submitted by GP principals and 93 (49%) by GP registrars.

The number of reports with no code, a single code or multiple codes for each section of the SEA is shown in Table 9.

8.3.2 *Types of significant event*

A classification summary of the most frequently occurring significant event codes with examples is outlined in Table 10. Events involving disease diagnoses and management (46, 24.1%), such as the care of terminally ill patients, and prescribing issues (46, 24.1%), for example an inappropriate drug prescription, were the most common subjects for analysis. Almost as prevalent were issues precipitated by the patient or their relative (43, 22.5%), such as unnecessary medication requests, and anger or upset at an aspect of their healthcare.

8.3.3 *Reasons for occurrence*

A classification summary with examples of the most frequently reasons given by GPs as to why events occurred is outlined in Table 11. The most prevalent cause of events identified was 'errors' related to deficiencies in knowledge and skills which were attributed to individual healthcare professionals (62, 32.5%). Substandard communication issues between the practice and the patient, or within and between care providers was the second most frequent cause of significant events occurring (58,

30.4%). Patient behaviour, such as non adherence to medication or refusal to attend for investigations, was also a significant contributory factor in over a quarter of events (55, 28.9%). These three self-reported reasons for occurrence were the most frequent contributing factors assigned by the researchers to the significant event using the NPSA categorisation (Table 12).

8.3.4 *External involvement in significant events*

104 SEA reports (54.5%) described the direct or indirect involvement of other health and social care agencies in the significant event. The frequency of involvement was as follows: secondary care (58, 30.4%); other [e.g. police or ambulance service (32, 16.8%); community pharmacy (13, 6.8%); out-of-hours services (7, 3.7%) and social services (3, 1.6%).

8.3.5 *Self reported learning issues identified*

182 reports (95.3%) identified learning needs, points or issues which required to be addressed as part of event analyses (Table 13). Over half of all SEA reports identified personal learning issues for the individual doctor who drafted the event analysis. These learning issues related mainly to 'generic' issues around diagnosis, clinical management and patient behaviour. Specific learning points relating to clinical knowledge in areas such as psychiatry or contraception were detailed in less than a sixth of reports.

8.3.6 *Actions agreed and purported to be implemented*

154 reports (80.1%) demonstrated that change(s) had been agreed and implemented by at least one member of the primary care team as a result of the SEA (Table 14). 32 reports (16.6%) detailed a change(s) in which other GPs, nurses and health visitors in

the GP practice (as well as the GP author of the report) were able to apply new or revised clinical knowledge and skills. In just under one sixth of reports this application of new knowledge, skills and changes in clinical behaviour was applied by the GP author only. The most frequently occurring action taken was the development of new and adaptation of existing protocols. 73 reports (44.5%) incorporated protocols into practice procedures and policies.

8.3.7 *Levels of patient harm as defined by the NPSA grading system*

48 SEA reports (25.1%) described incidents which led to patient harm (Table 15). In total, 109 reports (57.1%) outlined circumstances which had the potential to cause patient harm but were either prevented or ran to completion without harm occurring ('near misses'). A minority of reports (34, 17.8%) did not involve incidents which could have compromised patient safety.

Of the 48 reports of patient harm, 42 (87.5%) described the direct or indirect involvement of health care teams or agencies external to the general practice as well as members of the practice team. In comparison, for reports that did not describe a patient safety incident, a total of 62 involved health care teams or agencies (56.9%) external to the practice [difference 30.6%; 95% CI 17.4 to 43.8].

Non–medical primary care staff were statistically less likely to be involved in both the learning (Table 16) and the implementation of change that resulted from the analysis of a significant event which resulted in patient harm than those events which did not cause harm (Table 17).

8.4 Discussion

8.4.1 *Main findings*

The types of significant events described in this study are consistent with the broad range of clinical and administrative events previously identified in similar general practice studies (Pringle *et al*., 1995; Cox and Holden, 2007). The majority of events had the potential to cause patient harm, while one quarter described actual incidences of patient harm. Several of the major underlying reasons why significant events occurred, such as knowledge and skills errors or communication difficulties, are consistent with previously reported reasons why errors and adverse events occur (Ely *et al*., 1995; Bhasale, 1998; Baker *et al*., 2004). Although learning issues are identified in the majority of SEA reports, these frequently relate to non–specific personal learning issues which are not shared with the practice team. There is also limited direct involvement of the practice team members in implementing the changes required from the SEA. If, however, a significant event has resulted in patient harm, then medical colleagues tend to be involved in the implementation of change in almost all cases.

8.4.2 *Study limitations and strengths*

The reports voluntarily submitted to the SEA peer feedback model described may not be representative of those undertaken by the GP population as they are likely to be highly selective (Bowie *et al.*, 2005a). The reports reviewed were judged by trained peers to be ‘satisfactory’ event analyses, while a substantial minority was considered ‘unsatisfactory’. This provides an element of verification and assurance of SEA quality which does not exist in other similar studies. SEA reports are written retrospectively, normally by a single author. Personal recollections may therefore be affected by recall bias or misinterpretation of reasons for event occurrence or decisions made. In addition,

although the researchers did not examine inter-rater reliability, an attempt was made to minimise data coding variability through researcher triangulation.

The format of the SEA report lacks a theoretical categorisation framework with which to guide GPs as to the perceived causes of the events or the changes to be implemented. For example, Vincent *et al*. (1998) suggested that categories such as the organisation, work environment, team issues, individual and task factors as well patient characteristics could be used to aid analysis of risk and safety in medicine. Using an accepted model such as this or the Systems Engineering Initiative for Patient Safety (SEIPS) model could have strengthened the findings through clarifying and reducing overlap amongst categorisations (Carayon *et al.*, 2006). These models are, however, not widely used in general practice.

8.4.3 *Context and Implications:*

Types of Events

The range of significant events identified for analysis in this study confirms the potential for SEA to examine a breadth of patient-safety related subject matter which appears largely unlikely to be investigated or uncovered by quantitative audit. Although there is much to be learned from 'good practice', it is apparent that GPs participating in this model of SEA in the vast majority of cases choose to examine events that could or did highlight sub-optimal care. This is presumably because they find these events a more valuable learning experience. The most common clinical areas for analysis were cancers and acute psychiatric events and this most probably reflects their role as 'marker' events identified by the General Medical Services contract (GMS) in the UK as being worthy of SEA (DoH, 2004). As previously described, the benefits of using these 'marker' events are that they can provide opportunities for prevention, early detection

and inform the process of care but at an 'opportunity-cost' (Pringle, 2007). They may also highlight previously unknown learning needs. The frequent choice of events that relate to disease diagnosis and management and those that involve prescribing and drug issues is consistent with GPs selecting topics that reflect the routine case work of general practice. It also reflects the subject choice of SEA in other parts of the UK (Cox and Holden, 2007).

Reasons for occurrence

This study found that the two most common reasons cited for the significant event having taken place were self-reported individual errors by the doctor and communication issues. The role of individual error may reflect appropriate insight on the part of the doctor or could reflect a lack of understanding of the deeper reasons behind errors or violations. The roles of 'systems' and 'human factors' analyses are unlikely to be fully explored in most SEA and may require separate investigation by individuals trained in such techniques. Despite this, there was broad agreement between what GPs identified as the reason(s) for events occurring and the researchers' own interpretations of the event causation.

Given that these SEA reports are submitted as part of an educational exercise for the GP, it may be that the doctors feel they will learn more by analysing significant events attributable to themselves than events more directly connected to others or team members. Additionally, GPs are known to submit events that they feel responsible for as a form of 'personal catharsis' (Bowie *et al.*, 2005a). Patient behaviour was thought to be an underlying factor in over a quarter of significant events, which has rarely been cited as a key reason for significant event occurrence in general practice (Elder, 2004). This does not necessarily mean that there is 'blame' attached to a patient's role in the

significant event. Their role can be influenced by a host of factors, including illness behaviour, lack of knowledge on the part of the patient or their carers, or poor lines of communication between the practice and the patient. These issues may deserve further exploration.

External involvement in SEA

A study in one PCT in England found that nearly 19% of significant events originated out-with the general practice unit (Cox and Holden, 2007). Although this study did not look at the place of origin of the significant event, the finding that over half of the significant events involved an agency external to the practice highlights the potential for multi–disciplinary and multi-agency learning and collaboration in event analysis. This finding may also demonstrate that many GP teams are prepared to investigate what could be interpreted as difficult-to-resolve "interface" issues. Alternatively, there could be a degree of 'blame-shifting' attached to the event analyses.

Learning and change

The potential role of SEA as a reflective learning technique has been highlighted in Chapter 2. The findings provide further evidence of this but it is of note that much of the learning would appear to be personal to the SEA report author. There may be a reluctance to share this with colleagues if the GP was professionally embarrassed or felt that the learning point was not of sufficient interest to others. Another possibility is that the authors of the SEA reports failed to document that they shared their learning with other members of the practice team.

Although reflection and learning is recognised as an important part of improving the quality of care in the NHS (DoH, 2001a) it is the application of this learning into

sustained change that will enhance the quality of each individual's care. The evidence that SEA can improve the quality of care, and enhance patient safety in particular, is limited (Bowie *et al.*, 2008). However, perhaps it should be seen in the context of a raft of other quality improvement approaches - such as audit, prescribing reviews, referral analysis, complaints review and the nGMS contract itself - that are available to GPs and their teams as part of a multi-method approach to reducing harm and minimising risk.

Patient safety

The small percentage of SEA reports that involved severe harm or death is consistent with other recent UK data on the severity of events analysed by GPs (NPSA, 2008; Cox and Holden, 2007). With one in four events involving some form of harm (nearly all of which involve external agencies), it is also apparent that GPs are prepared to confront potentially difficult issues that may not reflect well on them or their practices. The majority of events had the potential to cause patient harm but did not actually do so. GPs have been encouraged to identify and use these types of 'near misses' to highlight learning and safety issues in practice rather than wait till a patient inevitably suffers harm (NPSA, 2005). Given the range of outcomes outlined in the reports, if properly applied, the SEA technique would appear to be well-placed to inform and educate on a wide range of patient safety-related issues in general practice.

Although SEA is encouraged as a team-based activity (Pringle *et al.*, 1995; DoH, 2004), if the event in question involved patient harm, non-medical staff were less likely to be involved in the learning and change as a result of the analysis. As most GPs will involve staff only where they deem it appropriate (Chapter 7), this may represent a pragmatic decision on the part of GPs to form a smaller, more relevant group when clinical care issues are involved. Alternatively, there may be a reluctance to open up discussion to

the wider primary care team if the doctor perceives patient harm to have been in some way related to their own practice.

Significant events are often described as a rich resource that could aid both local and national reporting and learning systems (NPSA, 2005). Indeed, most of the patient safety incidents amongst the SEA reports in this study could be applied to the NPSA reporting template and other international taxonomies (Dovey *et al*., 2002; Makeham *et al*., 2008). The narrative aspect of SEA should serve to enrich the basic factual details applied to reporting systems in health care and thus could potentially offer much more information than most reporting forms (Berwick, D., 2008; Vincent, 2006). A drawback may be that narratives could be more difficult to code and analyse. Additionally, it is highly likely that the patient harm described in this study is an under-representation of the true volume of patient harm identified by routine application of SEA. This is because GPs are selective in the type of significant events that they choose to submit to this peer review model and those GPs who do submit their SEA for feedback may not be representative of their GP colleagues. Given the large scale non-engagement of GPs in reporting systems (NPSA, 2008), it may be appropriate to encourage the sharing of these reports to enhance patient safety.

8.5 Conclusions

The study adds to the limited evidence of the potential of SEA to improve healthcare quality and safety. It is applied to investigate and learn from a wide variety of quality-related issues identified by GP teams, including those resulting in patient harm and which often involve other health care agencies. Learning and change reportedly occur, but more research is required to establish if sustainable improvement is possible.

Table 9: Number and percentage of reviewed SEA reports with zero, single and multiple codes in each section of the SEA report (n = 191)

SEA Report Section	Number and Percentage of Codes (with range)					
	0	**1**	**2**	**3**	**>3**	**Range**
Section 1 – What happened?	0	43 (22.5%)	80 (41.9%)	54 (28.3%)	14 (7.3%)	0-4
Section 2 – Why did it happen?	0	47 (24.6%)	72 (37.7%)	40 (20.9%)	32 (16.8%)	0-6
Section 3 – What have you learned (n>0, =182)	9 (4.7%)	60 (31.4%)	62 (32.5%)	41 (21.5%)	20 (10.5%)	0-7
Section 4 – What have you changed (n>0,=154)	37 (19.4%)	98 (51.3%)	44 (23.0%)	10 (5.2%)	2 (1.1%)	0-5

Table 10: Type and proportion of significant events based on 191 SEA reports

Significant Event Types	**n**	**%**
Disease diagnosis and management *(e.g. terminal care management, missed diagnosis)*	46	24.1
Prescribing, dispensing and other drug issues *(e.g. wrong/inappropriate drug prescribed/administered, warfarin issue)*	46	24.1
Patient and relatives *(e.g. patient behaviour, anger or upset)*	43	22.5
Investigations and results *(e.g. incorrect results given to patient, results not acted upon)*	37	19.4
Communication *(e.g. lack of communication, unsuccessful communication)*	23	12.0
Administration *(e.g. complaint, breach of protocol)*	16	8.4
Medical records and confidentiality *(e.g. breach of confidentiality, wrong records accessed)*	15	7.9
Appointments and surgeries *(e.g. patient did not arrive issues, running late)*	12	6.3
Home visits and external care *(e.g. delay in arrival, visit request not done)*	10	5.2
Equipment *(e.g. computer search facility ineffective, difficulty accessing cupboard containing medical supplies)*	7	3.7
Miscellaneous *(e.g. difficulty in signing death certificate, change in partnership personnel)*	2	1.1
Health and safety *(e.g. staff injury, unsuccessful procedure for dealing with clinical waste)*	2	1.1

* More than one classification may have been accorded to a single SEA report.

Table 11: Reasons for occurrence of significant events from 191 SEA reports

Reasons Given	n	%
Individual health care professional 'errors' *(e.g. lack of knowledge of practice/hospital protocols, poor clinical task delivery)*	62	32.5
Communication *(e.g. substandard communication between practice and patient, substandard communication between practice and hospital/out of hours/other agencies)*	58	30.4
Patient and relatives *(e.g. negative patient behaviour, illness behaviour)*	55	28.9
Disease/diagnosis/management *(e.g. difficult diagnosis, incomplete history/examination)*	44	23.0
Administration *(e.g. poor task delivery, ineffective administrative system/protocol)*	32	16.8
Medication *(e.g. error writing/prescribing/administering (wrong drug dosage/formulation prescribed), no system/protocol to check for out of date emergency tray/bag medicines)*	23	12.0
Tests/investigations/results *(e.g. no sample tracking/record, delay in checking blood tests results)*	22	11.5
Patient records *(e.g. failure to check notes adequately, failure to record in notes)*	18	9.4
Equipment *(e.g. ineffective emergency buzzer system for staff to identify location of emergency, inadequate search facility on computer system)*	13	6.8
General practice protocols/systems/guidelines *(e.g. no formal protocol for checking BHCGs, no system for emergency bag tracking)*	8	4.2
Clinical behaviour *(e.g. doctor avoidance of addressing a difficult situation, lack of clinical leadership of patient review)*	8	4.2
Reasons for event undetermined	7	3.7
Appointments *(e.g. delay in being seen, not enough time with patients)*	6	3.1
Visits/external care *(e.g. change in out of hrs service, delay in attending house visit)*	3	1.6

- More than one classification may have been accorded to a single SEA report.

Table 12: Contributory factors assigned by researchers to significant events using NPSA categorisation

Contributory factors	**n**	**%**
Individual Staff	100	54.9
Patient Factor	66	36.3
Communication	60	33
Task	28	15.4
Education & Training	26	14.3
Equipment & Resources	22	12.1
Team & Social	21	11.5
Working Condition & Environment	6	3.3
Other	3	1.7

Table 13: Type, range and number of learning needs identified from 191 SEA reports submitted

Learning Issues identified	**n**	**%**
Personal Awareness/Responsibilities and Change (*e.g. general issues on improving diagnosis and management, dealing with negative patient/family behaviour*)	98	*51.3*
Communication (*e.g. issues on communication with patient, issues on communication within team*)	54	*28.3*
Administration (*e.g. need for new/improved protocols/systems, staff training required*)	36	*18.9*
Clinical Knowledge (*e.g. psychiatry, contraception*)	30	*15.7*
Equipment/task aids/workspace (*e.g. become familiar with medical centre – whereabouts of drugs and equipment, the need to secure and check prescription pads*)	26	*13.6*
Case Notes (*e.g. the need for accurate detailed documentation, Read clinical notes*)	17	*8.9*
Whole Practice Awareness (*e.g. need for system to respond to emergency within practice, clarification of responsibilities*)	16	*8.4*
Medication/Prescription (*e.g. responsibilities for GP-secondary care interface prescribing, need for better supervision of PRHO prescribing*)	9	*4.7*
Patient/Carers (*e.g. effects of mental illness on carers, education required on self-management of asthma*)	6	*3.1*
Complaints (*e.g. dealing with complaints system, avoiding complaints being generated*)	4	*2.1*
GP and Partners Awareness (*e.g. the need for regular medication review with GPs, the importance of efficient and accurate results handling system*)	4	*2.1*
Health and Safety (*e.g. re-sheathing needles should not be undertaken, to ensure all clinical staff immunised against Hepatitis B*)	3	*1.6*

- 182 reports detailed at least one learning point
- More than one learning point may have been reported in a single report.

Table 14: Type and range of actions taken from 191 SEA reports

Changes implemented	**n**	**%**
Clinical Team Disease Diagnosis and Management (*e.g.* raised clinical awareness/knowledge by dissemination to others then actioned, *raised procedural awareness and dissemination to others for action*)	32	16.7
Doctors Personal Skills/Behaviour/Knowledge application (*e.g. change in behaviour, application of knowledge*)	28	14.6
Communication (*e.g. improved communication with patients, improved communication between practice staff or between staff and doctors*)	26	13.6
Administration (*e.g. clarification of staff duties, member of staff designated to a particular role*)	26	13.6
Medication (*e.g. change to prescribing software, medication change highlighted on discharge script from hospital*)	24	12.6
Results/Investigations/Tests (*e.g. stop doing in-house tests*)	18	9.4
Patient Records (*e.g. improved recording in notes - paper or electronic, important patient information highlighted in notes - electronic or paper*)	16	8.4
Equipment and Workspace (*e.g. improved storage of equipment, face mask added to medical bag*)	15	7.8
Appointments (*e.g. increase appointment time for emergency surgery*)	3	1.6
Miscellaneous (*e.g. equipment & stocking within consulting room*)	2	1.0
Staffing/Premises Issue (*e.g. book used to document leave, revision of supervisory arrangements*)	2	1.0

- 154 SEA reports detailed actions to implement change
- More than one change may have been reported in a single report.

Table 15: Level of patient harm as determined by NPSA grading system

NPSA Severity Grading	GP Principals (n)	GP Registrars (n)	Total n (%)
No Harm – Impact Prevented *Any patient safety incident that had the potential to cause harm but was prevented, resulting in no harm to people receiving NHS-funded care*	24	28	52 (27.2)
No Harm – Impact Not Prevented *Any patient safety incident that ran to completion but no harm occurred to people receiving NHS-funded care*	33	24	57 (29.8)
Harm – Low *Any patient safety incident that required extra observation or minor treatment and caused minimal harm to one or more persons receiving NHS-funded care*	7	7	14 (7.3)
Harm – Moderate *Any patient safety incident that resulted in a moderate increase in treatment and which caused significant but not permanent harm to one or more persons receiving NHS-funded care*	9	13	22 (11.5)
Harm – Severe *Any patient safety incident that appears to have resulted in permanent harm to one or more persons receiving NHS-funded care*	6	3	9 (4.7)
Harm – Death *Any patient safety incident that directly resulted in the death of one or more persons receiving NHS-funded care*	2	1	3 (1.6)
Not applicable *Non-patient safety incidents*	18	16	34 (17.8)
TOTAL			**191**

Table 16: Involvement of healthcare professionals in the learning from undertaking a SEA

Learning Needs Identified (182/191 = 95.3%)	Reporting GP	Partners	Practice Manager	Practice Nurse	Administration. Staff	Health Visitor	District Nurse	Comm. Pharmacist	Hospital
Patient Harm (n=44)	44 (100%)	13 (29.6%)	2 (4.6%)	2 (4.6%)	4 (9.1%)	0 (0%)	0 (0%)	1 (2.3%)	0 (0%)
Non-Patient Harm (n=138)	138 (100%)	57 (41.3%)	35 (25.4%)	29 (21.0%)	29 (21.0%)	2 (1.4%)	0 (0%)	6 (4.3%)	1 (0.7%)
Difference (95% Confidence Intervals) P-value	*NA*	*11.8 (4-27) P=0.14*	*21 (11-30) P<0.01*	*16 (7-26) P <0.01*	*12 (1-23) P=0.03*	*NA*	*NA*	*NA*	*NA*

Table 17: Involvement of healthcare professionals in the implementation of change from undertaking a SEA

Action Taken (154/191 = 80.6%)	ReportingGP	Partners	Practice Manager	Practice Nurse	Admin. Staff	Health Visitor	District Nurse	Comm. Pharmacist	Hospital
Patient Harm (n=34)	34 (100%)	28 (82.4%)	7 (20.6%)	3 (8.8%)	12 (35.3%)	0 (0%)	0 (0%)	0 (0%)	1 (2.9%)
Non-Patient Harm (n=120)	117 (97.5%)	87 (72.5%)	47 (39.2%)	33 (27.5%)	47 (39.2%)	2 (1.7%)	4 (3.3%)	0 (0%)	2 (1.7%)
Difference (95% Confidence Intervals) P-value	*NA*	*10 (5-25) P=0.20*	*18(2-35) P=0.02*	*19(6-31) P=0.003*	*4(14-22) P=0.628*	*NA*	*NA*	*NA*	*NA*

Box 6: Summary of standard SEA framework and report format recommended by NES Scotland

1. What happened?

- Collate and record as much factual information as possible about the event including, for example, what happened, when and where, what was the outcome and who was involved.
- Record the thoughts and opinions of those involved, including patients and relatives if appropriate, and attempt to form an accurate impression of what happened.

2. Why did it happen?

- Ensure the main reasons why the event occurred are fully established and recorded, e.g. was it a failure in a system or a failure to adhere to a protocol?
- Establish the underlying or contributory reasons as to why the event occurred, e.g. why was there a failure in a system or adherence to a protocol?

3. What has been learned?

- Agree and record the main learning issues for the health care team or individual team members.
- Ensure that insight into the event has been established by the team or the individuals concerned.

4**. What has been changed?**

- Agree and implement appropriate action in order to minimise the chances of the event recurring, where change is considered to be relevant.
- Monitor the implementation of any change introduced.

Box 7: NPSA contributing factors to potential patient harm (NPSA, 2005)

Patient factors
- These are unique to the patient(s) involved in the incident, such as the complexity of their condition or factors such as their age or language spoken.

Individual factors
- These are unique to the individual staff member(s) involved in the incident. They include psychological factors, home factors and work relationships.

Task factors
- These include aids that support the delivery of patient care, such as policies, guidelines and procedural documents. They need to be up-to-date, available, understandable, useable, relevant and correct.

Communication factors
- These include communication in all forms: written, verbal and non-verbal. Communication can contribute to an incident if it is inadequate, ineffective, confusing, or it is too late. These factors are relevant between individuals, within and between teams, and within and between organisations.

Team and social factors
- These factors can adversely affect the cohesiveness of a team. They involve communication within the team, management style, traditional hierarchical structures, lack of respect for less senior members of the team and perception of roles.

Education and training factors
- The availability and quality of training programmes for staff can directly affect their ability to perform their job or to respond to difficult or emergency circumstances. The effectiveness of training as a method of safety improvement is influenced by content, delivery style, understanding and assessment of skill acquisition, monitoring and updates.

Equipment and resources factors
- Equipment factors include whether the equipment is fit for purpose, whether staff know how to use the equipment, where it is stored and how often it is maintained. Resource factors include the capacity to deliver the care required, budget allocation, staffing allocation and skill mix.

Working conditions and environmental factors
- These factors affect ability to function at optimum levels in the workplace and include distractions, interruptions, uncomfortable heat, poor lighting, noise and lack of, or inappropriate use of space.

CHAPTER 9

The initial development and testing of an assessment instrument for the formative peer review of significant event analyses.

9.1 Introduction

Although general practitioners (GPs) and other primary healthcare staff were increasingly taking part in event analyses for the reasons described in Chapter 2, there was at this stage limited evidence of an individual GP's ability to verifiably undertake the technique effectively (Bowie *et al*., 2003; McKay *et al.,* 2003; McKay *et al.*, 2006). This is important because superficial or informal discussion of an event is unlikely to lead to a deeper understanding of causal factors, the identification of appropriate learning needs and the implementation of necessary change (Bowie, 2005b).

One method of informing on the quality of SEA is through external and independent review by trained peers. As described in Chapter 3, 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 arrangements for their continuing professional development (CPD) since 1998.

The assessment instrument, however, had been subject to critical comment. During quarterly periodic meetings, GP peer reviewers indicated that it was too limited to allow sufficient educational feedback because it did not adequately cover the broad spectrum of significant events or recognise the variation in the depth and complexity of some event analyses. The requirement for reviewers to make a dichotomous judgement on whether the SEA report was satisfactory or not (which was then reported back to the participant) was thought to emphasise a summative judgement on what was intended to be a formative educational exercise. In addition, external feedback from journal

referees and journal correspondents (Sykes, 2003) commented on the strong emphasis within the assessment process on the need to implement change as part of SEA to obtain a 'satisfactory' outcome.

Within the context of the utility of the whole peer review model, and given the perceived importance of the SEA technique to GP appraisal, the nGMS contract and clinical governance, the development of a feedback instrument examined for its validity and reliability would be highly desirable. In this way, a professional judgement could be made on the quality of the event analysis in question and formative feedback provided for consideration. By raising the quality of event analyses undertaken by GPs and their teams, there is clear potential to further enhance learning and the quality of patient care.

This initial study was undertaken to establish the content validity of an updated peer review instrument, elucidate aspects of its reliability and investigate possible sub-sample differences between GP principals and registrars, which would be relevant to generalising to a wider population of GPs.

9.2 Method

9.2.1 *Content Validity*

The developmental stage to assimilate the proposed items for the instrument was initially carried out independently by this author and two other colleagues. This work was informed by previous focus group interviews with the west of Scotland Audit Development Group (Lough, 2002). Agreement was then reached on four criteria considered "essential" for assessment of a significant event analysis. Together with

previous research (Pringle *et al.*, 1995) these criteria were prioritised to generate domains and items after discussion between the researchers.

These domains and items were discussed amongst the three researchers until consensus was achieved on the items to be included in the 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 (Figure 4).

The proposed instrument, a covering letter and guidance notes (Appendix 7) were sent to ten general practitioners, herein referred to as "experts", who were identified as being "well informed" in the SEA technique because they were either experienced peer assessors or had published research papers on this topic 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 (Figure 5). Content validity was then quantified by application of a content validity index (CVI). Eight 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 (Lynn, 1986). This was determined to provide sufficient evidence for inclusion of each item as part of the final instrument.

9.2.2 *Reliability Testing*

Participants and assessment exercise

The proposed instrument was introduced at a training day to the west of Scotland Audit Development Group from which all of the peer assessors were drawn (Box 3, Chapter 3). The role of the assessors and any clarification points around using the instrument

were discussed. All assessors agreed to use the new instrument concurrently with the existing assessment instrument as new SEA reports were received. Any further issues raised by assessors were to be communicated to the authors by electronic mail as they arose.

At three monthly follow-up meetings issues arising from the use of the instrument were also discussed with the authors and the rest of the assessment group. All 20 assessors also agreed to take part in a reliability marking exercise. A nested design consisting of five cells, each with four assessors was used and members of each cell reviewed 20 separate SEA reports, unique to that cell, using the proposed new assessment instrument. The exercise was repeated after one month with the raters in each cell marking the same unique 20 SEA reports. To explore any potential differences in the reliability of the instrument between the two groups of GPs most likely to submit SEA report to the peer review model, the 20 SEA reports for each cell consisted of 10 SEA reports undertaken and submitted by GP principals (i.e., experienced doctors) and 10 SEA reports from GP registrars (i.e., doctors in training).

Data Analysis:

A repeated measures ANOVA was undertaken using BMDP software. The analysis was to establish the variance attributable to each variable being included in the study design (SEA reports, n=100, 20 per cell), (raters, n=20, 4 per cell), (time, n=2), (items, n=10). Generalisability theory was then used to investigate the instrument's ability to consistently discriminate among SEA reports (Cronbach, 1972). The internal consistency (a measure of item homogenicity), intra-rater agreement (same rater on two occasions) and inter-rater (agreement among raters) were all calculated. Separate analyses are reported on the SEA report provided by experienced doctors (GP principals) and those provided by less experienced doctors (GP registrars).

9.3 Results

9.3.1 *Content Validity*

At least eight out of 10 experts endorsed all 10 items listed in Figure 5 and the overall instrument, thereby 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.

9.3.2 *Reliability*

The G co-efficients obtained for the overall test reliability, internal consistency, and inter/intra-rater reliability values for the instrument when used to assess SEA reports submitted by GP principals are shown in Table 18. Equivalent results based on the analysis of reports submitted by GP registrars are shown in Table 19.

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.

The high intra-rater co-efficients for SEA reports undertaken by GP principals (0.78) and GP registrars (0.71) suggest that individuals' opinions regarding the quality of each SEA are reasonably stable over time, thereby supporting the use of the instrument as a measure to provide formative educational feedback to the submitting practitioner on the standard of an event analysis report.

The moderate G co-efficients for inter-rater reliability, assessed using the average score of those 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 would be required for the average score to be deemed highly reliable (i.e. G > 0.8).

The overall test G co-efficient, though low for each group for any single item (0.25, 0.18), was good (0.73, 0.71) when considered for the average of all items, thereby indicating that the instrument can provide consistent information on the standard achieved by the SEA report.

A comparison of the mean scores between GP principals' and GP registrars' SEA reports is shown in Table 20 and demonstrates no difference in the mean scores between the two groups.

9.4 Discussion

The main findings of this study demonstrate that the content validity and reliability of the assessment instrument are adequate for use in a formative peer review model, providing the first steps towards developing 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 amongst peer assessors in their assessment of SEA reports. Consistent with previous research (McKay *et al.*, 2006), no difference was found in the quality ratings assigned to SEA reports completed by GP principals or GP registrars.

9.4.1 *Limitations of the study*

Validity testing

This instrument has been developed by general practitioners to help inform colleagues on the quality of significant event analyses. By definition, the assessment instrument is therefore doctor-centred, despite the frequent multi-disciplinary and multi-professional nature of significant events and their analyses (Bennett and Danzak, 1995; Westcott *et al.*, 2000).

For the purpose of this study, the "expert" raters were simply "well-informed" individuals as the number of individuals with sufficient knowledge and experience to be deemed truly expert is limited and, it must be acknowledged, poorly defined (Fink *et al.*, 1984; van der Vleuten, 1996).

The CVI exercise was adequate but a different consensus method approach, such as the Delphi technique or nominal group technique, could have been applied (Jones and Hunter, 1995). The former may have allowed a broader range of relevant individuals to have expressed opinions on what should have been included in the instrument, while both could have enabled more of a challenge to have been made on the researchers' justification(s) for inclusion of items. For all these methods, consensus in itself does not mean that the 'correct' items have been identified.

Reliability

The reliability exercise was carried out using written SEA reports. The significant events chosen for analysis by both the GP principals and GP registrars to submit for peer review were self-selected. These reports may not be representative of significant events that are analysed within practices (indeed, SEA reports can only ever act as a

'proxy' for the actual analysis) and so the instrument may be limited by being tested on a relatively narrow range of examples. The finding that most event analysis reports were rated as having a global score of 4 or above on the rating scale may indicate that there is a bias toward submission of reports with which the submitting doctor feels comfortable. Also, there was no way of knowing the extent to which each report was team-based, with ownership of the event analysis being assumed by the doctor (Bradley, 1992). The impact of these limitations, however, should have worked against the observation of sufficient reliability.

In addition, it should be noted that the reviewers were individuals with extensive experience with SEA and have a considerable amount of opportunity to discuss how to interpret the rating task they were to perform. Further study is required to determine whether or not similar findings would be achieved with less experienced reviewers. This is explored in more detail in Chapter 13. A further limitation of the study is that although the instrument is designed to provide written comments as well as numerical data, this study analysed only the latter. For a formative instrument, written comments may be at least as important to the submitting doctor (Sargeant *et al.*, 2005).

9.4.2 *Findings in Context*

This study has examined only a small part of the process of instrument evaluation. The reliability exercise places much emphasis on the quantifiable aspects of testing and does not consider the effect that the instrument has on learning and needs to be considered in a broader educational framework beyond the limits of psychometric testing (Moss, 1992; van der Vleuten, 1996). This aspect of the instrument, and indeed the whole peer review model, therefore required its own separate evaluation and is reported in Chapters 11 and 12.

Acceptable levels for reliability coefficients are based on opinions. For 'high' stakes examinations reliability, coefficients of between 0.85 and 0.9 have been suggested (Weiner and Stewart, 1984; Downing, 2004) while 'lower' stakes assessments and 'research' projects should aim for levels above 0.7 (Nunnally and Bernstein,1978; Downing, 2004). The interpretations given to the reliability coefficients are therefore just that. Instruments are not reliable or unreliable, rather they will have degrees of reliability that helps inform one aspect of utility (Streiner and Norman, 2003).

A strength of this instrument is that it is for use in the work place 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 encourages reflection and learning from actual events. This should add to the potential attractiveness and relevance of the instrument and therefore its impact.

The study demonstrated content validity, but further work would be required to examine other aspects such as criterion or construct validity. The high G co-efficients would indicate that the domains and items are inter-related and the CVI indicates that our judges considered the questions to be relevant. One criticism of the high intra-class correlation may be that it indicates that there are too many items in the instrument and thus the number of items could be reduced. Conversely, this may be of benefit in a formative instrument. However, as the instrument is relatively short, the current number of items should be feasible for reviewers**.** Currently, a standard universal SEA method does not yet exist, although several are proposed (Pringle *et al*., 1995; Lough, 2002). This instrument mirrors the emphasis on understanding, learning and change associated with these approaches, but is unique in providing written feedback by peers.

Further studies on the feasibility, acceptability and consequential validity testing as part of the overall utility of the instrument (and peer review model) would help inform this.

The testing process did not look at context specificity, therefore the instrument cannot identify whether a GP is proficient (or otherwise) in applying the SEA technique. It is not designed to summatively judge in terms of "passing" or "failing", or to support accreditation of GPs or their teams. The study merely set out to test whether SEA reports could be reliably evaluated and could thus facilitate written educational feedback. This is in keeping with the context of developing the educational component of the instrument as part of an overall formative model. Feedback itself is important in enhancing patient safety by highlighting issues that could further improve the submitted SEA report and the standard of future event analyses. As discussed in Chapters 3 and 5, formative assessments place more emphasis on an instrument's validity and educational impact than reliability. However, given that the various elements of an instrument's utility are multiplicative, it is still essential that reliability is evaluated (McKinley *et al.*, 2000).

The largest degree of error when providing feedback using this instrument is the variation among peer assessors. This is a difficulty for many assessment instruments, including those used to assess GPs' personal development plans (Roberts *et al.*, 2006) and information from multi-source feedback exercises (Ramsay and Wenrich, 1999). The moderately large G co-efficients for intra-rater reliability would imply that there is a reasonable degree of instrument stability when used by individual peer reviewers to assess reports at different points in time. The lower inter-rater reliability is more likely, therefore, to be related to training and calibration issues among the assessors rather than the robustness of the instrument itself.

As highlighted by Norman *et al*. (2004), there is a need to achieve a balance in continuing professional development between self-assessment and top-down regulatory approaches. An ideal tool would be something that is seen as "*supportive and individual, yet objective*". In addition, for a tool to be successful as a formative feedback instrument, it should give information via numerical scores and written feedback which should be used in conjunction with facilitated feed-back (Sargeant *et al.*, 2005). The format of the peer review system utilising this instrument would fit with both of the above concepts in that it promotes self (or team) directed learning and reflection. Assessors are encouraged to detail written formative feedback as well as to give numerical scores and it is known that GP participants in this model of peer would use review use their SEA reports as supporting evidence for appraisal, accreditation and contractual obligations (Bowie *et al.,* 2005a).

Although SEA is only one element in the spectrum of a general practitioner's work, participation in an educational model of peer review for the analyses of significant events may not only enhance an individual's learning but also demonstrate to patients, appraisers, health-care organisations and possibly future revalidation authorities a willingness for the doctor to submit aspects of their own work for external review as part of an educational process. This would serve to confirm that the GP is reflecting on how patient care could be improved, thus signifying verifiable participation in one important aspect of the clinical governance agenda.

9.5 Conclusion

Although this study makes only one small step in verifying two aspects of an assessment instrument’s utility, the findings justified further development of the

instrument. In particular, there was a need to strengthen the inter-rater reliability of reviewers and to investigate the ability of the instrument (and more generally the peer review model) to identify GPs' educational needs in relation to SEA.

Table 18: Calculated reliability co-efficients for GP principals' SEA reports marked using the peer review instrument (expressed with 95% confidence intervals)

	Overall	**Internal Consistency**	**Intra-Rater**	**Inter-Rater**
Single Item	0.25 (0.17-0.33)	0.62 (0.55-0.68)	0.64 (0.57-0.70)	0.31 (0.23-0.39)
Total Score (i.e., average of Average of all 9 items	0.73 (0.68-0.73)	0.94 (0.93-0.95)	0.78 (0.73-0.82)	0.64 (0.57-0.70)
Global Score (Item 10)	0.80 (0.76-0.84)	N/A	0.70 (0.64-0.75)	0.43 (0.35-0.51)

Table 19: Calculated reliability co-efficients for GP registrars' SEA reports marked using the peer review instrument (expressed with 95% confidence intervals)

	Overall	**Internal Consistency**	**Intra-Rater**	**Inter-Rater**
Single Item	0.18 (0.11-0.26)	0.48 (0.4-0.55)	0.55 (0.48-0.62)	0.27 (0.19–0.35)
Total score (i.e. Average of all 9 items)	0.71 (0.65-0.76)	0.89 (0.87-0.91)	0.71 (0.65-0.76)	0.6 (0.53-0.66)
Global score Item 10	0.83 (0.80-0.86)	N/A	0.58 (0.51-0.64)	0.42 (0.34-0.50)

Table 20: Comparison of the mean scores between GP principals' and GP registrars' SEA reports

	Principals	**GP registrars**	**P value**
Mean score over items 1-9 (+/- 1S.D.)	4.81 (0.79)	4.88 (0.63)	0.59
Mean global score of SEA reports (+/- 1S.D)	4.85 (0.80)	4.94 (0.99)	0.58

Figure 4: Significant Event Analysis - Proposed peer review and feedback instrument (cropped version)

Please rate the level of evidence contained in the SEA report for each of the following (using the rating scale where 1=Absent and 7=Excellent):

1. The description of what actually happened Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
2. The role(s) of all individual(s) involved in the events has been described: Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
3. The setting(s) where the event happened has been described: Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
4. The underlying reason(s) why the event happened has been described: Comments	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
5. The impact or potential impact of the event has been described: Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
6. Reflection on the event has been demonstrated: Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
7. Learning from the event has been demonstrated: Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
8. Appropriate action has been taken (where relevant or feasible): Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
9. Where possible, appropriate individual(s) have been involved in the analysis of the significant event: Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Absent Poor Good Excellent
10. GLOBAL RATING SCALE *Please rate the overall analysis of this significant event:* Comments:	**1 2 3 4 5 6 7** ☐ ☐ ☐ ☐ ☐ ☐ ☐ Highly Unsatisf. Unsatisfactory Satisfactory Highly Satisfactory

Figure 5: Content Validity Index (CVI) for proposed SEA feedback instrument

Key to Feedback

1. Not relevant - **2.** Unable to assess relevance without item revision or item in need of such revision that it would no longer be relevant - **3.** Relevant but needs minor alteration - **4.** Very relevant and succinct

Please rate each of the proposed items using the scale 1-4 outlined in the above Key. If the score is 1, 2 or 3 then please type comment in space provided

1. The description of what actually happened:

1 ☐ 2 ☐ 3 ☐ 4 ☐

2. Appropriate timescales involved have been described:

1 ☐ 2 ☐ 3 ☐ 4 ☐

3. The role(s) of all individual(s) involved in the events has been described:

1 ☐ 2 ☐ 3 ☐ 4 ☐

4. The setting(s) where the event happened has been described:

1 ☐ 2 ☐ 3 ☐ 4 ☐

5. The underlying reason(s) why the event happened has been described:

1 ☐ 2 ☐ 3 ☐ 4 ☐

6. The impact or potential impact on individual(s) or organisation(s) has been described:

1 ☐ 2 ☐ 3 ☐ 4 ☐

7. Adequate reflection on the event has been demonstrated:

1 ☐ 2 ☐ 3 ☐ 4 ☐

8. Adequate learning from the event has been demonstrated:

1 ☐ 2 ☐ 3 ☐ 4 ☐

9. Appropriate action has been taken

1 ☐ 2 ☐ 3 ☐ 4 ☐

10. Appropriate individual(s) have been involved in the analysis of the significant event

1 ☐ 2 ☐ 3 ☐ 4 ☐

11. **Please rate the Content Validity of the Whole Instrument**

1 ☐ 2 ☐ 3 ☐ 4 ☐

Please note below any alterations or suggested items you may have:

CHAPTER 10

A study on the inter-rater reliability of the peer assessment instrument after further training of reviewers.

10.1 Introduction

In Chapter 9 the findings demonstrated that the content validity and overall reliability of the peer review instrument were acceptable. However, in order to further improve the reliability of the instrument, there was a need to reduce the variation amongst peer reviewers in their assessments of SEA reports. In this way a more consistent judgement could be made on the quality of a GP's SEA report.

Following a presentation on the findings reported in the previous chapter to the ADG at one of the quarterly training meetings, a discussion on possible methods to improve the agreement between reviewers on their assessment of SEA reports occurred within the group. After consideration of several suggestions from different members of the group, it was agreed to pilot three major changes to the way in which reviewers would be fed-back information on their assessment of the reports.

Firstly, all reviewers would take part in an exercise to assess five SEA reports submitted to the peer review model using the updated review instrument. The assessment scores given to the five reports by each member of the ADG would be recorded, anonymised and fed-back to individual assessors so that he or she could make a comparison of their own assessments with those of the group. This exercise would be repeated after three months.

Secondly, as each SEA report submitted to the model was assessed by at least two reviewers (Figure 3), each assessor would receive a copy of their co-reviewer's

numerical and written feedback on the SEA report. Both changes, therefore, involved reviewers being able to compare their own assessments with other members of the ADG.

A third agreed proposal involved reviewers being given a copy of the peer review coordinator's feedback letter to the GP participant.

One additional change was also made to the format of the assessment instrument. As the majority of individual scores of the SEA reports assessed in Chapter 9 were high (i.e. >4), this would imply that a number of the adjectival scale points for each item were being underused and were potentially obsolete. To take account of the possibility of 'redundant' descriptors, a decision was taken to modify the adjectival scale so that the majority of SEA reports would be of a quality judged to be in the middle of the scale. The adjectives on the seven-point scale were therefore adjusted so that more descriptors related to 'good' performance. This in turn is likely to increase true variance between the SEA reports (Streiner and Norman, 2003). The 'final' peer review and feedback instrument is shown in Figure 6.

A further study was set up to test the impact of these changes on the inter-rater reliability of the peer review assessments using the updated instrument.

10.2 Methods

10.2.1 *Sample size calculation*

Ten reviewers would need to assess more than 50 SEA reports in order to achieve an inter-rater reliability of 0.75, with a standard error (SE) of 0.05 (Streiner and Norman, 2003). This assumed a standard error of 0.05 amounts to setting the 95% confidence

interval at 0.1. This indicates that we were 95% confident that the inter-rater reliability co-efficient is between 0.65 and 0.85. An ICC of 0.9 with SE of 0.1 would require 10 raters to assess twenty reports. Discussion amongst reviewers at an ADG meeting suggested that assessing 40 SEA reports would be the maximum achievable. A pragmatic decision was therefore taken to ask 10 raters to assess forty reports.

10.2.2 *Pilot training and calibration exercise*

The updated assessment instrument was introduced at an ADG training day in January 2006. Two separate training and 'calibration' (a check on inter-rater reliability) exercises involved all ADG members. Each involved the assessment of five SEA reports. These were undertaken in February/March 2006 and April/May 2006. Results were fed-back to all ADG members and discussed at the quarterly group meetings one month following each exercise. Any individual assessment and feedback issues not dealt with at these fora were discussed with the peer review co-ordinator by e-mail or telephone.

10.2.3 *Participants and assessment exercise*

In September 2006, 10 randomly selected reviewers agreed to take part in the further reliability exercise. Each reviewer assessed the same 40 SEA projects which were randomly selected from reports submitted by GP principals to the peer review model between August 2002 and April 2006. The assessments were returned to the peer review coordinator within four weeks of the exercise start date. A reminder had been sent to those who had not returned the assessment instruments one week prior to the completion date.

10.2.4 *Data Analysis*

This was carried out as described in Chapter 9 with the exception that the variables included were SEA report (n=40), reviewers (n=10) and items (n=10). Only inter-rater agreement and internal consistency were calculated.

10.3 Results

The G co-efficients for the internal consistency and inter-rater reliability are shown in Table 21. Internal consistency of the instrument remained high when averaged over all items (0.958). However, item reliability of a single item remained <0.8, again indicating that no single item was a reliable indicator of the quality of the SEA report.

When compared with the previous study reported in Chapter 9, the inter-rater G co-efficients assessed using the average score over all ten raters (0.86 versus 0.64) and the global score (0.96 versus 0.43) had improved. Decision studies indicated that six raters would be required for the mean total item score to be deemed highly reliable (G>0.8), with four raters required for a satisfactory level for formative assessment (G>0.7). Two raters would be required for the global score to be deemed highly reliable (i.e. G>0.8)

There was no statistical difference in the overall mean global scores for SEA reports between the two exercises, with each mean score equating to the adjectival scale description of ‘good’ (Table 22).

10.4 Discussion

The main finding demonstrated that the inter-rater reliability of the SEA review instrument had improved from G=0.43 to G=0.96. These results strongly indicate that using two reviewers to assess each SEA report and report a global score would appear to have sufficient inter-assessor reliability for formative assessment purposes

(Downing, 2004). The alteration of the item descriptors to enable a wider range of responses from reviewers had not impacted statistically on the global assessment score given to the SEA reports.

10.4.1 *Limitations of study*

The reviewers selected for this study were chosen at random and may not have been representative of the whole peer review group. The SEA reports were also randomly selected from those submitted to the peer review model and may not have been indicative of the range of SEA reports submitted by GP principals. The other limitations applying to reliability testing detailed in Chapter 9 also apply to this study.

10.4.2 *Context and implications*

Although inter-rater reliability has been improved, it is not possible to quantify whether (and by how much) this was due to the different approaches to improving inter-rater reliability that the reviewers were subject to or whether general experience through assessing more SEA reports over time had an impact on their performance.

Although it is known that experienced reviewers will vary in their assessment of the same 'case' (Noel *et al.,* 1992), strategies to minimise this variation are often non-specific (Schuwirth and van der Vleuten, 2007).

One specific method relevant to general practice was the oral examination that, until recently, was undertaken as part of the assessment for the membership examination for the RCGP. By carefully selecting and then calibrating examiners through the use of video recordings of assessment performance, the correlation between examiners'

judgements was improved (Wakeford *et al*., 1995). Although variation between examiners' judgements was still apparent, despite ongoing calibration when the exam format was revised in 1998, it was a candidate's performance and not the examiners' judgements that was the major explanation of the variance (Wass *et al.,* 2003).
Immediate feedback to examiners (Demsky and Sellers, 1995), 'recalibration' of assessors in using the assessment tool (List *et al.,* 2006) and training by calibration against experienced assessors opinions (Bank *et al*., 2002) have been shown to be effective in improving the reliability of assessors in clinical examinations of patients. Involving peer reviewers in feedback sessions advising those responsible for setting the assessment tasks can also improve inter-rater reliability (Norman *et al.,* 1993).

However, training of examiners is not always successful. Training of reviewers did not demonstrate long-term improvement when tried with those assessing manuscripts for journal peer review (Schrotter, 2004).

Newble *et al*. (1980) suggested that one further method of improving inter-rater reliability is simply to remove more 'outlier' assessors from participation. Where reviewers exhibit contrasting but consistent behaviour, such as being 'hawkish' or 'dovish', these behaviours - unless very extreme - can be accounted for (Raymond *et al.,* 1991; McManus *et al*., 2006). The results from this and the previous study indicated that there were no obvious 'outliers' amongst the reviewers. Further calibration training beyond that routinely undertaken for the quarterly meetings of the ADG was therefore not considered necessary.

10.5 Conclusion

The reliability of the assessment instrument was of a sufficient level to be used as a formative tool in the peer review model. Reliability, however, is not a static process (Downing, 2004) and requires ongoing training and calibration of reviewers. These last two chapters have investigated aspects of the reliability and validity of the peer review instrument, which are acknowledged as two important aspects within the overall utility of the peer review model. However, the views and experiences of those who participate in the model, either as GPs submitting SEA reports, or as reviewers of their colleagues' work, require further investigation to add to the evidence base for the acceptability and educational impact of this model.

Figure 6: Final Peer Review Instrument

NHS
Education
for
Scotland

PEER REVIEW FEEDBACK INSTRUMENT

SIGNIFICANT EVENT ANALYSIS REPORT

Project Number

Project Title

INSTRUCTIONS FOR PEER REVIEWERS

Please use the attached tool to critically review and rate each relevant area of the SEA report. Feedback on how to improve the event analysis should be constructive and given in the comments section at the end of each relevant area. Similarly, where an area of the analysis has been undertaken well, please comment on this so it too can be given as positive feedback to the submitting GP. Please remember that all educational feedback should be specific, informative, sensitive, and directed towards improving the event analysis

Please rate the level of evidence contained in the audit report for each of the criteria listed overleaf (using the rating scale where 1=Very Poor and 7=Outstanding):

Project Reviewer……

What happened?

	1. Very Poor	2. Poor	3. Fair	4. Good	5. Very Good	6. Excellent	7. Outstanding	Not appropriate
1. The description of what actually happened:	☐	☐	☐	☐	☐	☐	☐	
Comments:								
2. The role(s) of all individual(s) involved in the events has been described:	☐	☐	☐	☐	☐	☐	☐	
Comments:								
3. The setting(s) where the event happened has been described:	☐	☐	☐	☐	☐	☐	☐	
Comments:								
4. The impact or potential impact of the event has been described:	☐	☐	☐	☐	☐	☐	☐	☐
Comments:								

Why did it happen?

	1. Very Poor	2. Poor	3. Fair	4. Good	5. Very Good	6. Excellent	7. Outstanding
5. The underlying reason(s) why the event happened has been described:	☐	☐	☐	☐	☐	☐	☐

Comments:

Reflection and Learning

	1. Very Poor	2. Poor	3. Fair	4. Good	5. Very Good	6. Excellent	7. Outstanding
6. Reflection on the event has been demonstrated:	☐	☐	☐	☐	☐	☐	☐

Comments:

	1. Very Poor	2. Poor	3. Fair	4. Good	5. Very Good	6. Excellent	7. Outstanding
7. Where possible, appropriate individual(s) have been involved in the analysis of the significant event:	☐	☐	☐	☐	☐	☐	☐

Comments:

8. Learning from the event has been demonstrated:	☐	☐	☐	☐	☐	☐	☐

Comments:

Appropriate Action Taken

	1. Very Poor	2. Poor	3. Fair	4. Good	5. Very Good	6. Excellent	7. Outstanding	Not appropriate
9. Appropriate action has been taken (where relevant or feasible):	☐	☐	☐	☐	☐	☐	☐	☐

Comments:

Please add any general comments

Global Rating Scale

	1. Very Poor	2. Poor	3. Fair	4. Good	5. Very Good	6. Excellent	7. Outstanding
10. Please rate the overall analysis of the significant event	☐	☐	☐	☐	☐	☐	☐

Table 21: Calculated reliability co-efficients for GP principals' SEA reports marked using the peer review instrument (expressed with 95% confidence intervals)

	Internal consistency	**Inter-rater**	**Overall**
Single Observation	0.72 (0.67-0.77)	0.67 (0.61-0.72)	0.6 (0.61-0.72)
Multiple observations (i.e. Average of 9 items X 10 raters)	0.958 (0.95-0.97)	0.86 (0.83-0.89)	0.93 (0.92-0.95)
Global score	NA	0.964 (0.95-0.97)	NA

Table 22: Comparison of the mean scores between GP principals and SEA reports Jan 05 and Sept 06

	Jan 05	**Sept 06**	**F Value**	**P value**
Mean score over items 1-9 (+/- 1S.D.)	4.81 (0.79)	3.70 (1.24)	0.08	P=0.96
Mean global score of SEA reports (+/- 1S.D)	4.85 (0.80)	3.66 (1.17)	0.612	P=0.98

CHAPTER 11

The acceptability of the SEA peer review model for trained reviewers of significant event analyses; a focus group study.

11.1 Introduction

By 2006, increasing numbers of GPs were participating in the peer review model, with some reporting that they had used the feedback received as supporting evidence in their annual appraisal. In addition, aligning external peer review with the GP appraisal system may be one way of adding rigour and educational value to the latter.

To examine the potentially wide range of emotional, professional and ethical issues that the role of peer reviewer could expose and be exposed to, and to add to the evidence base for the updated peer review model, this study set out to explore the views and experiences of those GP peer reviewers who were involved in making educational judgements on colleagues' work. This investigation specifically set out to focus on aspects of the acceptability of the peer review model for SEA reviewers.

11.2 Methods

11.2.1 *Study Sample*

At the time of this study, 21 GP reviewers formed the ADG. All were invited by email to participate in the study and written consent to participation was obtained prior to the interview (Appendix 8).

11.2.2 *Focus group interviews*

Three group interviews were undertaken lasting between 60 and 70 minutes. These were audio-taped and transcribed with permission. This author acted as moderator and assisted the discussion using a topic guide (Appendix 9) developed from feedback from

previous ADG meetings. The guide was modified as new issues arose. All participants contributed satisfactorily to discussions which were free-flowing and lively.

11.2 3 *Analysis of transcripts*

The author read the transcripts iteratively to identify broad themes prior to subsequent group sessions. Data were analysed for content independently by both the author and a GP research colleague. This involved systematically coding and categorising data and identifying principal themes. Categories and themes were modified by merging and linking them after joint discussion.

11.2.4 *Data validation*

Transcripts and a summary of our preliminary findings were sent to all participants for comment. No disagreements were raised. Further validation of data analysis was undertaken by another research colleague who read the transcripts against the themes and categories generated.

11.3 Results

20 of the 21 GP reviewers were able to take part in one of the focus groups. One group consisted of eight participants with the other two having six participants each.

Seven principle themes were identified:

- Purpose of feedback
- Volume and depth of written feedback
- Professional and legal implications
- Emotional issues
- Educational and practice benefits
- Training and calibration
- Format and content of feedback instrument

11.3.1 *Purpose of feedback*

There was agreement that feedback should offer encouragement to improve the quality of SEA reports if required. Tension was apparent between those who perceived SEA peer review as being a mainly formative educational process and those who saw their role as verifying a SEA standard. It was acknowledged that there was limited evidence available concerning the experiences of doctors submitting SEA reports and that they would have differing expectations of the type and depth of feedback.

> *"This all started because we were promoting change and I am not sure at the end of the day that this is still not the driving force in terms of the mass of doctors that we want to use this exercise as a way of promoting change and their own education."* (F1, M3)

> *"…the educators want it to be formative. I think probably the recipients would be quite happy for it to be a summative (judgement)."* (F1, M1)

For some reviewers the fact that they were originally trained to offer a "summative" judgement on criterion audits as part of GP registrar training (and still did at this time) and were now also giving formative feedback on SEA was a stark contrast. It was important that they were in the "mind-set" for formative feedback.

Many participants assumed that a significant number of doctors would submit their SEA report in the expectation of using the feedback as evidence for appraisal. Differing views on how the feedback should be used to complement the appraisal process were advanced. A particular contention was whether a dichotomous 'satisfactory' or 'unsatisfactory' judgement should be made on SEA reports. Some thought this was essential as GPs could use it to add rigour to their appraisal. Others re-iterated that appraisal is a confidential, non-judgmental process and thus satisfactory or unsatisfactory judgements were unhelpful. The purpose of their feedback was to highlight educational issues an appraisee may wish to discuss further in appraisal.

> *"Appraisal doesn't make a decision so we should."* (F2, F1)

> *"…we are not looking at it for satisfactory or unsatisfactory…we are looking at it to highlight areas that are weak - much more formative."* (F2, M2)

Some reviewers were unsure of the focus of their feedback and whether this should be on the doctor's SEA "technique" only, or whether they should also add their personal views on how the issues raised by an event analysis could be addressed.

> "*I was marking [assessing] the structure (of the SEA) whereas now it is more of a quality judgement on the whole SEA*." (F2, M3)

11.3.2 *Volume and depth of written feedback*

It was agreed that poorer quality SEA required more in-depth feedback. Consistent with previous research (Bowie *et al*., 2005b) participants identified learning and change issues as the most frequently occurring problem areas requiring the most feedback to submitting doctors. High quality SEA required less in-depth feedback because these invariably demonstrated effective event analyses.

> *"…if a person is scoring 5 or 6 [out of 7 on the rating scale] I tend not to say why. If they are scoring further down I would tend to say why…"* (F1, F3)

> *"…the poorer the significant event [analysis] the more difficult it is to mark the individual questions and to try and tease out what they have actually said, whether they address a particular question or not."* (F2, F1)

The volume and depth of the feedback provided by some reviewers reflected their own personal preferences for this level of feedback. Although some felt this should be brief, others countered that appropriate recognition of colleagues' efforts with in-depth feedback should be given to all.

> *"If I submitted an SEA that was fine and I had to read a page telling me it was fine I would be really annoyed."* (F3, F2)

> *"Someone may have sweated over the SEA and are dying to hear back… what is acceptable for them may be different than for you."* (F3, M3)

Participants were aware that the final feedback report given to submitting GPs is an edited version of reviewers' comments combined with the peer review coordinator's input. A few were aware that not all of their comments would be given as verbatim feedback to the submitting doctor. They felt that this allowed a degree of openness on their real thoughts but also security in knowing that these would be put into more diplomatic language when summarised. In contrast, the instructions on the feedback instrument state that comments should be constructive and fair. This highlighted a training issue to be discussed at the ADG group.

> *"…we are putting down exactly what we think about that aspect of the significant event analysis and then you are interpreting that in a way that is digestible."*
> (F2, M4)

Practical issues also influenced the type of feedback given. All participants are GP partners and none has specific "sessions" set aside for peer review. They have to make time to undertake reviews and the volume and depth of their feedback could reflect their competing work and social pressures.

> *"It depends on how many things I am chasing on that particular day."*
> – general agreement in group. (F3, M1)

11.3.3 *Professional and legal implications*

Some participants expressed concerns around their professional responsibilities as a peer reviewer. Of particular concern were SEA reports outlining serious patient safety issues where the analysis was limited by poor professional insight, inadequate understanding or the ineffective implementation of change.

"…what if Dr X did something terrible and it has all been covered up and it comes out in the SEA? What do you do?" (F3, M4)

"…the SEA – where you look at what conclusions have been drawn and you go whoa, whoa – you just really missed the point here." (F3, M2)

These issues had previously been discussed at length in ADG meetings. For some it was clear that the duties of a doctor outlined in the General Medical Council's (GMC) guidance *Good Medical Practice* (GMC, 2006) gave appropriate direction on how to respond. The issues around poor performance and possible GMC referral clearly continued to make some reviewers uncomfortable. One raised the issue of whether participating doctors should be informed of areas felt to be inappropriate such as substandard care by other professionals and events which could lead to disciplinary action.

"…it could be applied on a whole range of levels from the trivial to the massive, and the legal for instance, which is something I assume we are not going to drift into - areas which might be used in court if it were to get serious." (F3, M4)

11.3.4 *Emotional Issues*

A number of emotional issues were raised. Despite the system being confidential for participating doctor and the reviewer, there was apprehension for some about whether their feedback was truly anonymous. While they appreciated that this was superficially the case, there was a feeling that the system had never been truly tested for anonymity. Some participants were concerned that a doctor referred for poor performance may wish to see individual feedback reports as well as the summarised feedback letter. In

contrast, others were happy to have their feedback quotations openly attributed to them.

> *"…you have got to be careful about our comments; the person may want a very significant discussion about the comment."* (F1, M4)

> *"I don't mind being quoted."* (F3, F2)

SEA reports in which the submitting doctor had appeared to make little effort caused frustration for all. These tended to be very brief and superficial. While brevity could be an acknowledged virtue, it was only relevant if the analysis was effective. There was agreement that these reports reflected poorly on the doctor's professionalism and demonstrated a lack of commitment to the educational ethos of SEA and peer review. Many reviewers were puzzled as to why these doctors would participate in such a voluntary exercise. Conversely, there was reluctance by some to rate these SEA reports as 'very poor' as they felt that going to the effort of writing and submitting a report was a positive attribute.

> *"You sometimes feel the report you are going to give back is going to take longer than some of the poor ones take to do!"* (F1, F3)

> *"When I started marking things I thought I am going to mark things, most people up – I feel quite sympathetic, but then I found myself thinking – what the hell did you submit this for? What was the point?"* (F3, M2)

> *"…part of the reason we don't give out [feedback] 'rubbish' is because we have a feeling for the person getting the feedback and we also have experience that says – this person has gone as far as producing this report – already they are not at the lowest end of the scale."* (F2, M1)

Many reviewers did not use the extreme ends of the feedback instrument scale [very poor or excellent]. This is a well-recognised phenomenon in the use of assessment instruments (Striener and Norman, 2003). One reason highlighted and generally agreed was the perceived reluctance of Scots in general to offer high praise for achievements.

> "*…in Scotland we never say that something is outstanding!*" (F3, M2)

Making a "satisfactory" or "unsatisfactory" judgement on the SEA caused division. Many found little difficulty in what they perceive as a low stakes judgement, while others were more troubled by the prospect of being 'black and white'.

> *"…sometimes we are asked to give a satisfactory/unsatisfactory judgement - sometimes I do not want to do that, I find it difficult sometimes but accept sometimes I may be asked and as long as written feedback is sensitively and appropriately handled, that is okay."* (F1, M4)

11.3.5 *Educational and practice benefits*

Most participants agreed that their role as a peer reviewer was a positive personal learning experience. One perceived learning issue described was in being able to apply the SEA technique much more effectively as a result of being a reviewer. Being subject to initial training in SEA and use of the instrument combined with in-depth exposure to a

large volume of SEA reports which they had assessed, provided them with a greater knowledge and understanding of how to apply the technique and in particular what 'not to do'.

Several participants expressed the opinion that the learning accrued from the analysis of events was at a 'deeper' level than that which they had previously experienced from reading about patient safety events incidents described, for example, in medical defence union publications or general medical journals.

> "…*instead of there but for the grace of God go I… I enquire as to how the doctor plans to prevent this happening and think how I would do it in our practice."*
> (F1, M2)

Reviewers in all three groups were able to describe examples of changes that had been implemented in their own practices as a result of reading SEA reports.

> "M*y staff must love it when I have been reviewing SEA… they [SEA reports] are stimulating and I often review and sometimes change things or just check the system we have is working.*" (F2, F2)

Other reviewers described using the SEA reports as templates for their own teaching of GP registrars either in formal one-to-one tutorial sessions, GP registrar group day release programmes or in informal single case reviews. They were perceived to be especially useful in prompting registrars to consider the difficulty of implementing change in practice and dealing with other members of the primary care team or health professionals outwith the immediate practice team.

However, there were some conflicting opinions about the learning associated with being a peer reviewer of SEA. A few reviewers did not agree that assessing SEA reports resulted in substantial learning. Although in the 'early days' [i.e. first five years] of being a reviewer they found it a positive learning experience, this had become less frequent and more superficial. This was attributed to the common origin of many significant events, such as communication issues, or the relatively straightforward and 'low challenge' analysis described.

> *"…when we first started assessing reports most of it was interesting and you could learn things but not now, no, not really…well only very occasionally.*"
> (F3, M3)

11.3.6 *Training and calibration*

Initial feedback training was viewed as either adequate or more than necessary. There was agreement that although it was appropriate to train new reviewers, it was the calibration process which was a priority for more experienced reviewers. Calibration had previously been described to participants at ADG meetings as a process to maximise levels of agreement between reviewers. This was perceived to be an essential part of a reviewer's professional responsibility to themselves on a personal level and to the submitting doctors.

> *"…have enough training, just need recalibrated once in a while".* (F2, M2)

> *"I think if you are taking on responsibility, you are being responsible for other people's learning then you are duty bound to recalibrate yourself."* (F1, F3)

The three calibration formats used (formal assessment of SEA reports by the entire group with individual and group feedback, comparison of assessments with co-reviewers and reading of the peer review co-ordinator's feedback) were seen as valuable and pragmatic in the circumstances. Participants perceived group discussions at quarterly meetings to be helpful in highlighting specific difficulties for reviewers. Examples included obtaining the views of fellow reviewers on what actually constituted a 'significant event' and gaining consensus on the need for SEA reports to explicitly detail the impact or potential impact of the event, which was often absent.

> *"…helps collective learning – you know and discuss out the issues of why one person saw it one way and why another saw it another way."* (F2, F2)

The use of previous SEA reports (with consent) for calibration was helpful. Reviewers would assess these individually and then subsequently share their assessment with the group.

> *"…useful to see how other people are marking [assessing] as well – it's not just me being a hawk!"* (F1, M4)

There was general agreement that receiving 'feedback on your own feedback' was a positive step. This is achieved by sending each reviewer copies of their own and their co-reviewer's feedback, as well as the summarised feedback letter sent to the participating doctor. This allowed for personal self-reflection and also gave reviewers the opportunity to query the appropriateness of the coordinator's final feedback.

"…the very best thing educationally for me is when you get the feedback and you see what your co-marker has marked and you see how the comments have been collated into the final document …without a shadow of doubt." (F2, M2)

"it is interesting to get your feedback coming to us because if we then disagree with actually what you have said - I think we will let you know." (F2, M3)

11.3.7 *Format and content of feedback instrument*

In the twelve-month period prior to the study, the reviewers had been using two feedback instruments concurrently - the 'original' instrument (Figure 2) and the updated instrument (Figure 6). The latter was preferred as it was felt it "*encourages comments*" and allowed more targeted feedback, despite the additional time this took to do.

"…makes us think more clearly and specifically what the feedback is about." (F1, F1)

"I guess in some ways it is a bit more hassle because you have to consider and write a bit more." (F2, F2)

Several reviewers thought there may be a difficulty in applying the new instrument to a "positive" event analysis, such as the early detection and successful treatment of a cancer. These types of events may not necessarily fit well with the review instrument. Others found no difficulty in applying the instrument for this purpose, as long as reports were in the recommended format. Allowances were made by reviewers to take account of the unlikely requirement to implement change for these cases.

Difficulties understanding, interpreting and applying some of the feedback instrument's ten measurable items were raised. Two items - demonstrating 'reflection' and 'adequate learning' - were identified as requiring further discussion. There were conflicting views on the application of the instrument's rating scale, for example in the adjectival descriptors and the use of a seven-point scale. However, it was accepted that no instrument can be perfect and that the existing training and calibration mechanisms could overcome any difficulties.

> "*You know, discuss out the issues at the group of why one person saw it one way and why another person saw it another way.*" (F2, M4)

Participants acknowledged the need to undertake ongoing validity and reliability testing of the instrument and felt that their efforts in this undertaking were acknowledged. They appreciated that their contribution to developing the instrument and the wider peer review model were sought and felt that this enhanced a sense of ownership in the process

> "*It keeps us interested as part of the process and it was good to hear you acknowledge our experience and input"* – general agreement (F3, M4)

11.4 Discussion

For this group of external peer reviewers the opportunity to give feedback to their colleagues on their SEA attempts is seen as an important professional undertaking. It was surprising given the ongoing training and calibration of reviewers that some provided feedback only on the "process" of SEA rather than using their clinical knowledge and experience to suggest solutions to significant events. Indeed, as the

purpose of formative peer review is to improve the quality of a given piece of work (Rennie, 2003), it is essential that all involved are given clear guidance on the purpose of the assessment.

The study also highlights the difficulty for reviewers if they are asked to make two concurrent judgments – one on the quality of the SEA report and the second on whether it reaches a "satisfactory" standard. The emotional difficulty encountered by some in this (formative) situation highlights a problem that may be encountered in a higher stakes assessment where judgments may be "*marked up"* (Hay, 1995). As well as compromising the validity and reliability of ratings (Norcini, 2003), this is an issue which would require to be addressed if GP appraisal is to embrace both formative and summative judgements.

It is known that doctors submitting their work for this kind of external peer review are stepping outside the relatively "safe" environment of appraisal or practice-based discussion and exposing their work to a potentially more 'threatening' review (Bowie *et al.,* 2005a; McMillan and Cameron, 2006). This engenders in the reviewers a professional respect for their colleagues that manifests itself as a desire to maintain their own skills as peer reviewers to avoid being overly critical of presented work and to be constructive in their comments. A negative effect of the high expectation of colleagues was the frustration expressed by reviewers at those doctors who demonstrated little professionalism in their efforts.

The findings show that some reviewers tailored feedback according to their own perception of what the doctor wished from the feedback process. This was particularly apparent for high quality submissions where there was disagreement on the type and

volume of feedback. Simple comments rather than descriptions of why the work was 'good' were deemed appropriate by some reviewers. Effective formative feedback is recognised as crucial to the role of peer review as a teaching and learning exercise (Nicol and McFarlane-Dick, 2006). Therefore, despite the group having had previous discussions on the attributes of 'good' feedback practice (Ende, 1983), this may represent a training issue for some. Alternatively, the reviewers may be reflecting the reality that motivation to take part in this form of peer review for some doctors is primarily for an endorsement rather than for written formative feedback (Nicol and McFarlane-Dick, 2006).

As discussed in Chapter 9, an ideal educational tool would be "supportive and individualised, yet uniformly applied" (Norman *et al*., 2004). To this end, on–going training and feedback to reviewers on their performance is an essential component of the peer review process (Norcini, 2003). These activities serve to act as a quality assurance and governance marker in the system. Using multiple systems of training and calibration was reported by reviewers to enhance their own learning and subsequent feedback skills. This would support the argument that peer review is an applied skill honed through experiential learning (Davidoff, 2004) and that short didactic educational interventions are probably unlikely to succeed (Schrotter *et al.,* 2004). Most reviewers were reassured by the ongoing training that their assessments were generally not at variance with their co-reviewers. Where this occurred, they could reflect on and learn from it, which should also help improve the reliability of the peer review process.

With no specified "protected" time for assessment, reviewers admit that they can become harassed at their work or home with many other competing issues requiring

attention that can affect the quality of peer review. This is a complex issue as the idea of 'protected' time was not supported by all reviewers and reflects the need to balance the cost-benefits in any system. The educational model described in this study allowed for a small payment to be made to the reviewers. We did not explore the practical and financial implications if the relatively small number of SEA reports reviewed annually was overtaken by a "compulsory" system for all doctors which would generate higher volumes of work.

The concerns of reviewers regarding the feedback instrument were ameliorated by an appreciation that they had important input into the development of the instrument and were taking part in ongoing validity and reliability testing. This helps engender a sense of ownership in the process and an appreciation that a feedback instrument will always contain imperfections.

The positive educational impact for many reviewers is perceived as a personal benefit to participation in this model. Whether this compensated for any negative workload and emotional implications was not explored. For those who found that the educational benefits of reviewing had diminished as a result of assessing many similar reports, it is important that such 'SEA fatigue' does not lead to a 'blasé' attitude to reviewing or a less discriminating outlook. It perhaps offers the opportunity for the peer review coordinator to identify and offer further learning experiences, such as educational sessions on issues identified by group members, for example critical thinking skills.

An issue of concern for some reviewers was the potential ramifications of SEA reports that exposed poor performance for themselves, the submitting doctor and the patient. This is despite an agreement with the Medical and Dental Defence Union of Scotland

(MDDUS) on the appropriate action for reviewers to take where SEA reports raised questions of 'fitness to practice' or 'severe patient harm' (see Appendices 10a and 10b). These actions were also accepted by both a local and a national research ethics committee. There were also persistent misgivings and concerns around anonymity for reviewers, despite the highly confidential nature of the system. This mirrors the findings of previous work with certain doctors submitting SEA reports into this system who, despite reassurances regarding anonymity, were still reticent to submit selective reports that resulted in a poor outcome for a patient as a result of their actions (Bowie *et al.,* 2005a).

11.4.1 *Strengths and limitations of the study*

With one exception, all peer reviewers took part in the study. The fact that the participants were both experienced in peer review and all knew each other well allowed for free dialogue, interaction and frank discussion. A limitation is that not all issues identified by groups could be fully explored in the time available, constraints on time and resources militated against further group sessions being reconvened.

The groups were moderated by this author, the peer review coordinator, whose presence may have introduced bias and limited the critical nature of discussions. The author is a founder member of the group who is responsible for convening meetings, training reviewers and drawing up contracts. He may have been seen as having 'seniority' over 'ordinary' members. An independent moderator was not employed as it was felt important that practical issues relating to the peer review model raised by study participants would require a stakeholder with an understanding of the system. The findings need to be viewed in this context.

11.5 Conclusion

This study reports on the professional, workload, educational and training issues that concern peer reviewers in this particular external feedback model. The findings suggest that the model is broadly acceptable to reviewers since participation is viewed positively by most. The importance placed on peer review as a professional duty and the positive educational impact for most reviewers are positive factors that should help inform the feasibility and acceptability of other peer review models for general practice.

However, if external peer review is to be offered as an option in a modified appraisal and revalidation system, it is important that the doctors who undertake reviews are able to make evaluations that are universally and consistently applied. The training of peer reviewers therefore needs to acknowledge the difficulties, emotions and tensions experienced when making professional judgements on aspects of colleagues' work.

CHAPTER 12

The acceptability to and educational impact on participants of the peer feedback model for significant event analyses.

12.1 Background and aims.

Although participation in the original SEA peer feedback model was thought by participants to potentially offer reassurance to agencies that the GP had participated in independent, external appraisal of aspects of their personal practice (Bowie, 2005a), difficulties were also apparent. Evidence of the feasibility and acceptability of this model was limited by selectivity of events due to concerns about confidentiality, litigation and professional embarrassment. As described in Chapter 9, the value of educational feedback provided was also thought to be minimal for those SEA reports that were judged to be of a 'satisfactory' standard or where the significant event described 'good practice' rather than an actual or potential adverse occurrence.

To address some of these criticisms, changes in the SEA assessment instrument, as described in Chapters 9 and 10, were implemented and evaluated. In addition, changes were made to the training and calibration of reviewers (Chapters 10 and11). The format and content of the feedback letter sent to participants was also altered. Previously, feedback on 'satisfactory' SEA reports was limited to a sentence confirming an acceptable analysis. A more substantial narrative based on an amalgamation of the two reviewers' assessments on the adjectival scales of the updated instrument and their free-text comments was introduced to give more specific detail to the participating doctor (an example of an SEA report and feedback is given in appendix 11).

In view of these changes it was decided to investigate the views and experiences of GPs who had participated in the updated peer review model for SEA. This study specifically aimed to explore aspects of the acceptability and educational impact of the review process and the feedback received.

12.2 Method

12.2.1 *Study sample*

54 GPs from the west of Scotland had participated in the revised peer feedback model for SEA between July 2006 and February 2007, with 24 submitting SEA reports in the previous six months. Participants were sampled from this group of 24 by their level of identified learning need. This was determined by the global score on the SEA feedback instrument, which employs a continuous rating scale where 1 is 'very poor' and 7 is 'excellent'. The level of learning need is inextricably linked to the rating given (Chapter 11). We wished to ensure that we included a range of participants given the full spectrum of feedback. Potential participants were sub-divided into three groups by rating score: 1-2 (very poor-poor; group 1), 3-4 (fair-good; group 2), and 5-7 (very good to outstanding; group 3). GPs in each group were invited by e-mail or telephone to participate in the study in chronological order of submission.

12.2.2 *Data collection*

Nine one-to-one semi-structured interviews were undertaken by this author. The interviews, which were audio-taped and transcribed with permission (see Appendix 12), took place in each GP's place of work and lasted between 50 and 70 minutes. A topic guide developed from issues raised in previous literature (Hewison and Little, 1998; Bowie *et al.,* 2005a) was used to assist the interview process (see Appendix 13). Data saturation was achieved after seven interviews as no new themes emerged thereafter.

12.2.3 *Analysis of transcripts*

The author read the transcripts iteratively to identify themes prior to subsequent interviews and modified the interview schedule accordingly. Data were analysed for content independently by the author and a GP research colleague. This involved systematically coding and categorising data and identifying principal themes. Categories and themes were modified by merging and linking them after joint discussion.

12.2.4 *Data validation*

Transcripts and a summary of initial findings were sent to participants for comment. None was raised. Data validation was enhanced by a third researcher who reviewed the transcripts and cross checked the identified themes and categories.

12.3 Results

Nine GPs were interviewed and eight principle themes were identified:

- The submission process
- Perception of benefit
- Selectivity of events
- Peer review organisation and peer reviewers
- Format of feedback and timescale
- Learning and change
- Role in supporting educational appraisal
- Sharing learning through the peer review model

The themes generated were applicable to all GP groups, unless otherwise stated.

12.3.1 *The submission process*

All participants were aware of the standard SEA report format recommended for submission to the peer review model (which was unsurprising given it is also the recommended SEA 'template' for GP appraisal in NHS Scotland). There was consensus that the template allowed most types of significant event to be analysed and written up to highlight learning issues and demonstrate change, where this was required.

> *"I suddenly realised what the importance of having a very clear structure to it and a very clear chain for these kind of events. Involved more than I thought to do it properly."* (M1, group 2)

A practical issue raised by some participants was that the suggested report format was difficult to access and download through the NES web site and most had to access the template either through an alternative website or request a hard copy. It was thought that this had the potential to reduce the 'enthusiasm' of some GPs to participate.

The routine cost of submitting a report for peer review is £25, payable by the submitting doctor. Some GPs had taken part in an 'Educational Partnership' initiative that allowed participants to purchase unlimited SEA/audit/video peer reviews for £125.

There were contrasting views as to the cost involved. The majority thought that the sum involved was relatively minor and was "*value for money*", while several others thought that it was an "*expensive*" form of education, particularly if sending in multiple reports, where the feedback was similar for each SEA report.

> "Unless I was in the partnership I wouldn't send in a pile – put it that way." (F2, group 1)

The importance of confidentiality was emphasised by one participant who warned:

> *"if people aren't sure it's confidential they are not going to write anything that is too challenging."* (M2, group 3*)*

There was a general assumption amongst participants that the system was confidential and indeed an expectation that because it was a "peer review" system, the coordinating body and individual peer reviewers would ensure this as a professional responsibility.

12.3.2 *Perception of benefit*

Several participants reported that going through the feedback process formalised the SEA and provided reassurance from peers that the action taken as a result of the event was adequate. Others suggested that the act of participation would enhance their appraisal and give validation to their work or provide an additional opinion on how the significant event could have been handled by themselves and the practice team.

Most participants agreed that periodically submitting SEA reports for peer feedback would be of educational benefit, although how regular this should be was not discussed. A contrary opinion was expressed by one participant who felt that repeated submissions were unnecessary, and that as long as the doctor could demonstrate competence in the SEA technique, then repeated submissions were of limited value (F3, group2).

12.3.3 *Selectivity of Events*

There was disagreement among participants as to what type of significant events they would submit for external review. Some participants admitted that although there was a 'no-blame' culture in their practices and they accepted that the peer review model was both confidential and educational, they were still uncomfortable with analysing and then submitting events that reflected poorly on themselves or their practice professionally. However, others were more equivocal, appreciating that peer review could offer support even in a difficult situation, while several participants stated that serious significant events should be given priority for submission.

> *" If you don't send in the really bad stuff you are missing the whole point – imagine if you were a pilot disguising your SEA –'oh never talk about that one – the wheel fell off, never talk about the wheels falling off.' In fact the really bad ones is when you benefit most."* (M3, group2)

Selectivity of events was also influenced by the contractual requirement for GPs to identify and analyse specific types of significant events, such as a new cancer diagnosis or a suicide. There was a consensus that it was simpler to submit these reports for peer review than to analyse and submit a different event. This was despite general agreement that the analysis of 'prescribed' significant events was lacking in educational value when compared to self-selected events.

There was also a perception amongst some participants that "positive" significant events (those which demonstrate exceptional practice and that are worth sharing) would not be encouraged for peer review, despite awareness that these could contain learning points for the practice and others.

"People are very reluctant to put things forward that they think they have done well probably because I think you don't want them to think you're 'bumming' [bragging]." (M1, group 3)

12.3.4 *Peer Review Organisation and Peer Reviewers*

Although participants were content to submit their SEA reports to NES for peer review, several raised the issue of the relevance of the organisation to themselves and colleagues. Different perceptions of the professional status of the peer reviewers were also apparent. A few participants were aware of the peer review group and their normal professional roles. The others either assumed that reviewers were a group of "part-time academics" or admitted giving "little thought" to who was reviewing their work. When it emerged that reviewers were largely 'frontline' GPs who were trained and experienced in giving peer feedback, sceptical participants found this encouraging and reassuring.

"There was a feeling (within the health centre) that there was a barrier there, perceptions of an ivory tower – it's not seen as true 'peer' review." (M2, group 2)

12.3.5 *Format of feedback and timescale*

There was overwhelming agreement that the SEA feedback format was appropriate. Other alternatives such as face-to-face feedback or a combination of written and face-to-face feedback were considered impracticable. Participants were generally satisfied with a turnaround time of approximately six weeks to receive feedback from peers. Beyond this timescale was considered a problem because the SEA had probably lost its emotional impact, educational relevance and would not be considered a "priority for action". Given the frequency of events, it was considered likely that other more pressing concerns for the practice would eventually take precedence.

In discussing the feedback instrument, there was general agreement that the inclusion of a numerical global rating as part of the feedback was not a priority. Some believed this should not be shared with participants as additional feedback would be required to explain the global score against the appropriate descriptor – *"Does a '3' mean 'fair'?"* (F1, group 1). Others agreed that they had not taken part in the exercise for the purpose of receiving a "score" which carried connotations of a "summative" judgement.

The majority of participants thought that receiving a 'low score' would be demoralising even if constructive written feedback was given and that this was likely to act as a barrier to taking part in future external peer review.

> "*I would rather be told something was 'fair' rather than a three out of seven.*" (F4, group 2)

> *"Not sure if worth extra work [giving numerical values] – because you are looking at how to do it differently next time rather than how to get a score – nice if you get a good one you might be disappointed if you get bad one."*
> *(F2, group 3)*

Conflicting views were also apparent on whether an overall dichotomous 'satisfactory' or 'unsatisfactory' judgement on SEA report(s) should be made. Some wished to know this 'judgment call' so they could be provided with a professional assurance of the quality of the SEA. In turn, this could provide evidence, if required, for their appraisers, GP partners, clinical governance leads and the wider public that they had reached a set "standard". The others were ambivalent about this aspect of peer judgement or had no

interest in such judgments as this approach conflicted with their perception of peer review as a purely educational and non-judgmental exercise.

> *"It's not just the public it is anyone outside the practice. It is just a little bit too cosy to just mention your errors to one person and expect that to necessarily promote things forward… I can see for the general public you know, if you want things to have any kind of rigour then you want outside comments."*
>
> (M4, group1)

12.3.6 *Learning and change*

There was variation in the perceived educational gain from the review process. For some this was related to their initial perception of what peer review would offer. All agreed that the feedback narrative was strong at highlighting best practice on how to write-up an SEA report, but some expected more to be offered in the way of solutions or alternative methods to addressing the significant event under scrutiny. The others thought that the feedback did address this where it was appropriate. There was also a recognition that if the SEA was well-written and had identified appropriate issues around learning and change, then there was little additional input that the reviewers could provide, other than confirm good practice in the application of technique and in analysing the specific event in question. There was agreement amongst most that comments which highlighted areas for improvement were much more useful than solely 'positive' and 'encouraging' feedback. Some believed, however, that where there is in-depth feedback that outlines a number of learning needs, which can be perceived as 'negative' feedback, it can *"cause your hackles to raise a bit"* (M3, group 2). Indeed one interviewee (F3, group 1) had written back to the peer feedback coordinator to disagree with the reviewers' comments (a highly unusual occurrence).

"…it [feedback] would have been more useful if I hadn't done one before…The comments were more about my writing up of the SEA rather than how we could change all different things - want suggestions, not for the way you laid out SEA but for ways of improving or avoiding event happening again." (F1, group 3)

"There were points raised, you know, of dealing with things I hadn't thought of." (M2, group 2)

A small majority of participants were prepared to share the feedback with other practice colleagues to highlight learning and change issues that had not been previously considered by the team. They reported that where the suggested change was thought to be appropriate and practical, this was taken on board by the team and implemented to improve practice.

12.3.7 *Role in supporting educational appraisal*

Participants who used their peer feedback to support their appraisal did so because they thought that it demonstrated to their appraiser that they were reflecting and learning at a high level. The feedback was reported by some to form the basis of discussions between appraiser and appraisee on clinical issues not directly related to the SEA in question. Others reported that their appraiser also seemed to look on an SEA which attracted positive feedback as a 'stamp of approval'. In addition, gaining external peer feedback was thought by some to offer a much more frank assessment of the educational issues involved in a significant event than would be received from appraisers or colleagues.

*"Appraisers might not be quite as honest and quite as frank as someone completely independent." (*M1 group 3)

12.3.8 *Sharing learning points from SEA reports*

The majority of participants agreed that other GPs could learn from at least some of their SEA reports that had been submitted for peer review and vice-versa. However, there was little enthusiasm for sharing these reports with other doctors or for receiving information on their colleagues' SEA. A few participants thought that because of the manifest differences between individual practices, the ability to learn and change from others' SEA attempts was limited, while there was general agreement that they were time-limited in terms of reading the SEA reports of other practices and that it would not be a priority.

There was also conflict over potential confidentiality issues between those participants who preferred the peer review system to be coordinated nationally and those who would opt for a local system. It was perceived by some participants that local systems would not be popular with some doctors who had experienced a practice split or worked in a multi-practice health centre where it may be relatively easy for others GPs to be aware of where the SEA report had originated. The majority thought that a national system would not be as relevant or would contain too wide a spectrum of events to identify reports where the learning points could be applied to their own practice

"*I don't know probably would be worried though. I wouldn't do without discussing with my partners.. Senior partner wouldn't like baggage from previous partnership splits."* (F2, group 1)

12.4 Discussion

12.4.1 *Main findings*

For this group of GPs taking part in the updated model, this small study demonstrates that it is broadly acceptable and had positive educational outcomes. Submitting work for external feedback is seen as an acceptable form of learning and complementary to the appraisal process. The findings highlight areas for improving peer feedback and the SEA learning experience for others.

12.4.2 *Strengths and limitations of study*

To obtain a more comprehensive picture of the acceptability and educational impact of the model we would need to gain insights from larger GP samples, including non-participants as well as other professional GP groups. Interviews were conducted by the peer review coordinator instead of an independent researcher, which may have introduced bias and limited critical discussion. However, as with the focus group interviews in Chapter 11, an interviewer possessing an in-depth understanding of the political, emotional and practical issues relating to the model was judged to be of high importance. Participants may have submitted their SEA report up to six months prior to the interview and thus recall bias may have influenced their responses.

12.4.3 *Context and implications of findings*

Peer review is recognised as an important educational process (Helfer, 1972; Grol, 1995). For UK family doctors this mostly takes place within the relatively “safe” environment of practice-based discussion or one-to-one appraisal (Scottish Executive, 2003). Informal discussions with colleagues can often be unstructured and subjective (Bradley, 1992; Grol, 1995) whereas doctors who submit SEA for external peer review are exposing themselves to a ‘more threatening’ review, which is external, independent

and aims to be objective. Despite this, our findings suggest this type of external formative review of SEA is feasible and potentially acceptable as part of a doctor's CPD, provided participants are assured as to the confidentiality of the system, can access the system easily, and perceive a low personal financial cost to participating. The feasibility of using such a system as more than a personal learning exercise, for instance to act as a patient safety tool, is likely to be limited by the selectivity of events chosen for submission by some of the doctors, the limited follow-up of action points suggested by reviewers and the lack of desire to share their own and read others' SEA reports.

Although the purpose of feedback is most often to highlight learning needs and facilitate improvement (van der Ridder *et al.,* 2008), the finding from our study - that some participants were primarily interested in receiving confirmation that their SEA was of an acceptable standard - was of particular interest. For these GPs, this gave external verification to the quality of their work to be used in a future appraisal interview and as evidence for potential revalidation. However, this may give false reassurance to the GP and others, as both the feedback instrument and the overall model facilitates a formative review rather than summative assessment (Chapter 9). In contrast, others who take part in this model, in common with participants in other peer review models (Ringstead *et al.,* 2004), do not think that documentation of competence was an important issue. These conflicting perceptions on the value of peer judgements, in addition to the peer feedback, presents a difficulty for this model. In mitigation however, the exercise is a purely formative review and is promoted as such.

An advantage of this model is that learning is not compromised by having a prescribed list of events that has to be analysed irrespective of perceived learning opportunity - as

is the case with their contractual obligations (DoH, 2006). Nor do events analysed have to involve patient harm, as is the case with significant events reported to the National Patient Safety Agency (NPSA, 2005). However, all three systems are likely to be limited by personal and recall bias. It is also likely that educational opportunities are lost in this model because some GPs avoid submitting particularly sensitive events. These potentially deleterious effects on the educational value to participants need to be considered within the overall context of the feasibility and acceptability of the model.

The importance of the acceptability of NES as an organisation and the professional status of the peer reviewers was highlighted. As discussed in Chapters 3 and 9, although there is debate on who constitutes a true 'peer', it was apparent that participants valued the fact that the reviewers were active 'frontline' practitioners. Conversely, this led to disappointment for some that although their feedback was informative on 'how to' carry out and analyse a significant event, some of their peers did not use their own knowledge of practice to suggest alternative solutions to problems raised by the significant event. This may be related, at least in part, to the peer reviewers' interpretation of the "feedback rules" and can be addressed through further education and training of the reviewers.

This peer review model was thought by participants to be complementary to the appraisal process. Appraisal as practised in Scotland at the time of this study was (and still is) an educational, reflective, non-judgemental discussion between the GP and a fellow independently nominated GP appraiser (Scottish Executive, 2003). There is an assumption that this process to help GPs self reflect on their professional practice will improve patient care – although the evidence for this is limited (Driessen *et al.,* 2008). Participation in external feedback prior to their appraisal means that the efforts of GPs

are scrutinised by doctors who have specific skills in SEA feedback beyond those of their appraisers who are not specifically trained in providing this. Although learning can be driven by taking part in a relevant assessment exercise, this model's combination of external peer review and facilitated feedback should enhance the educational process (Sargeant *et al.,* 2005). Since SEA is a 'core' component of GP appraisal in Scotland, participation in this activity must be demonstrated, even if the GP has more pressing educational needs. The finding that the peer review process released time for the GP and appraiser to discuss other issues arising from the significant event itself or move on to other identified educational issues is an additional benefit of participation.

12.5 Conclusion

The findings add to the evidence for the acceptability of the external peer review model as an activity for GPs and the associated educational gain of participation. Aligning the external peer review of SEA with appraisal is arguably one method of adding rigour and educational value. The findings also highlight areas of concern amongst GPs in sharing SEA reports to learn from and improve their professional practice and enhance patient safety.

CHAPTER 13

The feasibility of using NHS audit specialists in the external peer review model of SEA: a pilot study.

13.1 Introduction

At the time of this study in 2006 it was likely that objective evidence of performance would be required to support the annual appraisal of GPs (DoH, 2006). Although the system had yet to be established for determining how and what evidence would be gathered and verified, the updated model of peer review had the potential of offering a feasible and acceptable solution for making judgments on the quality of some of the necessary evidence that was likely to be presented to appraisers.

If judgements on the quality of evidence were to be made, then a recognised challenge would be in meeting the demand for review and managing the associated costs and resources supporting the system

From both a criterion audit and an SEA perspective there is potential to support the review of this activity from existing resources within the NHS. The great majority of NHS organisations – across all health sectors – typically employ a small operational team of experienced (mainly but not exclusively non-clinical) clinical audit specialists. It is a core responsibility of these specialists to support, facilitate and advise health care teams on all aspects of the clinical audit process, including significant event analysis. These individuals are expected to be skilled and knowledgeable in audit methods and the management of change. Given these circumstances, there is clear potential for clinical audit specialists to become involved in the review of SEA and provision of educational feedback in support of this core category of the appraisal system.

However, there is a lack of evidence supporting whether non-medical professionals are as competent as medical peers at judging the quality of any clinical audit project undertaken by medical practitioners. In recognition of this, a journal communication put forward the case for NHS clinical audit specialists being as competent as trained GP assessors at assessing the audit projects of GP registrars (Dolman *et al.,* 2005). Taken at face value this would appear to many with a professional interest in this area to be a reasonable proposition. However, the study methods employed and the interpretation of preliminary data reported appeared to be flawed, mainly because GP audit assessors are trained and experienced in the assessment process, whereas clinical audit professionals (although specialists) have not had this advantage.

The submission of SEA reports by GPs as part of our peer review model does, however, provide a useful opportunity to test the assumption made that clinical audit specialists are as competent as medical peer assessors at formally assessing this form of audit activity. As part of a study to quantify any differences between clinical audit specialists and experienced medical peer assessors in their assessment of two types of audit (criterion audit and SEA) their scoring of SEA reports was examined using the revised feedback schedule detailed in Chapter 10 (Figure 6).

13.2 Methods

13.2.1 *Study participants and setting*

This pilot study was undertaken involving a comparison of the assessment outcomes of SEA reports by two small groups (one experienced and one novice).

Group one (the experienced peer assessors) consisted of a convenience sample of 11 GP audit assessors voluntarily recruited from the previously described Audit

Development Group of 21 GPs based in the west of Scotland. Group two (the novice assessors) consisted of a convenience sample of ten specialists with a minimum of four-years' front-line experience of advising health care professionals in clinical audit, who were recruited by email from across all health care sectors in NHS Scotland. The email containing an explanatory note of the study was sent out by NHS Quality Improvement Scotland (a national health authority responsible for patient care quality and safety initiatives), which has a database of 156 known clinical audit specialists employed in NHS Scotland. Positive email replies were received from 27 audit specialists, and ten individuals were randomly selected and approached to participate. All possessed the minimum experience required and voluntarily consented to enrolment in the study.

13.2.2 *Feedback instrument and study timescale:*

SEA reports were formally assessed by the members of two study groups using the previously described instrument (Figure 6). Each group was given a four-week period to complete this task in their own personal time.

13.2.3 *Power calculation*

It was determined that using 12 significant event analysis (SEA) reports to assess agreement between groups would have just under 90% power to detect a standardised difference of 1 in the global judgement score, at the 5% significance level, using a two sample t-test. An average difference between scores of 2 would be relevant, and an average difference of 1 or less unimportant. If the standard deviation of the differences between the groups is small (similar to the difference deemed relevant from an assessment comparison) then even this small pilot study will be sufficiently powered to determine whether the novice Audit Specialist Group give results similar to the

experienced Audit Development Group. If the pilot study is not sufficiently powered then it will provide the information for an informed power calculation for a further study.

13.2.4 *SEA reports being assessed*

The 12 SEA reports were chosen at random from those voluntarily submitted for peer review over a five-year period between 2002 and 2006 by GPs as part of their continuing professional development. A total of 833 SEA reports were submitted in this period in standard report formats. Brief details of the significant events analyses selected for the study are outlined in Box 8.

13.2.5 *Statistical analysis*

The mean scores for the SEA reports were then compared for both groups. Confidence intervals (95%) were calculated to quantify any difference (i.e. one group marking consistently higher than the other) in the scoring between the groups and 95% limits of agreement calculated to determine whether the two groups were interchangeable.

13.2.6 *Training of clinical audit specialists*

Prior to the study, the novice assessor group received three hours training in the assessment of simulated scenarios and 'real-life' examples of criterion audit and SEA reports using the appropriate instruments. This training was delivered by the author and a colleague who was experienced in clinical audit and in related educational assessment and research.

13.3 Results

Eleven experienced GP audit assessors (Group 1) and 10 novice assessors (Group 2) rated the 12 SEA reports.

13.3.1 *Assessment of SEA Reports*

The mean assessment scores for each of the two groups and the differences in means for all 12 SEA projects are outlined in Table 23. The difference in mean scores between both groups for all projects was less than 1.0. No group consistently rated SEA projects higher or lower than the other group. The 95% confidence interval for bias is -0.1 to 0.5 (p=0.14 i.e., the null hypothesis for no difference between the groups was upheld). This indicates that the Audit Specialist Group, after the provided training, is capable of assessing as consistently as the experienced Audit Development Group. The 95% limits of agreement between the groups ranged from -0.7 to 1.2, which is safely within the 2-point limit agreed prior to start of the study.

The relatively close comparison in SEA assessment scores between groups is of particular interest as most members of the Audit Specialist Group had limited exposure to SEA prior to the study when compared with their experience of criterion audit.

13.4 Discussion

13.4.1 *Main findings*

This study provides sufficient information to suggest that a sample of NHS Clinical Audit Specialists can give numerically accurate formative feedback scores as part of the peer review model described, which are generally comparable with the ratings of experienced medical peer assessors with respect to SEA reports.

13.4.2 *Limitations of the study*

This small study focused on only a single numerical aspect of assessment. Further evaluation of the audit specialists' written educational feedback would be required to provide more evidence of their ability to participate as assessors in this model. However, as audit specialists, they should by definition be skilled in identifying deficiencies in the application of audit methods by health care professionals and providing educational solutions.

For the purposes of this study, the Audit Development Group was accorded 'expert' status and their assessments were judged to be closer to the 'truth' than those of the Clinical Audit Specialists. In reality, the opposite may be true.

A larger study involving a greater range of examples of SEA reports for assessment would provide stronger evidence to confirm the findings, but is unlikely to have a substantial influence on the limits of agreement. These are primarily dependent on the standard deviation of the differences between the groups, which is not a function of sample size.

13.4.3 *Context / Implications*

A pragmatic decision was taken that a *consistent* mean difference of more than 2 points would raise serious issues regarding the viability of the novice Audit Specialist Group, with the level of training given, to reliably undertake these assessments. In reality, when this situation does occur in the existing model of peer review, a further assessment is undertaken independently by two experienced reviewers who would then jointly agree a consensus rating score and provide appropriate developmental feedback. If a mean score differences of between 1.1 and 1.9 arose, it was felt that this would raise issues

which could be adequately dealt with through further training and calibration of the Audit Specialist Group, as is the case in the day-to-day operation of the model. Consistent mean scores of 1.0 or less (as were found here) would not be considered important from an assessment perspective and may indicate an ability to adequately assess clinical audit activity to a level comparable with the 'gold standard' medical peer group.

Experience of this form of external peer review in general practice is more advanced in Scotland than the other UK countries. However, although there is limited information on its acceptability, feasibility and educational impact, there is recognition that stronger evidence is required if policy-makers are to be convinced of its value, particularly in linking it with appraisal as a means of providing objective comment on performance.

13.5 Conclusion

In summary, this study supports further exploration of the feasibility of trained NHS audit specialists providing formative educational feedback to medical practitioners on the quality of their SEA reports, if this were to be required. It is clear that this specialist group also has the potential to support audit in general, and SEA in particular, as part of the peer review model through arrangements for CPD in the region. Although not strictly 'peers' as defined in this thesis, they have specialist knowledge in audit methods, advising on the undertaking and reviewing of clinical audit projects and could perhaps be thought of as 'audit peers'. If stronger evidence is forthcoming, then issues over the acceptability of using what many GPs would consider a 'non-peer', and the potential service implications, would require investigation.

Box 8: Identification number and brief description of individual SEA reports

No.	Significant Event Analysis
1.	Complaint – Refusal to complete a sickness benefit claim form
2.	Confusion over status of referral letters
3.	Upper GI Bleed in an elderly lady on aspirin and NSAID
4.	Verbal complaint over appointment waiting time
5.	Delayed medical reports
6.	Cardiac arrest – problem with defibrillator data card
7.	Potentially fatal overdose
8.	Failure to inform patient of positive chlamydia result
9.	Amiodarone blood monitoring
10.	Interface communication failure
11.	Complaint over treatment of minor wound
12.	Complaint over delayed clinic appointment

Table 23: Significant event analysis judgements: Mean scores (range 1-7) and standard deviations by Assessment Group for individual SEA reports

Assessment Group	**SEA Report Identification Number (1-12)**												
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***Mean (SD)***
Group 1 Audit Development Group (n=11)	4.6	1.9	3.8	3.2	2.5	2.4	4.1	4.1	5.2	1.1	4.7	4.3	3.5 (1.3)
Group 2 NHS Clinical Audit Specialists (n=10)	4.0	2.1	3.2	3.5	2.3	2.6	3.2	3.2	4.9	1.4	4.1	4.6	3.3 (1.0)
Differences in Means **Group 1 - Group 2**	+0.6	-0.2	+0.6	-0.3	+0.2	-0.2	-0.9	+0.9	+0.3	-0.3	+0.6	-.3	0.2 (0.5)

CHAPTER 14

Discussion

14.1 Summary of main findings

The studies undertaken in this thesis have contributed to the existing body of knowledge and understanding of the potential roles of significant events and their analyses in general medical practice. In addition, new evidence has contributed towards the utility of a peer review SEA model.

14.1.1 *Significant events: identification, grading and reporting*

The evidence from the two questionnaire surveys (Chapters 6 and 7) suggest that promoting the identification and analysis of significant events as a method to engage GPs in incident reporting systems will be challenging. The first survey found that three-quarters of GP principals indicated a willingness to take part in a local anonymised reporting system for identified adverse events. However, a large majority of GP principals did not support mandatory reporting and would be selective if this were the case. These findings provide evidence for two possiblel obstacles for potential reporting systems in primary care.

If incident reporting systems are to be of value in general practice, then the finding in Chapter 6 of a potential difficulty for a sizeable minority of GPs in determining when an incident is ‘significant’ needs to be addressed. This would be important as deciding when an event is sufficiently ‘significant’ enough to merit formal notification to the appropriate NHS organisation may cause uncertainty and confusion for GPs. This difficulty applied particularly to those who were not principals in a GP training practice. The second survey - undertaken after the introduction of the nGMS contract –

highlighted that these concerns about identifying and grading 'significance' may be misplaced. The overwhelming majority of GP respondents were able to appropriately identify six event scenarios regarded as 'significant.' In this study, there was no difference between GPs of different educational groupings in grading the severity of significant event scenarios. Overall, these GPs were able to distinguish less serious significant events from those perceived to be more severe. This could be of importance for local or national systems where consistency of advice on what should be reported may be required to avoid both under- and over-reporting. Perhaps not surprisingly, the participants agreed that events graded as being of 'high severity' offer more learning opportunities to other primary healthcare team members than those perceived to be of lesser importance.

The preparedness of GPs to report selected significant event scenarios was associated with their grading 'severity'. The more 'significant' the event grading, the more likely GPs would be to report this to an anonymous and confidential reporting system.

A possible further difficulty for NHS Scotland healthcare authorities, professionals and regulators in developing incident reporting and learning systems is highlighted by the clear preference for events to be reported to NES. Exempting this option, just over half of GPs reported a preparedness to notify even those events thought to be 'highly significant' to either a local medical committee (LMC) or a national reporting system.

14.1.2 *Examination of SEA reports submitted for peer review*

The content analysis of SEA reports demonstrated that the broad range of subjects chosen for analysis were similar to those reportedly undertaken in other general practice studies (Pringle *et al*., 2005; Cox and Holden, 2007). The wide variety of

events identified add to the evidence that SEA is a complementary way to examine areas of healthcare not easily investigated by the orthodox cohort and criterion audit methods.

The majority of SEA reports outlined the potential for patients to come to harm as a direct result of their interaction with the healthcare system (with one in four describing actual harm). This suggests that GPs are prepared to confront potentially difficult issues that may not always reflect well on either themselves or their practice. The association between the possible 'severity' of an event and preparedness to analyse such a scenario is supported by the findings in the second survey. This demonstrated that an analysis was more likely to be undertaken for those scenarios graded to be most 'significant.' As the vast majority of event analyses reported that learning needs had been identified and appropriate change implemented, then there is potential to inform and educate others in patient safety related issues.

The full potential of SEA to fulfil its proposed 'learning' role is likely to be restricted by three issues identified from the studies. Firstly, there is a lack of engagement of GPs with existing reporting systems, as described previously. Secondly, the reported learning was predominantly related to personal professional issues. This is undoubtedly important and worthwhile but there was limited evidence of issues being widely shared with other healthcare professionals. Thirdly, the reticence of some GPs interviewed (Chapter 12) to share SEA either locally or nationally with other healthcare teams was expressed. Fear of exposure and criticism from other practices contributed to this reluctance. There was also a perceived lack of relevance of SEA reports from other practices. A further factor was the prioritisation within professional practice whereby the 'cost' of reading other reports outweighed the 'opportunity.'

For the scenarios given in the questionnaire study (Chapter 7), GPs were willing to involve relevant practice team members in any analysis that was to be undertaken. However, if the event in question involved patient harm then non-medical staff were less likely to be involved in the learning and change as a result of the analysis. In addition, where issues relating to learning and change were identified by peer review, only some GPs would feed this back to the practice team for further consideration (Chapter 12).

14.1.3 *Peer review of SEA reports*

The findings suggested that the updated model of peer review was considered professionally acceptable to the majority of GPs. Participants reported that the feedback from peers had a positive educational outcome, particularly in imparting technical knowledge on how to analyse a significant event in a structured way. There were differing perceptions on the level of learning achieved between GPs on other aspects of received feedback. Some participants thought that the feedback was appropriate and enhanced their understanding of how to analyse a significant event. In addition, it could also highlight matters or issues not previously considered as part of the analysis. For some participants there was an expectation of more emphasis being placed on how the significant event itself could have been avoided or addressed. Alternative suggestions from reviewers as to how they would have used their own experiences in general practice to improve outcomes from the analysis would have enhanced the perceived value of the feedback. This also highlighted potential training issues for reviewers.

Focus group findings suggest that the breadth and depth of feedback on individual SEA reports varied between reviewers. Despite being aware of the recommendations for good feedback practice, some chose to offer comment based on their own perception of

what constituted appropriate levels of information. This is likely to have influenced the quality of advice for the participants. Indeed, reviewers considered it essential to have regular training, preferably using several different methods. This was thought to augment or maintain their feedback skills. Participants submitting SEA reports to the system found it reassuring that reviewers were regularly trained and updated. The explanation that reviewers were also active GPs led to a perception of them as 'true peers', which served to increase the acceptability of the model.

There were different motivating factors for participants in deciding to submit their SEA reports for review. Some were primarily concerned with receiving confirmation that their SEA was of an acceptable standard for appraisal purposes and/or contractual requirements. Others either had no interest in such a judgment being made or gave it a much lower priority. Disagreement was also apparent amongst reviewers on whether a 'validation' judgement should be offered or routinely given to participants. Some participants and reviewers thought that a dichotomous 'satisfactory/unsatisfactory' decision on the report analysis was useful. This would provide reassurance to appraisers and, if necessary, regulatory bodies and the public that the GP could properly apply the SEA technique in this instance. However, a substantial minority of reviewers would be uncomfortable about making such a decision. This was partly because they thought that this would marginalise and overshadow the intended formative educational nature of the process. It could also engender a negative emotional feeling within themselves.

Overall, participants viewed the model as being complementary to the appraisal process. The issues raised in the feedback letter were used either as a platform for further discussion of significant events with the appraiser or as a way of demonstrating

that they had participated satisfactorily in a ‘core category.’ This was perceived to act as a stimulus to the interview being developed to encompass other educational issues. For some participants, participation in peer review was thought of as contributing an added degree of rigour to the appraisal process.

Peer reviewers undertake their task primarily because they value this activity as an important professional undertaking. In addition, they valued their own part in developing the model, such as having input into the content and format of the review instrument. For some members, undertaking peer review itself was an educational process. However, despite guidelines given to the group, there was still concern expressed about their own professional and legal obligations to colleagues and to patients seriously harmed as a result of significant events.

In terms of the psychometric development of the peer feedback instrument, an acceptable level of content validity was established (see Chapter 9). One finding also suggested a ‘good’ level of reliability for the instrument in differentiating between the varying qualities of SEA reports but highlighted the scope for increasing inter-rater reliability. The follow up study, conducted after further reviewer training, demonstrated an improved reliability coefficient. These findings taken together strongly supported the use of two independent reviewers using the feedback instrument as having sufficient reliability for the formative feedback to GPs on their SEA reports.

The updated peer review instrument was found to be more acceptable to reviewers, despite requiring an increased time input from some. There was a general acknowledgement that the new version led to more specific feedback. In addition, the

updated instrument could be more easily applied to the review of all significant events analyses rather than just those outlining sub-optimal care.

A further study demonstrated that a small number of clinical audit specialists were able to review SEA reports comparably with members of the ADG when using the adjectival scale on the peer review instrument. However, the numerical assessment of a report is only one small part of the overall model. Although the input of such a group is not required at present, this finding does raise the potential for an alternative group of SEA reviewers if the feasibility of the current peer review model were to be affected by a significant increase in participation in the future.

14.2 Limitations of the study methods and the peer review model

Many of the advantages and disadvantages of the different research methods applied in this study have been discussed in the earlier chapters. Within the overall context of the thesis, it is acknowledged that various professional, contractual and policy issues evolved throughout the timescale of the studies and as a result had implications for the findings uncovered.

For example, by 2002 both local and national incident reporting and learning systems were becoming an expectation in England and Wales (DoH, 2001a; Harrison *et al*., 2002). However, local incident reporting in primary care in Scotland has remained poorly developed (NHS QIS, 2008), if implemented at all, particularly in general practice. Indeed, there are no immediate plans to implement a national incident reporting system for Scotland (Burns, personal communication, 2007). As such, the questionnaire surveys undertaken required GPs to express attitudes and opinions on reporting systems that did not exist or were either largely unknown or poorly engaged

with. As with all questionnaire surveys, the opinions expressed do not necessarily reflect what would be the respondents' behaviour at that particular point in time or in the future.

Two factors that were likely to have implications for the findings were the introduction of the nGMS contract and the implementation of a national GP appraisal system for Scotland. The questionnaire survey described in Chapter 6 was undertaken prior to the introduction of both these initiatives, while the survey discussed in Chapter 7 was carried out afterwards. It is highly likely that these initiatives resulted in increased participation of both GPs and their teams in SEA. Consequently, this may have influenced the attitudes and abilities of GPs with regard to the analysis and reporting of significant event. However, this aspect was not explored.

In the study reporting on the evaluation of the contents of SEA reports, findings depend at least partly on the ability of participating GPs to articulate a clear and convincing narrative outlining their competence in SEA. No attempt was made by this researcher to confirm the veracity of the levels of patient harm, or the learning and changes reportedly implemented by participants in the SEA reports. An assumption is made, based on the professional status of the submitting practitioner, that information given is factual and represents a consensus of opinion if more than one individual has participated in the analysis.

A criticism of the studies described in Chapters 7-13 is that GPs were often asked to describe or interpret 'learning'. Although a simple description of learning was given in Chapter 2.2.3 (Pitts, 2007), there are many theories of what constitutes adult learning in relation to medical education. No single theory is comprehensive (Merriam and

Caffarella, 1999; Kaufmann *et al*., 2000). SEA for individuals and their teams, could be seen to involve aspects of reflective practice (Schon, 1983), self-directed learning (Candy, 1991) and experiential learning (Kolb, 1984). The outcome of a learning process for individuals varies, depending on their existing knowledge (Spencer, 2003) and learning styles (Newble and Entwistle, 1986). No learning process or event is likely to mean the same to two individuals. Educational gain has also often been used synonymously with learning throughout the thesis. It is acknowledged that although some researchers define these terms as being interchangeable, others view them as distinct entities (Garavan, 2007).

Several studies related to the reliability of the peer review model, which, as previously described, could be viewed as a 'less important' aspect of utility for a formative feedback system (McKinley *et al*., 2001). While this is acknowledged, the reliability of the review instrument and its associated use by reviewers was considered crucial to inform the overall feasibility, acceptability and educational impact of the model.

The peer review model itself is not designed to assess whether a practitioner is competent or not in the SEA technique. Although the study described in Chapter 8 involved a review of SEA reports judged as 'satisfactory' by the trained peers, this was an assessment undertaken for research purposes. While participants in the model may ask whether their report(s) are judged as 'satisfactory' or not ,neither they nor an appraiser can take this as a reflection on their competence in SEA. Such assessments would require further 'cut-point' studies of multiple SEA report submissions, which may be a topic for further research.

14.3 Implications and findings within current context.

14.3.1 *SEA and the quality and safety of patient care.*

Prior to the start of this thesis, the SEA technique had been proposed as one method, amongst many, that could potentially contribute to improving the quality and safety of health care (Pringle *et al*., 1995; Harrison *et al*., 2002; Stead *et al.,* 2001; Sweeney, 2003). Following the commencement of this research, it was introduced as an optional contractual duty for GPs on the basis that it was an educational audit activity which promoted reflective learning. Evidence for its impact on the quality and safety of patient care was limited, but there was an assumption that SEA would facilitate learning and consequently lead to improvement in care (GMS, 2004).

SEA was also introduced in 2003 as a 'core category' in GP appraisal in Scotland based on its perceived strength as a reflective learning technique which would demonstrate participation in one aspect of Good Medical Practice (Scottish Executive *et al*., 2003). Furthermore, it was recommended in primary care by the NPSA in 2005 as a quality improvement technique to be undertaken when incidents occurred in which patient safety was compromised and/or low to moderate harm had occurred (NPSA, 2005). Again, there was an expectation that patient care would be improved as a consequence of undertaking an event analysis. This was despite there being very limited evidence for its effectiveness or impact on improving the quality and safety of patient care at this time (Chapter 2) or subsequently (Bowie *et al*., 2008).

The breadth of healthcare and professional issues identified by GPs and their teams through SEA may be particularly important given that the subject matter for criterion audit is likely to be dictated by the financial imperatives of the QOF. Audit of patient care in common clinical topics such as diabetes and asthma (as assessed using

surrogate outcome measures) has been shown to improve with the introduction of the nGMS contract (Campbell *et al*., 2007). This is likely to relate at least in part to the substantial financial incentive associated with chronic disease management in the QOF. Arguably, for some it has come at a cost to practice dynamics and the nature of the doctor-patient interaction (Campbell *et al*., 2008) and could potentially compromise care of conditions not covered by the framework. SEA has the advantage of not being as limited to specified areas of care and may serve to highlight deficiencies or potential problems in healthcare not covered by this aspect of the GP contract. However, improvement in patient care brought about by SEA is likely to be more difficult to quantify. For SEA, 'surrogate' markers of good care, such as improved team work, personal professional knowledge and behaviour, as well as individual and team learning, are more humanistic qualities that do not lend themselves easily to quantitative outcomes.

Whether GPs should be directed to analyse more serious significant events is open to debate. It is known both from a previous study (Murie and McGhee, 2003) and the finding in Chapter 8 that analyses of incidents which highlight positive aspects of healthcare can also lead to learning and change. The QOF 'marker' events, many of which will demonstrate positive aspects of care, are often not perceived as educationally valuable to individual practitioners or to the practice in improving the quality of patient care (see Chapter 12). The opportunity-cost of reviewing specified events therefore requires evaluation. For example, incident reporting systems and other clinical governance requirements may wish to encourage GPs to analyse and report events which result in specified levels of harm. It may be that the 'marker' events in the QOF should be reviewed to focus on areas and aspects of care where things are known

to go wrong, such as medication errors (Esmail *et al.,* 2004). This may maximise learning opportunities.

The findings do provide some evidence that a majority of GPs may use SEA as a method to identify and address learning needs. What is not known is whether this learning is 'deep learning' that can lead to a prolonged understanding of the issues highlighted by the SEA, or 'superficial' learning that is more likely to be forgotten once the analysis is completed (Marton and Saljo, 1976). There is still, however, a paucity of evidence on whether changes reported as being implemented as a result of SEA actually take place. For example, the most common change as a result of SEA, described in Chapter 8, was the introduction of protocols particularly for administrative systems. Whether these were implemented and followed up within practice is not known. There is also no evidence on whether any improvements are sustained. This was also a substantial difficulty with the initial introduction of criterion audit, where relatively few audits ever completed a full cycle of two data collections with even smaller numbers demonstrating a third data collection (Nelson, 1976).

A verification process to ensure the implementation and monitoring of suggested changes in practice is likely to be difficult to achieve. It may be that primary care organisations through QOF review visits could corroborate whether or not reported changes are in place. In practice this is likely to be difficult to do. The reviewers may not have sufficient time and it could be difficult to observe change in what is a snapshot of practice on a particular day. In addition, without training in SEA, reviewers' judgements on the analysis of and outcomes from reports are likely to be inconsistent and not uniformly applied. This in turn could potentially lead to unfairness in financial remuneration. Additionally, a reviewer may sanction a change that is inappropriate.

The need for verification may also be seen as an attack on the 'professionalism' and probity of the GPs involved.

The evidence base described in Chapter 2 suggested that SEA could be used as a method for team-based learning. However, the findings from the studies in Chapters 7 and 8 demonstrated a variation in the participation of the practice team in SEA. There was clear evidence that GPs would be prepared to analyse the significant event scenarios described in Chapter 7 with appropriate members of the primary care team. However, the content analysis of reports did not find evidence of widespread team learning, particularly for those events that involved patient harm. It is known that participation in SEA can be hindered by professional embarrassment, fear of litigation (Bowie *et al*., 2005a; Lord and Lillis, 2005) and poor team dynamics (Fox *et al*., 2001). One explanation for the apparent difference in findings in these two chapters may be one of the limitations of the peer review model. GP authors of the SEA reports may simply have forgotten or omitted to mention all the individuals involved in the analysis.

The value of SEA as a learning and change activity is also likely to be overly emphasised from the results in this study. It is known that one third of SEA reports submitted to the peer review model did not reach a satisfactory standard of analysis when judged by their trained colleagues. There is an assumption that GPs can apply the SEA technique when there is evidence that this is not the case for a substantial minority who have participated in the peer review model (Bowie *et al*., 2005b). This mirrors the assumption when medical audit was first introduced to healthcare that medical professionals could apply the technique. One subsequent review of the audit research evidence base demonstrated that this was not necessarily correct (Johnstone

et al., 2000). Indeed, evidence from the woS peer review model also demonstrated that GPs had difficulty applying one technique for criterion audit (McKay *et al*., 2006).

14.3.2 *Linking SEA with incident reporting systems*

As described in Chapter 1, the overall role of incident reporting in improving patient safety in healthcare is a matter of debate. Where such reporting systems did exist, SEA was promoted as one way of facilitating engagement with general practice. At this time there was also deliberation as to whether reporting by healthcare professionals to such systems should be voluntary or mandatory. As previously highlighted, some researchers view all reporting systems as being essentially voluntary (Billings, 1998; Barach, 2000). The evidence from this thesis would add weight to this opinion. In the study reported in Chapter 6, GPs replied that they would be selective in the choice of significant events that they would notify to a local reporting system. In addition, there is emerging evidence from the structured review of patient records that some incidents of patient harm in general practice go unrecognised and are potentially preventable (De Wet and Bowie, in press). Reporting systems will therefore invariably be incomplete as databases of potential and actual harm to patients, irrespective of what level of engagement is achieved.

While acknowledging the limitations in identification and reporting of events, the findings in Chapter 7 suggest that it may be possible to develop guidance on common significant events or specified outcomes from an event that should be analysed by a practice and reported to a local healthcare organisation. In a similar way to the existing system in England and Wales, those of a particular severity could be reported nationally if such a system existed. Consistency in reporting is important so that advice can be given on what to report and what not to report (thus avoiding under and over reporting).

Esmail (2006) suggested that general practice lacks a 'reporting culture' and thus engagement with such systems is likely to remain poor, at least in the short term. He argues that reporting models where notifications are categorised quantitatively rather than exploring the narrative of the learning are unlikely to appeal to "*non corporate structures"* such as general practices. The educational value of narrative text in reports submitted to such systems is also highlighted by Vincent (2004). He emphasised that it is learning associated with the incident rather than the number and consistency of reports that matter. It is, therefore, perhaps unsurprising that the majority of GPs surveyed in the study reported in Chapter 7 would prefer to report to an educational organisation. The provision of feedback to the person who has submitted the report is also known to be important (Firth Cozens *et al*., 2003) since lack of action from reporting bodies is known to dissuade people from engaging with such systems. Clinicians may be more likely to report to agencies that do not have contractual or financial leverage over them. However, it is notable that appraisers and training practice GPs were as prepared to report to NES as other GPs. This is despite NES being either their employer (in the case of appraisers) or making decisions on their training status (in the case of training practices). In addition, to whom and to where a GP would choose to report an incident are also likely to be influenced by other factors. For example, the accessibility of the reporting form, the level of detail required and the ease of the submission process could all affect the willingness to participate (Battles *et al*., 1998).

Further, in clinical medicine there can be a "*cultural censorship*" on incident reporting (Hart and Hazelgrove, 2001). The authors describe this as the finding that within an organisation an adverse event or problem can be simultaneously known about but also concealed. Problems of consensus on the issues, "*bonds of transgression*" (where through mutual self interest doctors will cover up deficiencies in one another's care),

and a failure to resolve the issue all contribute. Whether this would be the case in general practice is not apparent as this subject does not appear to have been widely investigated. Furthermore, incident reporting can also have negative consequences for some individuals within an organisation (Firth Cozens *et al*., 2003). This may be particularly difficult in a small general practice team where reporting an event could leave open the possibility of being traced back to an individual through its relatively unique set of circumstances.

The issue(s) around who has 'ownership' of significant events were not explored in any of the studies undertaken. This could have a strong influence over who analyses and reports incidents**.** Many events may involve several different healthcare professionals, in different systems with varying degrees of involvement in the incident. Indeed, some of the events analyses studied as part of this thesis originated in and were largely concerned with quality and safety of care issues in secondary care. The analyses of such events may not necessarily take account of all those involved. Conversely, those involved in the analysis discussion may not have been party to the actual event. It may be hampered by hierarchical structures within the practice that may influence findings (Hillier, 2002). In addition, the incident will often involve a patient who subsequently has no input to the analysis.

The optimal strategy to improve the quality and safety of patient care has been described as being through a multiple-method approach (Olsen *et al*., 2007). For the foreseeable future, SEA will remain as one quality improvement method in general practice through its role in nGMS and GP appraisal. A potential additional role in local incident reporting systems also exists. In general practice, other methods such as criterion audit and analysis of complaints are also relatively well-established. As well as

these methods, root cause analysis (RCA) is recommended in primary care for identified events that result in severe harm or death to patients (NPSA, 2005). Despite being well-known in both the acute and mental health sectors as a retrospective patient safety technique in secondary care, the feasibility and acceptability of RCA by general practice teams is not known. Other emerging quality improvement techniques such as trigger tools (Resar *et al*., 2003), care bundles (Institute for Healthcare Improvement, 2009) and real-time patient safety audits (Ursprung *et al*., 2005) are likely to be studied for their potential role in general practice. These techniques were originally developed for use in industry. They were recently adapted for healthcare settings in the United States (www.IHI.org) and have subsequently been used as part of patient safety initiatives in the UK, mainly in secondary care (NHS Modernisation Agency, 2004; Scottish Patient Safety Programme, 2007). The evidence for their uptake and value to general practice is poorly researched because of the relatively recent introduction of these methods. An evidence base is important since one study in secondary care reported that doctors' and nurses' ideas on how to achieve quality in patient care are quite distinct from the techniques endorsed by 'quality experts' (Hudelson *et al.,* 2008). It is likely that whatever the technical approach taken to quality improvement, it is the attitudes and behaviours of healthcare staff that will ultimately determine success or failure (Powell et al., 2009).

14.3.3 *The potential role of the SEA peer review model*

There is an inherent assumption that peer review will help to identify unmet educational needs (Grol, 1988). The converse is that these would otherwise remain unchecked by health professionals or hidden in a self–reflective process. As previously discussed in Chapter 3, the effectiveness of such feedback on professional learning and clinical practice is limited. Peer review does have the potential to be a quality improvement

technique and combining this with SEA could enhance the latter's ability to deliver improvements in the care of patients. The research and evaluation of the SEA peer review model was undertaken to examine and subsequently strengthen its utility as an educational process for GPs. In this way, individual participants, professional bodies and healthcare authorities could be informed about the evidence base for some of the main strengths and limitations of the model. By making the model as scientifically robust as possible it was perceived that this would enhance its credibility as an educational activity.

Although it is now a contractual duty, SEA as part of appraisal does not currently have the 'high stakes' status that would be associated with future revalidation. There is evidence that some GPs viewed (falsely) a positive feedback report as providing evidence of satisfactory participation in their core category (Chapter 12). This perhaps emphasised the confusion for some GPs in attempting to understand the place of appraisal within the context of the ongoing discussions and developments in the revalidation of doctors.

This is not surprising. In 2006 the CMO for England recommended that appraisal should "*make explicit judgement about performance against the generic standards*" (Donaldson, 2006). Subsequently, the white paper '*Trust, Assurance and Safety'* (DoH, 2007) put appraisal at the core of revalidation, and advised that it should be both "*formative and summative*." The RCGP, in their discussion on the '*Principals of GP Appraisal',* propose that appraisers will make a decision on the quality of evidence submitted by the GP (RCGP, 2008d). This judgement will be made in relation to explicit, although as yet undetermined, criteria and standards when revalidation does come into force. How this will be done has not been specified. Participation in SEA is likely to

contribute to the evidence presented by a GP for recertification. It is likely that five SEA reports will be submitted as part of the appraisal evidence (RCGP, 2008e). Although this may increase the quantity of SEA reports, it does not necessarily address the issue of the quality of analyses.

The peer review model is one way through which a judgment on the quality of a doctor's work can be given. Potential advantages of this model are that the judgement is external to both the appraisee and appraiser and can be placed in the context of the limitations of the system. Whether peer reviewers will be prepared to make dichotomous 'satisfactory/unsatisfactory' judgements if required in a modified appraisal system is not clear. Accepting the task of peer review is an important professional duty. Yet, models such as the one described may be limited by the number of GP reviewers willing to make a 'decision' on a fellow professional's work. Making such an assessment is currently a difficult task for some reviewers in what is a low stakes, formative process. A further issue is the need to consider whether there are any associated legal and ethical implications for reviewers. In addition, if external peer review such as that described here is to be offered as an option in a modified appraisal system, it is important that the doctors who undertake reviews are able to make evaluations that are consistently applied. The training of peer reviewers, including appraisers, needs therefore to acknowledge the difficulties, emotions and tensions experienced when making professional judgements on aspects of a colleague's work.

A further benefit of such an external model would be that regulatory authorities and the public can be assured that this is not a 'self–assessment.' This may be of particular importance if in a future modified appraisal system GPs are to self-assess their learning. Proposals have been put forward for GPs to award themselves 'credits' based

on the challenge and outcome of the learning activity (RCGP, 2008f). However, in a training context, doctors, are known to be inconsistent at self-assessment when compared with an external standard (Colthart *et al*., 2008b). For example, family physicians were asked to 'score' their own knowledge level in specific clinical areas prior to a formal assessment of that knowledge. Correlation between self assessment and actual performance was poor (Parker *et al*., 2004). This was particularly marked for those in the lowest quartile of the examination who greatly overestimated their ability. Unlike doctors who under-estimated their performance and could subsequently be trained to achieve a better correlation between their predicted and actual score, those who over-estimated their likely achievements had a tendency to maintain the dichotomy between self-predicted and actual outcome. This is essentially a lack of insight and has implications as to whether the doctor will be able to fully engage with their professional duties (Bosk, 1988). One method proposed to address difficulties in self assessment is through engaging in feedback from peers (Dunning, 2006).

Participation in an external review of SEA is viewed by participants (Bowie *et al.,* 2005a) and peer reviewers (Chapter 11) as demonstrating a willingness to undertake an educational activity that is beyond the 'safe' confines of normal practice. As such, it may be considered more challenging, and participation could be recognised by appraisers and RCGP, if a new system of 'credits' as described above is introduced. However, this is likely to require further evidence to make a truly persuasive case.

All voluntary peer feedback models will be limited by selectivity of submissions. A GP could have several serious significant events in a year and it may not be feasible to submit all of these for peer review or indeed discuss all of these with an appraiser. It could be strongly argued that identification of serious events is not the principal

objective of such systems. Rather, it is the learning associated with the reflection and review that is important. However, there needs to be an acknowledgement of the professional and ethical issues involved if GPs avoid identifying, analysing and submitting particular significant events because they may reflect poorly on their professional practice. Many may avoid highlighting 'technical' or 'knowledge' errors even though these are accepted by fellow doctors as part of professional practice (Bosk, 1988). If these incidents are avoided, then arguably the only other opportunity for these to be addressed is if they happen to be highlighted in complaints or litigation processes.

In terms of the nGMS contract, it is the responsibility of the primary care authority to ensure that SEA reports are of a sufficient quality to ensure payment to practices. There does not appear to be a standard training procedure for QOF reviewers in order that they can consistently inform participating practices as well as the appropriate healthcare authority that SEA reports are of a sufficient standard. Opportunities to address educational and patient care issues may therefore be lost. There may be merit in aligning the model with the QOF to ensure a consistent and fair approach to practices that have undertaken SEA. This may act to lessen a potential conflict between contractual and governance issues (that are likely to centre on the quality and safety of patient care) and the education and training aspect of SEA originally envisaged when it was included in the general practice contract.

14.4 Suggestions for further research

The research presented in this thesis adds a small amount of detail to the still very limited evidence base for SEA and its application in general medical practice. The

findings highlight areas that could be explored further in developing the evidence for SEA, particularly its role as part of the GMS contract and in GP appraisal.

SEA was introduced to these roles with little evidence for its effectiveness as a reflective learning technique or a quality improvement measure. There is now some evidence for its reported benefits as a method for achieving learning and change but this remains limited because of the self-reported and unverifiable nature of the findings. Research could be targeted to examine whether this leads to sustained improvement in the aspects of care analysed. In addition, whether analysis of 'marker' significant events is an effective use of time and other resources allocated to SEA in the nGMS contract requires evaluation.

In Scotland and most probably throughout the UK, participation in SEA is likely to be a key component of GP appraisal. The evidence of learning presented at appraisal will, in turn, inform a major part of their portfolio for recertification. The value of SEA as part of this process remains largely unknown both for appraisers and appraisees. Whether SEA offers more to healthcare practice than other learning activities could help update any future reviews of the appraisal process.

If SEA is to remain in appraisal for the foreseeable future, then the value of the external peer review system needs further evaluation. The views of appraisers need to be sought as to whether this kind of approach to peer review can enhance the educational value of the appraisal process. If it can, then their opinions on how this would be best achieved should be explored.

In addition, the opinions and attitudes of GPs who have not participated in the model require exploration. This would aim to identify barriers and obstacles to such systems. It could also inform whether it is likely to be anything more than a minority activity while such systems remain voluntary.

14.5 Conclusions

This thesis has presented findings that contribute to the evidence base on the analysis and reporting of significant events in general medical practice. In addition, aspects of the utility of one model for external peer review to support and potentially enhance SEA have been investigated. The studies have all been undertaken with GP principals in the west of Scotland deanery. The peer review model was examined within the context of the nGMS contract and the Scottish GP appraisal model. However, the issues identified and the interpretation of results could inform GPs, professional bodies and healthcare organisations throughout the UK on some of the strengths and limitations of significant events and their analyses.

REFERENCES

Aday, L.A. & Correlius, L.J. (2005) *Designing and conducting health surveys* Ch. 7, pp.124-153, California, Jossey Bass

Baker, G.R., Norton, P.G., Flintoft, V. Blais, R., Brown, A., Cox, J. *et al*. (2004) The Canadian Adverse Events study: the incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal*, 170(11), pp. 1678-1686

Baker, R. (1985) Comparison of standards in training and non training practices, *Journal of the Royal college of General Practitioners,* 35, pp.330-332.

Baker, R. & Thompson, J. (1995) Innovation in general practice: is the gap between training and non-training practices getting wider? *British Journal of General Practice*, 45, pp. 297-300.

Baker, R., Lakhani, M., Fraser, R., & Cheater, F. (1999) 'A model for clinical governance in primary care groups', *British Medical Journal*, 318, pp.779-783.

Balint, M. (1957) *'The doctors, his patient and the illness',* London, Tavistock.

Bank, A.L., MacNeill, S.E., Hall, E.M., Nadjarian, R.K., Zaccagnini, A.V. & Lichtenberg, P.A. (2002) More than meets the eye: How examiner training affects the reliability of the MacNeill-Lichtenberg decision tree in geriatric rehabilitation*, Archives of Physical Medicine and Rehabilitation*, 83, pp.405-411.

Barbour, S. (2001) Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *British Medical Journal*, 322 (7294), pp.1115-1117.

Barbour, R.S. (2005) Making sense of focus groups, *Medical Education*, 39(7), pp.742-750.

Barach, P. & Small, S.D. (2000) Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems, *British Medical Journal*, 320, pp.759-763.

Battles, J.B., Kaplan, H.S., Van der Schaaf, T.W. & Shea, C. (1998) The attributes of medical event-reporting systems - Experience with a prototype medical event-reporting system for Transfusion Medicine, *Archives of Pathological Laboratory Medicine*, 122, pp.231-238.

Beckmann, U., West, L.F. Groombridge, G.J. Baldwin, I. Hart, G.K. Clayton, D.G *et al.* (1996) The Australian Incident Monitoring Study in Intensive Care: AIMS-ICU. The development and evaluation of an incident reporting system in intensive care. *Anaesthesia and Intensive Care,* 24, pp.314-319.

Beebe, T., Stoner, S., Anderson, K. & Williams, A. (2007) Selected questionnaire size and color combinations were significantly related to mailed survey response rates, *Journal of Clinical Epidemiology*, 60,(11), pp. 1184-1189.

Bennett, I. & Danczak, A. (1994) Terminal Care: improving teamwork in primary care using significant Event Analysis, *European Journal of Cancer Care*, 1, pp. 119-125.

Berwick D. (2008) The science of improvement, *Journal of the American Medical Association,* 299*,* pp*.* 182-184.

Bhasale, A. (1998) The wrong diagnosis: identifying causes of potentially adverse events in general practice using incident monitoring, *Family Practice*, *15*, pp. 308-18.

Billings, C. (1998) *Incident reporting systems in medicine and experience with the aviation reporting system,* in *A tale of two stories: contrasting views of patient safety*, Cook, R.I., Woods, D.D. & Miller, C.A. pp 55-56. Massachusetts, North Adams.

Bing-You, R.G., Paterson & J Levine, M.A. (1997) Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility, *Medical Teacher*, 19 (1), pp. 40-44.

Bosk, C.L. (1988) *Forgive and remember: managing medical failure*, in *Professional judgment* eds. Dowie. J. & Elstein, A., Ch.29, pp. 523-543. Cambridge, Cambridge University Press.

Bowie, P., Garvie, A., Oliver, J. & Lough, M. (2002) Impact on non-principals in general practice of the summative assessment audit project, *Education for Primary Care*, 13, pp. 356-361.

Bowie, P., McKay, J., & Lough M. (2003) Peer assessment of significant event analyses: Being a trainer confers an advantage, *Education for Primary Care*, 14(3), pp. 338-344.

Bowie, P., McKay, J., Norrie, J. & Lough, M. (2004) Awareness and analysis of a significant event by general practitioners: A cross sectional survey, *Quality and Safety in Health Care*, 13(2), pp. 102-107.

Bowie, P., McKay, J., Dalgetty, E. & Lough, M. (2005a) A qualitative study of why general practitioners may participate in significant event analysis and educational peer review, *Quality & Safety in Health Care*, 14, pp 185-189.

Bowie, P., McCoy, S., McKay, J. & Lough, M. (2005b) Learning issues raised by the educational peer review of significant event analyses in general practice. *Quality in Primary Care*, 13, pp. 75–84.

Bowie, P. (2005c) *A study of the analysis of significant events by general medical practitioners and the role of educational peer review*. MD Thesis, University of Glasgow

Bowie P & Kelly D. (2007) *Educational Peer Review: A Strategic Document (2007-2010)*. NHS Education for Scotland: Edinburgh.

Bowie, P., Pope, L. & Lough, M. (2008) A review of the current evidence base for significant event analysis, *Journal of Evaluation in Clinical Practice*, 14 (4), pp. 520-536**.**

Boud, D., Keogh, R. & Walker, D. (1985) *Reflection Turning Experience into Leaning*, pp 7-18 Kegan Page, London.

Boynton, P.M. (2004) Hands-on guide to questionnaire research: Administering, analysing, and reporting your questionnaire, *British Medical Journal*, 328, pp.1372-5.

Boynton, P.M. & Greenhalgh, T. (2004a) Selecting, designing, and developing your questionnaire, *British Medical Journal*, 328, pp. 1312-1315.

Boynton P.M., Wood G.W. & Greenhalgh T. (2004b) Hands-on guide to questionnaire research: Reaching beyond the white middle classes, *British Medical Journal.* 328(7453) pp.1433-1436.

Bradley, C. P. (1992), Turning anecdotes into data - the critical incident technique. *Family Practice* 9(1), pp.98-103.

Brigley, S., Young, Y., Littlejohns, P. & McEwan, J. (1997) Continuing education for medical professionals: A reflective model, *Postgraduate Medical Journal*, 73(855), pp. 23-26.

Britt, H., Miller, G.C., Steven, I.D., *et al*. (1997) Collecting data on potentially harmful events: a method for monitoring incidents in general practice, *Family Practice*, 14(1) pp.01-06.

Britten, N. (1995) Qualitative research: Qualitative interviews in medical research, *British Medical Journal*, 311 (6999), pp.251-253.

Buckley, G. (1990) *Clinically Significant Events*, in *Medical Audit and General Practice* (Ed. Marinker, M), London, BMJ Publishing Group

Buetow, S. Bonnie. S, Cantrill, J. & Crombie, I.K. (1996) Prevalence of potentially inappropriate long term prescribing in general practice in the United Kingdom, *British Medical Journal,* 331, pp. 1371-1374.

Burroughs, L. & Atkins, S. (2005) The Doncaster West approach to learning from significant events, *Clinical Governance Bulletin*, 6 (2), pp. 6–8.

Callaham, M. (2003) *The evaluation and training of peer reviewers*, in *Peer Review in Health Sciences* 2nd Edition, eds. Godlee, F. & Jefferson, T. pp .164-182, London, BMJ Publishing Group.

Campanelli, P. (2008) *The Survey Introduction*, in *Encyclopaedia of Survey Research Methods,* ed. Lavrakas, P. London, Sage publications.

Campbell, J.L., Richards, S.H., Dickens, A., Greco, M., Narayanan, A. & Brearley S. (2008) Assessing the professional performance of UK doctors: an evaluation of the utility of the General Medical Council patient and colleague questionnaires, *Quality and Safety in Health Care*, 17(3), pp.187-93.

Campbell, S., Reeves, D., Kontopantelis, E., Middleton, E., Sibbald, B. & Roland, M. (2007) Quality of primary care in England with the introduction of pay for performance, *New England Journal of Medicine*, 357, pp.181-190.

Campbell, S.M., McDonald, R. & Lester, H. (2008) The experience of pay for performance in English family practice: A qualitative study, *Annals of Family Medicine*, 6, pp. 228-234

Candy, P.C. (1991) *Self-direction in Lifelong Learning.* San Francisco (CA), Jossey Bass

Carayon, P., Schoofs Hundt, A., Karsh, B.T., Gurses, A.P., Alvarado, C.J., Smith M. *et al*. (2006) Work system design for patient safety: the SEIPS model, *Quality and Safety in Health Care*, 15 (Supplement 1), pp.50-58.

Carifio, J. & Perla, R. (2008) Resolving the 50 year debate around using and misusing Likert scales, *Medical Education*, 42, pp.1150-1152**.**

Carmines, E. & Zeller, R.A. (1979) *Reliability and Validity Assessment,* London, Sage Publications.

Caughey, J.L. (1967) Nonintellectual components of medical education. *Journal of Medical Education*, 42, pp.619-625.

Cohen, M. R. (2000) "Why error reporting systems should remain voluntary – they provide better information for reducing errors", *British Medical Journal*, 319, pp.136-137.

Collins English Dictionary (2000). Harper Collins, Glasgow.

Colthart, I., Cameron, N., McKinstry, B. & Blaney D. (2008a) What do doctors really think about the relevance and impact of GP appraisal **three** years on?: A survey of Scottish GPs, *British Journal of General Practice*. 58(547), pp. 82-87. section 5.2.1

Colthart, I., Bagnall, G., Evans, A. Haig, A., Illing, J., & McKinstry B. (2008b) The effectiveness of self-assessment on the identification of learner needs, learner activity and impact on clinical practice, Best Evidence Medical Education Systematic Review guide no.10 *Medical Teacher*, 30, 124-145.

Committee of Public Accounts: (2007) *Improving quality and safety - progress in implementing clinical governance in primary care: lessons for the new Primary Care Trusts*. Publication of 47th Report, 12 Sept 2007. available at: http://www.parliament.uk/parliamentary_committees/committee_of_public_accounts/pacpn070913.cfm accessed 21 jan 2008.

Confidential enquiry into maternal and child health available at: www.cemach.org.uk accessed 20 Nov 2008.

Cook, R.I., Woods, D.D. & Miller C. (1998) A tale of two stories: contrasting views of patient safety. *US National Patient Safety Foundation*, pp. 1-5 North Adams, MA. www.npsf.org/rc/tts/front.html. accessed 6 July 2008

Cook, A., Netuveli, G. & Sheikh, A. (2004) *Basic Skills in Statistics*, London, Class Heath

Cox, S.J. & Holden, J.D. (*2007)* A retrospective review of significant events reported in one district in 2004-2005. *British Journal of General Practice,* 57, pp.732-736.

Cresswell, J. (1998a) *Qualitative inquiry and research design: Choosing among five traditions*. Ch.1 Introduction, pp. 1-26 Thousand Oaks, CA: Sage

Cresswell, J. (1998b) Data collection, In, *Qualitative inquiry and research design: Choosing among five traditions*. Ch.7 Data collection, pp 109-137, California, Sage Publications.

Cronbach, L.T. (1960) *Essentials of Psychological Testing*, 2nd edition, New York, Harper and Rowe.

Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972) *The dependability of behavioural measurements: Theory of generalizability of scores and profiles*, New York, John Wiley.

Crossley, J., Humphris, G. & Jolly, B. (2002a) Assessing health professionals, *Medical Education*, 36, pp.800-804.

Crossley, J., Davies, H., Humphris, G. & Jolly, B. (2002b) Generalisability: a key to unlock professional assessment, *Medical Education*, 36, pp. 972-978

Crossley, J., Russell, J., Jolly, B., Ricketts, C., Roberts, C. *et al*. (2007) 'I'm picking up good regressions': the governance of generalisability analyses, *Medical Education*, 41, pp. 926-934.

Davidoff, F. (2004) Improving peer review; who's responsible? *Medical Education*, 328, pp. 657-658.

Davis, D. A., Thomson, M.A., Oxman, A. D. & Haynes R. B. (1995) Changing Physician Performance: A systematic review of the effect of continuing medical education strategies, *Journal of the American Medical Association*, 274(9), pp.700-705.

Demsky, Y. & Sellers, A. (1995) Improving examiner reliability on the visual memory span subset of the weschler memory scale, *Clinical Neuropsychologist*, 9(1), 79-82.

Denzin, N. & Lincoln, Y. (2005) *The discipline and practice of qualitative research*, in eds. Denzin, N. & Lincoln, Y., *The Sage Handbook of Qualitative Research*, Ch 1 Introduction, pp.1-32, California, Sage publications.

Department of Health. (1998), *A First Class Service*, London, Department of Health.

Department of Health. (1999) *Supporting doctors, protecting patients: a consultation paper,* London: Department of Health.

Department of Health. (2000) *An organisation with a memory: report of an expert group on learning from adverse events in the NHS*, London, Department of Health.

Department of Health. (2001a*) Building a safer NHS for patients: implementing an organisation with a memory*, London, Department of Health.

Department of Health. (2001b) *Doing less harm: improving the safety and quality of care through reporting, analysing and learning from adverse incidents involving NHS patients – key requirements for health care providers*, London, Department of Health.

Department of Health. (2004) *New GMS contract 2004/5.* London, Stationery Office.

Department of Health. (2006a) *Good doctors, safer patients: proposals to strengthen the system to assure and improve the performance of doctors and to protect the safety of patients. A report by the Chief Medical Officer.* London: Department of Health.

Department of Health. (2006b) *New GMS contract 2006/7*, London, Stationery Office.

Department of Health. (2007) *Trust, Assurance and Safety—The Regulation of Health Professionals in the 21st Centur,y* London, Stationery Office.

Department of Health. (2008) *Medical Revalidation – Principles and Next Steps*
The Report of the Chief Medical Officer for England's Working Group, London: Department of Health

De Wet, C. & Bowie, P. (In Press) A preliminary study to develop and test a global trigger tool to identify undetected error and patient harm in primary care records, *Postgraduate Medical Journal.*

Dewey, J. (1933) *How We Think: a restatement of the relation between reflective thinking to the educative process*, London, DC Heath.

DiCicco-Bloom B, & Crabtree B. (2006) The qualitative research interview. *Medical Education*, 40, pp. 314– 21.

Dixon-Woods, M., Fitzpatrick, R. & Roberts, K. (2001) Including qualitative research in systematic reviews: opportunities and problems, *Journal of Evaluation in Clinical Practice*, 7 (2), pp. 125-133.

Dolman, E. Hazelwood, L., Naidoo, S. & Purkiss, V. (2005) Assessment of clinical audit: lessons learned from Trent. *Education for Primary Care* 16, pp. 506-7.

Donabedian, A. (1980) The measurement of quality: a conceptual explanation, in *Explorations in Quality Assessment and Monitoring*. Volume 1: The Definition of Quality and Approaches to Assessment, Michigan, Ann Arbor, Health Administration Press.

Donaldson, L. (2006) *Good Doctors, Safer Patients London*: Department of Health.

Dovey, S.M., Meyers, D.S., Phillips, R, L., Green, L.A., Fryer, G.E., Galliher, J.M *et al*., (2002) A preliminary taxonomy of medical errors in family practice, *Quality and Safety in Health Care*, 11, pp.233-238.

Dowling, S., Rouse, M., Farrell, J., Gannon, L., Sullivan G., & Cussen K. (2007) The acceptability, feasibility and educational impact of a new tool for formative assessment of the consultation performance of specialty registrars in an Irish general practice training scheme. *Education for Primary Care*. 18(6) pp. 724-735.

Downing, S.M. (2003) Validity: On the meaningful interpretation of assessment data. *Medical Education*. 37(9) pp. 830-837.

Downing, S.M. (2004) Reliability: on the reproducibility of assessment data. *Medical Education*, 38, pp. 1006-1012.

Drewnowski A. (2001) Diet image: a new perspective on the food-frequency questionnaire. *Nutrition Review,* 59, pp. 370-2.

Driessen, E., van Tartwijk, J. & Dornan, T. (2008) The self critical doctor: helping students become more reflective. *British Medical Journal,* 336, pp.827-830.

Dunning, D. (2006) Strangers to ourselves? *Psychologist*, 19 (10), pp. 600-603.

Elder, N., Vonder Meulin, M.V., & Cassidy A. (2004) The identification of medical errors by family physicians during outpatient visits. *Annals of Family Medicine*, 2, pp. 125-129.

Elwyn, G., Edwards, & A. Grol, R. (1999) Towards a feasible model for shared decision making: a focus group study with general practice registrars. *British Medical Journal*, 319 (7212), pp.753-756.

Ely, J.W., Levinson, W., Elder, N.C., Mainous III, A.G., & Vinson, D.C. (1995) Perceived causes of family physicians' errors, *Journal of Family Practice*, 40, pp.337-344.

Ende, J. (1983) Feedback in clinical medical education, *Journal of the American Medical Association*, 250, pp.777-781.

Esmail, A., Neale, G., Elstein, M., Firth-Cozens, J., Davy, C., & Vincent, C. (2004) Case studies in litigation: claims reviews in four specialities pp5-27. *Report published by the Manchester Centre for Healthcare Management*. Manchester: University of Manchester.

Esmail. A. (2006) *Patient safety in primary care: on a wing and a prayer?* available at: http.//www.Saferhealthcare.org.uk/IHI/Topics?safetyCulture/Literature/patient safety.htm accessed 12 Dec 2006.

Evans, R., Elwyn, G., & Edwards, A. (2004), Review of instruments for peer assessment of physicians, *British Medical Journal*, 328, pp.1240-1243

Evans, S.M., Berry, J.G., Smith, B.J., Esterman, A., Selim, P., O'Shaughnessy J. *et al*. (2006) Attitudes and barriers to incident reporting: a collaborative hospital study. *Quality and Safety in Health Care,* 15, pp. 39-43

Fallowfield, l. (1995) Questionnaire design, *Archives of Disease in Childhood*, 72 (1), pp.76-79.

Fink A., Kosecoff J., Chassin M. & Brook, R.H. (1984) Consensus methods, characteristics and guidelines for use. *American Journal of Public Health,* 74, pp.729-734.

Firth-Cozens, J., Firth, R.A. & Booth, S. (2003) Attitudes to and experiences of reporting poor care. *Clinical Governance: An international Journal*, 8, pp.331-336.

Fischer, G., Fetters, M.D., Munro, A.P. & Goldman, E.B. (1997) Adverse events in primary care identified from a risk management database. *Journal of Family Practice,* 45, pp. 40-7.

Flanagan, J.C. (1957), The Critical Incident Technique, *Psychological Bulletin*, 1957, 51, pp.327-358.

Fox, M., Sweeney. G., Howells, C., & Stead, J. (2001), Significant event audit in prison healthcare: changing a culture for clinical governance – a qualitative study, *Journal of Clinical Governance*, 9, pp.123-128.

Garavan, T.N. (1997) Training, development, education and learning: different or the same? *Journal of European Industrial Training*, 21(2), 39-50.

General Medical Council (2001) *Good Medical Practice*, London, GMC

Goldberg, L.R. (1971) A historical survey of personality scales and inventories*,* in *Advances in psychological assessment*, vol 2. ed. McReynolds, P. pp293-336. CA., Palo Alto.

Goppe, N. (2001) The role of peer assessment in nursing, *British Journal of Nursing*, 10(2), pp. 115-121.

Grant, J. (2002) Learning needs assessment: assessing the need, *British Medical Journal*, 324, pp.156-159.

Greco, P.J. & Eisenberg, J.M. (1993) Changing physicians' practices. *New England Journal of Medicine*; 329, pp.1271–1273.

Grol, R., Mokkink, M. & Schellevis, F. (1988) The effects of peer review in general practice, *Journal of the Royal College of General Practitioners*, 38, pp. 10-13.

Grol, R., (1990) Peer review in primary care, *Quality Assurance in Healthcare*, 2, pp. 119-126

Grol, R. & Lawrence, M. (1995) A model for peer review and audit, in *Quality Improvement by Peer Review,* eds. Grol, R. & Lawrence, M. Ch.1, pp.3-13 Oxford, Oxford University Press.

Grol, R., Wensing, M. & Eccles, M. (2005) *Improving patient care*, London, Elsevier.

Hancock, B., (2002) An introduction to qualitative research. Trent focus for research and development in Primary health Care, pp.17-18, available at: http://faculty.capebretonu.ca/pmacintyre/course_pages/MBA603/MBA603_files/IntroQualitativeResearch.pdf accessed Jan 14 2009.

Harrison, P., Joesbury, H., Martin, D., Wilson, R. & Fewtrell, C. (2002) *Significant event audit and reporting in general practice*. Commissioned Report by the School of Health and Related Research, University of Sheffield, available at: http://www.shef.ac.uk/uni/academic/R-Z/scharr/html accessed 22 Jan 2007

Hart, E. & Hazelgrove, J. (2001) Understanding the organisational context for adverse events in health services: the role of cultural censorship, *Quality in Health Care*, 10, pp. 257-262.

Hay, J.A. (1995) *Tutorial reports and ratings*. In: Shannon, S. and Nocterm, G. (eds). *Evaluation Methods: A Resource Handbook*, Hamilton, Ontario, McMaster University.

Helfer, R. (1972) Peer evaluation: its potential usefulness in medical education. *British Journal of Medical Education*, 6, pp. 224-231.

Henderson, E., Berlin, A., Freeman, G. *et al*. (2002) Twelve tips for promoting significant event analysis to enhance reflection in undergraduate medical students. *Medical Teacher*, 24 (2), pp.121–124.

Hewson, M.G. & Little, M.L. (1998) Giving feedback in medical education: Verification of recommended techniques. *Journal of General Internal Medicine,* 13(2), pp.111-6.

Hider, P., Lay-yee, R., Davis, P., Briant, R. & Scott, A. (2005) Monitoring the quality of primary care: Use of hospital-based audit studies. *International Journal of Risk & Safety in Medicine*, 17, pp. 81–89.

Hillier, L. (2002) Introducing significant event analysis to independent contractors, *Journal of Clinical Excellence*, 4 (1 Suppl.), pp. 168–171.

Holden, J.D. (2001) Hawthorne effects and research into professional practice, *Journal of Evaluation in Clinical Practice*, 7(1), pp.65-70.

Holden, J. & Kay, A. (2005) Fellowship by assessment: 15 years of peer assessment of high-quality general practice, *Quality in Primary Care*, 13 (1), pp.3-8.

House of Commons Committee of Public Accounts (2006) *A safer place for patients: learning to improve patient safety* (HC 456, Session 2005–06) House of Commons, London: Stationery Office.

Hudelson, P., Cleopas, A., Kolly, V., Chopard, P. & Perneger, T. (2008) What is quality and how is it achieved? Practitioners' views versus quality models, *Quality and Safety in Health Care*, 17, pp.31-36

Hudson, P. (2003) Applying the lessons of high risk industries to healthcare, *Quality and safety in Healthcare*, 12 (S1), pp.7-12.

Hunter C., Blaney D., McKinstry B., Shaw J., & Denholm C. (2005) GP appraisal: A Scottish solution, *Education for Primary Care*, 16(4), pp. 434-439.

Hutchinson, L., Aitken, P. & Hayes, T. (2002) Are medical postgraduate certification processes valid? A systematic review of the published evidence, *Medical Education*, 36 (1), pp.73-91.

Illing, J. (2007) *Thinking about research: frameworks, ethics and scholarship,* in *Understanding Medical Education*, ed. Swanick, T., Edinburgh, Association for the Study of Medical Education.

Institute for Healthcare Improvement. 2009 Care bundles http://www.ihi.org/IHI/Topics/CriticalCare/IntensiveCare/ImprovementStories/BundleUpforSafety.htm. accessed jan 20, 2009

Irvine, D.H. (1983) Quality of care in general practice: our outstanding problem, *Journal of the Royal College of General Practitioners*, 33, pp.521-523

Irvine, D. & Irvine, S. (Eds), (1991) *Making sense of audit*, Radcliffe Medical Press, Oxford.

Jacobson, L., Elwyn, G., Robling, G. & Tudor Jones, R. (2003) Error and safety in primary care: no clear boundaries, *Family Practice*, 20, pp.237-241.

Jamtvedt, G., Young, J.M., Kristoffersen, D.T., O'Brien, M.A. & Oxman, A.D. (2007) *Audit and feedback: effects on professional practice and health care outcomes* (Review) The Cochrane Library, issue 4 available at: http://www.who.int/rhl/reviews/CD000259.pdf accessed 13 May 2008

Johnson, C.W. & Holloway, C.M. (2007) A Historical Perspective on Aviation Accident Investigation. *@inproceedings IET 2nd Conference on System Safety*.

Johnston, G., Crombie, I.K., Alder, E.M., Davies, H.T.O., & Millard, A. (2000) Reviewing audit: barriers and facilitating factors for effective clinical audit, *Quality in Healthcare*, 9, pp.23-36.

Joint Committee on Postgraduate Training for General Practice (1992*) Recommendations to regions for the establishment of criteria for the approval and re-approval of trainers in general practice,* London, JCPTGP.

Jones, J. & Hunter, D. (1995) Consensus methods for medical and health services research, *British Medical Journal*, 311, pp. 376-380.

Kasje, W.N., Denig, P., de Graeff, P.A. & Haaijer-Ruskanp, F.M. (2004) Physicians views on joint treatment guidelines for primary and secondary care, *International Journal for Quality in Health Care*, 16(3), pp. 229-236.

Kasje, W.N., Denig, P., Stewart, R.E., de Graeff, P.A. & Haaijer-Ruskamp, F.M. (2005) An educational programme for peer review groups to improve treatment of chronic heart failure and diabetes mellitus type 2 in general practice, *Journal of Evaluation in Clinical Practice*, 12 (6), pp.613-621.

Kaufman, D.M., Mann, K.M. & Jennet P.A. (2000) *Teaching and learning in medical education,* Association for the study of medical education, Edinburgh.

Kaufman, D.M. (2003) Applying educational theory in practice, *British Medical Journal*, 326, pp. 213-216.

Kirk, S., Parker, D., Claridge, T., Esmail, A. & Marshall, M. (2007) Patient safety culture in primary care: developing a theoretical framework for practical use, *Quality and Safety in Healthcare,* 16, pp.313-320.

Kirkpatrick, D. L. (1967) *Evaluation of training.* In *Training and Development Handbook* eds Craig, R. & Mittel, I. pp. 87–112. New York: McGraw Hill.

Kitzinger, J. & Barbour, R. (1999) *Developing focus group research, politics theory and practice*, Ch.1, London, Sage Publications.

Knowles, M. (*1980) The modern practice of adult education: from pedagogy to andragogy*. Cambridge Books. New York.

Kohn, L., Corrigan, J. & Donaldson, M.E. (1999) *To err is human*, National Academy Press, Washington, DC.

Kolb, D.A. (1984) *Experiential Learning: experience as the source of learning and development*, Englewood Cliffs, New Jersey, Prentice Hall.

Kostopoulou, O. (2006) A taxonomy of patient safety in primary care, *Ergonomics,* 49, pp.486-502.

Krippendorff, K. (2004) *Content analysis: an introduction to its methodology*, CA, Sage publications.

Kroll, L., Singleton, A., Collier, J. & Rees Jones, I. (2008) Learning not to take it seriously: junior doctors' accounts of error, *Medical Education*, 42, pp. 982-990.

Krueger, R.A. (1994) *Focus groups: a Practical Guide for Applied Research*, 2nd edn. Sage, London, UK.

Kubany, A.J. (1957) Use of sociometric peer nominations in medical education research, *Journal of Applied Psychology*, 41, pp.389-394.

Kuper, A., Reeves, S. & Levinson, W. (2008a) An introduction to reading and appraising qualitative research, *British Medical Journal*, 337 (973), 404-407.

Kuper, A., Lingard, L. & Levinson, W. (2008b) Critically appraising qualitative research, *British Medical Journal, 337,* pp. 687-689.

Leape, L.L. & Berwick, D.M. (2000) Safe health care: are we up to it? *British Medical Journal,* 320, pp.725-726.

Leape, L.L. (2000) Institute of Medicine figures are not exaggerated, *Journal of the American Medical Association*, 284, pp. 85-97.

Lingard, l. & Kennedy, T.J. (2007) *Qualitative research in medical education*, In *Understanding Medical Education*, ed. Swanick, T., Edinburgh, Association for the Study of Medical Education.

Liske, R., Ort, R.S. & Ford, A.G. (1964) Clinical performance and related traits. *Journal of Medical Education*, 39, pp. 69-80.

List, T.J.M., Dworkin, S.F. & Svensson, P. (2006) Recalibration improves inter-examiner reliability of TMD examination, *Acta Odontologica Scandanavia*, 64(3), pp. 146-152.

Lord, H. & Lillis, S. (2005) Implementing significant event management in general practice – potential barriers and solutions, *New Zealand Family Practice*, 32(4), 247-250.

Lough, J.R., McKay, J. & Murray, T.S. (1995) Audit and summative assessment: two years' pilot experience, *Medical Education*, 29, pp.101-103.

Lough, J.R.M., Kelly, D.R., Taylor, M., Snadden, D., Patterson, B., McNamara, I., *et al*. (2002) Clinical governance in Scotland – an educational model*, British Journal of General Practice*, 52, pp.322-328.

Lough, J.R.M. (2002) *The development of integrated audit for the training of general practitioners*. MD Thesis, University of Glasgow.

Lowe, D. (1993) *Planning for medical research: a practical guide to research methods*, Great Britain, Astraglobe Ltd.

Lynn, M.R. (1986) Determination and quantification of content validity. *Nursing Research*, 35(6), pp. 383-385.

Madridejos-Mora, R., Amado-Guirado, E., & Pérez-Rodríguez, M.T. (2004) Effectiveness of the combination of feedback and educational recommendations for improving drug prescription in general practice, *Medical Care*, 42 (7), pp. 643-648.

Makeham, M.A.B., Dovey, S.M., County, M, *et al*., (2002) An international taxonomy for errors in general practice: a pilot study, *Medical Journal of Australia* 177, pp.68-72.

Makeham, M.A.B., Stromer, S., Bridges-Webb, C., Mira, M., Saltman, D.C., Cooper, C. *et al*., (2008) Patient safety events reported in general practice: a taxonomy *Quality and Safety in Health Care*, 17, pp. 53 - 57.

Mamede, S & Schmidt, H.G. (2004) The structure of reflective practice in medicine, *Medical Education*, 38(12), pp.1302-1308.

Marton, F. & Saljo, R. (1976) On qualitative differences in learning. 1 – Outcome and process, *British Journal of Educational Psychology*, 46, pp. 4-11.

Mayor, S., (2001), "NHS introduces new patient safety agency", *British Medical Journal*, 322 p.1013.

Mays, N. & Pope, C. (2000) *Quality in qualitative health research*, in *Qualitative research in health care* eds. Pope, C. & Mays, N. 2nd edition, Ch.9 London, BMJ Publishing Group.

McColl, E., & Thomas, R. (2000), '*The Use and Design of Questionnaires',* in (eds.) Carter, Y., Shaw, S., & Thomas, S. *Master Classes in Primary Care Research No.2,* London, Royal College of General Practitioners

McIntyre, N. & Popper, H. (1983) The critical attitude in medicine: The need for a new ethic, *British Medical Journal*, 287, pp. 1919-1923.

McKay, J., Bowie, P. & Lough, JRM. (2003) Evaluating significant event analyses: implementing change is a measure of success. *Education for Primary Care*, 14, pp. 34-38.

McKay, J., Bowie, P., & Lough, J.R.M. (2006) Variation in the ability of general medical practitioners to apply two methods of clinical audit: a 5-year study of assessment by peer review, *Journal of Evaluation in Clinical Practice*, 12, pp. 622-629.

McKinley, R.K., Fraser, R.C., van der Vleuten, C.P.M. & Hastings, A.M. (2000) Formative assessment of the consultation performance of medical students in the

setting of general practice using a modified version of the Leicester Assessment Package, *Medical Education*, 34(7), pp. 573-579.

McKinley, R.K., Fraser, R.C. & Baker, R. (2001) Model for directly assessing and improving clinical competence and performance in revalidation of clinicians, *British Medical Journal*, 322, pp. 712-715.

McLeod, P.J., Meagher, T.W., Steinert, Y. & Boudreau, D. (2000) Using focus groups to design a valid questionnaire, *Academic Medicine*, 2000, p.671

McManus, I.C., Thompson, M. & Mollon, J. (2006) Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modeling, *BMC Medical Education*, 6,42 doi:10.1186/1472-6920-6-42

McMillan, R. & Cameron, N. (2006) Factors influencing the submission of videotaped consultations by general practitioners for peer review and educational feedback. *Quality in Primary Care,* 14 (2), 85-89.

Merriam, S.B. & Caffarella, R.S. (1999). *Learning in adulthood*. 2nd ed. San Francisco, Jossey-Bass.

Messick, S. (1980) Test Validity and the Ethics of Assessment, *American Psychologist,* 35(11), pp.1012-1027.

Messick, S. (1995) Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning, *American Psychologist*, 50(9), pp. 741-749.

Miles, A. Holden, J. (2007) Fellowship by assessment 1989-2006, *The British Journal of General Practice*, 57 (534), pp. 74-75.

Morgan, D.L. (1987) *Focus groups as qualitative research*, 2nd Ed. Ch.5, pp.45-64, London, Sage Publications.

Morgan D.L. (1995) Why things sometimes go wrong in focus group interviews, *Qualitative Health Research*, 5, pp. 516-523.

Morrison, J. (1999) 'Clinical governance – implications for medical education', *Medical Education*, 33, pp. 162-164.

Moss, P.A. (1992) Shifting conceptions of validity in educational measurement: Implications for performance assessment, *Review of Educational Research*, 62(3), pp. 229-258.

Murie J. & McGhee C. (2003) Assessing risk by analysing significant events in primary care, *Quality in Primary Care*. 11(3) pp. 205-210.

National Confidential Enquiry into Patient Outcome and Death. Available at www.ncepod.org.uk accessed 20 November 2008

National Confidential Enquiry into Suicide and Homicide, Available at www.medicine.manchester.ac.uk/psychiatry/research/suicide/prevention/nci/ accessed 20 Nov 2008.

National Patient Safety Agency (2005) *Seven steps to patient safety for primary care:* London, NPSA.

National Patient Safety Agency (2006) *Quarterly National Reporting and Learning System Data Summary. Autumn 2006.* London, NPSA.

National Patient Safety Agency (2008) *Reporting and learning system (RLS), data summary,* available at www.npsa.nhs.uk/nrls/patient-safety-incident-data/quarterly-data-reports/ accessed 10 October 2008

Nelson, A.R. (1976) Orphan data and the unclosed loop: a dilemma in PSRO and medical audit, *New England Journal of Medicine*, 6 (295), pp.617-619.

Ness, G., & Cordess, C. (2002), "Honesty and openness may not be best policy", [Letter*] British Medical Journal*, 324, p.109.

Neville, R.G., Robertson, F., Livingstone, S. & Crombie, I.K. (1989) A classification of prescription errors, *Journal Royal College of Gen Practitioners*, 39, pp.110-112.

Newble, D.I., Hoare, J. & Sheldrake, P.F. (1980) The selection and training of examiners for clinical examinations, *Medical Education*, 4, pp. 345-349.

Newble, D.I. & Jaeger, K. (1983) The effects of assessments and examinations on the learning of medical students, *Medical Education*, 41 (6), pp. 165-171.

Newble, D.I. & Entwistle N.J. (1986) Learning styles and approaches: implications for medical education, *Medical Education*, 20 (3), pp. 162-75.

NHS Information Centre and Department of Health: *Trends in consultation rates in general practice 1995 to 2007: Analysis of the QResearch database. September 2008.* http://www.ic.nhs.uk/webfiles/publication/trends accessed 4 January 2009.

NHS Modernisation Agency (2004) *10 High Impact Changes for Service Improvement and Delivery.* NHS Modernisation Agency. London

NHS Quality Improvement Scotland (2007) *NHS Scotland incident reporting culture*, available at http://www.nhshealthquality.org/nhsqis/3869.html accessed 20 Oct 2008.

Nicol, D.J. & McFarlane-Dick, D. (2006) Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education.* 31, pp. 199-218.

Noel,, G.L., Herbers, J.E., Caplow, M.P., Cooper, G.S., Pangaro, L.N. & Harvey, J. (1992) How well do internal medicine faculty members evaluate the clinical skills of residents? *Annals of Internal Medicine,* 117 pp. 757–765.

Norcini, J.J. (2003) Peer assessment of competence, *Medical Education*, 37(6), 539-543.

Norman, G.R., Davis, D.A., Lamb, S., Hanna, E., Caulford, P. & Kaigas, T. (1993) Competency assessment of primary care physicians as part of a peer review programme, *Journal of the American Medical Association,* 270, pp. 1046–51.

Norman, G.R., Shannon, S. & Marrin, M.L. (2004) The need for needs assessment in continuing medical education, *British Medical Journal,* 328, pp. 999-1001.

Norman, G & Eva, K.W. (2008) *Quantitative research methods in medical education*,in *Understanding Medical Education,* ed. Swanick, T., Edinburgh, Association for the Study of Medical Education.

Nunally, J.C. & Bernstein, I.H. (1978) *Psychometric theory*, 2nd ed. New York, McGraw Hill

O'Brien, K. (1993) *Improving survey questionnaires through focus groups*, Morgan, David L (Ed). (1993). *Successful focus groups: Advancing the state of the art*. pp.105-111. California, Sage Publications,

Olsen S., Neal G., Schwab K., Psaila B., Patel T., Chapman EJ. *et al*. (2007) Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place, *Quality & Safety in Health Care*, 16, pp. 40-44.

Oppenheim, A.N. (1966) *Questionnaire design and attitude measurement*, Aldershot, Gower publishing.

Oxman, A.D., Thomson, A.D., Davis, D.A. & Haynes, R.B. (1995) No magic bullets: a systematic review of 102 trials of interventions to improve professional practice, *Canadian Medical Association Journal*, 153 (10), pp.1423-1431.

Parker, D. & Lawton, R. (2003) Psychological contribution to the understanding of adverse events in health care, *Quality and Safety in Health Care*, 12, pp. 453-457.3

Parker, R.W., Alford, C. & Passmore, C. (2004) Can family medicine residents predict their performance on the in-training examination? *Family Medicine*, 36, pp. 705-709.

Patterson, E. (1996) The analysis and application of peer assessment in nurse education, like beauty is in the eye of the beholder. *Nurse Education Today* 16, pp. 49-55.

Patton, Q.M. (2002) *'Qualitative Evaluation and Research Methods* (3rd Edition)', London, Sage Publications Ltd.

Peacock, J. & Kerry S. (2007) *Presenting medical statistics from proposal to publication*, Ch.10, pp109-114, Oxford, Oxford University Press.

Pearce, J.L. & Porter, L.W. (1986) Employee response to formal performance appraisal feedback, *Journal of Applied Psychology*, 71(2), pp. 211-218.

Pendleton, D., Scholfield, T., Tate, P. & Havelock, P. (1984) *The consultation; an approach to learning and teaching*, Oxford, Oxford Medical Publications.

Pitts J, (2007) Portfolios, personal development and reflective practice, *Association for the Study of Medical Education* p.12.

Pope, C., **Ziebland, S. & Mays, N.** (2000) *Quality in qualitative health research*, in *Qualitative research in health care,* eds. Pope, C. & Mays, N. 2nd edition, Ch.8, London, BMJ Publishing Group.

Powell, A.E., Rushmer, R.K. & Davies, H.T.O. (2009) *A systematic narrative review of quality improvement models in health care*, Universities of Dundee and St. Andrews, NHS QIS.

Pringle, M., Bradley, C.P., Carmichael, C.M., Wallis, H., & Moore, A. (1995) *Significant event auditing. A study of the feasibility and potential of case-based auditing in primary medical care*. Occasional Paper No.70, Royal College of General Practitioners, London: RCGP

Pringle, M. (1998) Preventing ischaemic heart disease in one general practice: from one patient, through clinical audit, needs assessment and commissioning into quality improvement, *British Medical Journal*, 317, pp.1120-1124.

Pringle, M. (2001), Ensuring patient safety, *British Journal of General Practice*, 51(10), pp. 876-877.

Pringle, M. (2007) Significant event auditing and reflective learning in primary care. *Safer healthcare* http://www.saferhealthcare.org.uk/IHI/Topics/HealthcareSettings/PrimaryCare.html accessed 24 July 2008.

Ramsay, P.G. & Wenrich, M.D. (1999) Peer ratings: an assessment tool whose time has come. *Journal of General Internal Medicine*, 14, pp. 581-582.

Raymond, M.R., Webb, L.C. & Houston, W.M. (1991) Correcting performance-rating errors in oral examinations, *Evaluation and the Health Professions*, 14, pp.100-122.

Reason, J.T. (2001) *Understanding adverse events; the human factor*, in Vincent, C. *Clinical risk management: enhancing patient safety* 2nd Ed. BMJ Publications, London.

Reeves, S., Lewin, S. & Zwarenstein, M. (2006) Using qualitative interviews within medical education research: why we must raise the 'quality bar,' *Medical Education*, 40(4), pp. 292-292.

Reeves, S., Albert, M., Kuper, A., & Hodges B.D. (2008) Qualitative research: Why use theories in qualitative research? *British Medical Journal*, 337(7670) pp. 631-634

Rennie D. (2003) *Editorial peer review: Its development and rationale*. in: Godlee F, Jefferson T, eds. *Peer review in health sciences*, pp.1-14, London: BMJ Books

Renshaw, M., Vaughan, C., Ottewill, M., Ireland A. & Carmody, J. (2008) Clinical Incident Reporting: wrong time, wrong place, *International Journal of Health Care Quality Assurance*, 21(4), pp. 380-384.

Resar R.K., Rozich J.D, & Classen D. (2003) Methodology and rationale for the measurement of harm with trigger tools, *Quality & Safety in Health Care* Dec; 12 (Suppl 2):39-45.

Rethans, J.J., Sturmans, F., Drop, R., van der Vleuten, C.P.M., & Hobus, P. (1991) 'Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice', *British Medical Journal,* 303, pp.1377-1380.

Ringsted, C., Henriksen, A.H., Skaarup, A.M. & van der Vleuten, C.P.M. (2004) Educational impact of in-training assessment (ITA) in postgraduate medical education: a qualitative study of an ITA programme in actual practice, *Medical Education*, 38, pp. 767-777.

Roberts, C., Cromarty, I., Crossley, J. *et al*. (2006) The reliability and validity of a matrix to assess completed reflective personal development plans of general practitioners, *Medical Education,* 40, pp. 363 -370.

Robinson L.A. & Stacy, R. (1994) Palliative care in the community: setting practice guidelines for primary care teams, *British Journal of General Practice* 44: pp. 461-464.

Robinson, L.A., Stacy, R., Spencer, J. A. & Bhopal, R.S. (1995) Use of facilitated case discussions for significant event auditing, *British Medical Journal*, 311, pp. 315-318.

Robinson, L. & Drinkwater, C. (2000) A significant case audit of a community-based elderly resource team – an opportunity for multidisciplinary teams to introduce clinical governance? *Journal of Clinical Governance*, 8 (2), pp. 89–96.

Robson, C. (2002) *Surveys and Questionnaires*, in Real World Research, Ch.8, Oxford, Blackwell.

Royal College of General Practitioners. (2000) RCGP standards and criteria for accreditation. In: Walshe K, Walsh N, Schofield T, *et al.,* eds. *Accreditation in primary care: towards clinical governance*. Oxford: Radcliffe Medical Press, pp. 127–47.

Royal College of General Practitioners (1983), *The Quality Initiative,* London, RCGP

Royal College of General Practitioners (1990) Fellowship by assessment, *Occasional Paper* 50, London, RCGP

Royal College of General Practitioners (1997) *Fellowship by Assessment.* Occasional Paper 50. London: RCGP, (revised).

Royal College of General Practitioners (2008a) *Fellowship Guidance Notes*, available at http://www.rcgp.org.uk/membership/fellowship.aspx accessed 20 Feb 2009.

Royal College of General Practitioners (2008b) *Interim membership by assessment of performance* (iMAP) available at http://www.rcgp.org.uk/gp_training/imap.aspx accessed 20 Feb 2009.

Royal College of General Practitioners (2008c) *Quality Practice Award* available at http://www.rcgp.org.uk/practising_as_a_gp/team_quality/qpa.aspx accessed 20 Feb 2009.

Royal College of General Practitioners (2008d) *Principals of GP Appraisal* available at www.rcgp.org.uk/pdf/corp_principles accessed 12 Dec 2008

Royal College of General Practitioners (2008e) *Revalidation for General Practitioners Consultation Document*, available at www.rcgp.org.uk/docs/Appendix%20%revalidation. accessed 12 Dec 2008,

Royal College of General Practitioners (2008f) *RCGP credit based system for continuing professional development*, available at www.rcgp.org.uk/docs/cpd_council accessed Dec 7, 2009.

Royal College of General Practitioners Scotland. *Practice accreditation*. 2002.

Rubin, G., George, A., Chin D *et al*. (2003) Errors in general practice: development of an error classification and pilot study of a method for detecting errors. *Quality and Safety in Health Care*,12, pp. 443-7.

Runciman, W.B., Webb, R.K., Barker, L. *et al.* (1993*)* The Australian Incident Monitoring Study. The pulse oximeter: applications and limitations – an analysis of 2000 incident reports. *Anaesthesia and Intensive Care,* 21, pp.543-550. out

Runciman, B., Merry, A. & McCall Smith, A. (2001) "Improving patients' safety by gathering information: anonymous reporting has an important role", *British Medical Journal*, 323, p.298.

Rushton, A. (2005) Formative assessment: a key to deep learning? *Medical Teacher*, 27, pp. 509-13.

Russel, I.T., Addington-Hall, J.M., Avery, P.J., Bamford, C.H., Barton, A.G., Donaldson C. et al. (1992) Medical audit in general practice. I: Effects on doctors' clinical behaviour for common childhood conditions. North of England Study of Standards and Performance in General Practice, *British Medical Journal*. 304(6840), pp. 1480-1484.

Sanazaro, P.J. (1983) Determining physicians' performance: continuing medical education and other interacting variables, *Evaluation and the Health Professions*, 6(2), pp.197-210.

Sanders, J. & Esmail, A. (2003) The frequency and nature of medical error in primary care: understanding the diversity across studies, *Family Practice*, 20(3), pp. 231-236.

Sargeant, J., Mann, K. & Ferrier, S. (2005) Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness, *Medical Education,* 39, pp. 497-504.

Scally, G., & Donaldson, L. (1998) Clinical governance and the drive for quality improvement in the new NHS in England, *British Medical Journal*, 317, pp. 61-65.

Schon D.A. (1983) *The reflective practitioner: how professionals think in action*. pp.57-60, New York, Basic Books.

Schroter, S. Black. N., Evans. S., Carpenter. J, Godlee, F. & Smith. R. (2004) Effects of training on quality of peer review: randomised control trial, *British Medical Journal*, 328, pp. 673-675.

Schroter, S. & Groves, T. (2004). *BMJ training for peer reviewers*, *British Medical Journal,* 328, pp. 658-658.

Schuwirth, L.W.T. & van der Vleuten, C.P.M. (2006) *How to design a useful test: the principles of assessment*, in *Understanding Medical Education,* ed. Swanick, T., Edinburgh, Association for the Study of Medical Education

Scottish Executive, NHS Education for Scotland, RCGP (Scotland) & BMA (Scotland). (2003) *GP Appraisal: A Brief Guide*. Edinburgh, Scottish Executive.

Scottish Executive Department of Health. (2004) *General Medical Services Contract*, Edinburgh, Scottish Executive.

Scottish Executive Department of Health. (2004), *General Medical Services Contract*, Edinburgh, Scottish Executive.

Scottish Executive (2006) NHS Circular: PCA(M)(2006)8 *Publication of 'revision to the GMS contract 2006/7 – delivering investment in general practice: Scottish guidance' available at* http:// www.sehd.scot.nhs.uk/pca/PCA2006 (M)08 accessed 18 June 2008

Scottish Office, Department of Health. (1998), *Clinical Governance,* MEL 75, Edinburgh, Scottish Office.

Scottish Patient Safety Programme (2007) available at: http://www.patientsafetyalliance.scot.nhs.uk/programme/ accessed May 11 2008.

Shackles, D., Ward, G. & Skinner, L. (2007) Learning to train: developing the framework for a Scottish national general practice trainers' course. *Education for Primary Care*, 18, pp.616-623 (8)

Shavelson, R.J. & Webb, N.M. (1991) *Generalizability theory: a primer*, California, Sage Publications.

Sheikh, A. & Hurwitz, B. (2001) Setting up a database of medical error in general practice: conceptual and methodological considerations, *British Journal of General Practice*, 51, pp. 57-60.

Sheldon, M. (1982) *Medical Audit in General Practice*, Occasional Paper 20, London, RCGP

Silverman, D. (2005) *Doing qualitative research*, Ch.11, pp149-168, London, Sage Publications

Sim, J, (1998) Collecting and analyzing qualitative data: issues raised by the focus group, *Journal of Advanced Nursing*, 28(2), pp. 345-352.

Slotnick, H.B. (*1996)* How doctors learn: the role of clinical problem solving across the medical school-to-practice continuum. *Academic Medicine, 71*, pp. *28*-34.

Smeele, I.J., Grol, R.P., van Schayck, C.P., van den Bosch, W.J., van den Hoogen H.J. & Muris J.W. (1999) Can small group education and peer review improve care for patients with asthma/chronic obstructive pulmonary disease? *Quality in Health Care*, 8, pp. 92-98.

Smith, J. (2003) *How to set up a peer review system*, in *Peer Review in Health Sciences* 2nd Edition, eds. Godlee, F. & Jefferson, T. London, pp.151-163, BMJ Publishing Group,

Southgate, L., Cox, J., Hatch, D., Howes, A., Johnson, N. *et al*. (2001) The assessment of poorly performing doctors: the development of the assessment programmes for the General Medical Council's Performance Procedures, *Medical Education*. 35, (s1), pp. 2–8.

Spencer, J. (2003) Learning and teaching in the clinical environment, *British Medical Journal*, 326, pp.591-594

Standing Committee on Postgraduate Medical Education. (1989), *Medical Audit: The Educational Implications*, London, HMSO

Stanhope, N., Crowley-Murphy, M., Vincent, C. O'Connor, A.M Taylor-Adam, S.E. (1999) An evaluation of adverse incident reporting. *Journal of Evaluation in Clinical Practice*, 5, pp. 5-12.

Stead, J., Sweeney, G., & Westcott, R. (2001) Significant event audit - a key tool for clinical governance, *Clinical Governance Bulletin* 1, pp.13-14.

Streiner, D.L. & Norman, G.R. (2003) *Health Measurement Scales* Ch.3, pp. 14-28. New York, Oxford University Press.

Sweeney, G., Westcott, R. & Stead, J. (2000) The benefits of significant event audit in primary care: a case study. *Journal of Clinical Governance*, 8, pp. 128–134.

Sykes, A., McKay. J,, Bowie, P., Lough, J.R.M. & Murray, T.S. (2003) Significant event analysis (Letter). Education for General Practice, 14(3) pp. 377-379.
Taft, R. (1955) The ability to judge people, *Psychological Bulletin*, 52, pp.1-23

The Shipman Inquiry. (2004) *Fifth Report safeguarding patients: Lessons from the past – proposals for the future,* available at http://www.the-shipman-inquiry.org.uk/fifthreport, accessed 18 June 2008

University of Leicester, (2009) *Online statistics: Descriptive statistics*, available at: http://www.le.ac.uk/bl/gat/virtualfc/stats/descrip.html accessed 10 Jan 2009.

University of the west of England. (2007) *Quantitative research*, available at: http://hsc.uwe.ac.uk/net/student accessed 10 Jan. 2009

Ursprung, R., Gray, J.E., Edwards, W.H., Horbar, J.D., Nickerson, J., Plesk, P. *et al.* (2005) Real time patient safety audits: improving safety every day. *Quality and Safety in Health Care,* 14(4), pp.284-289.

van den Hombergh, P., Grol, R., van den Hoogen, H.J.M., & van den Bosch, W.J.H.M. (1998) Practice visits as a tool in quality improvement: mutual visits and feedback by peers compared with visits and feedback by non-physician observers, *Quality in Health Care*, 8, pp.161-166.

van de Ridder, J.M.M., Stokking, K.M., McGaghie, W.C. & ten Cate, Olle T.J. (2008) What is feedback in clinical education? *Medical Education*, 42(2), pp.189-197.

van der Vleuten, C.P.M. (1996) The assessment of professional competence: developments, research and practical implications, *Advances in Health Sciences Education*, 1 pp. 41-67.

van der Vleuten, C.P.M., & Schuwirth L.W.T. (2005) Assessing professional competence: From methods to programmes, *Medical Education*, 39(3), pp. 309-317.

Vautrey, R. & Neal, R. (2000) Using peer review visits to carry out the clinical governance baseline assessment in one primary care group, *Journal of Clinical Governance*, 8 (3), pp.135-140

Veninga, C.M., Denig, P., Zwaagstra, R. & Haaijer-Ruskamp, F.M. (2000) Improving drug treatment in general practice, *Journal of Clinical Epidemiology*, 53 (7), pp. 762-772.

Verby, J., Holden, P. & Davies, R. (1979) Peer review of consultations in primary care: the use of audiovisual recordings. *British Medical Journal*, 1(6179), 1686-1688**.**

Verstappen W.H.J.M., van der Weijden,T., Sijbrandij, J., Smeele, I., Hermsen, J., Grimshaw, J. *et al.* (2003) Effect of practice based strategy on test ordering performance of primary care physicians, *Journal of the American Medical association,* 289, pp. 2407-2412.

Vincent, **C., Taylor-Adams, S. & Stanhope**, **N.** (1998) Framework for analysing risk and safety in clinical medicine, *British Medical Journal*, 316, pp.1154-1157.

Vincent, C., Stanhope, N. & Crowley-Murphy, M. (1999) Reasons for not reporting adverse incidents: an empirical study, *Journal of Evaluation in Clinical Practice*, 5, pp. 13-21.

Vincent, **C., Neale, G., & Woloshynowych, M.** (2001) Adverse events in British hospitals: preliminary retrospective record review, *British Medical Journal*, 322, pp. 517-519.

Vincent, C. (2004) Analysis of clinical incidents: a window on the system not a search for root causes, *Quality and Safety in Health Care*, 13, pp. 242-243.

Vincent, C. (2006) *Patient safety*, London, Elsevier. Vincent, C.(2006) *Reporting and learning system,* in Vincent, C, ed. *Patient Safety* pp. 57-74, Elsevier. Churchill Livingstone. (check).

Vincent, C. (2007) Incident reporting and patient safety. *British Medical Journal*, 334, pp. 51-52.

Wakeford, R., Southgate, L. & Wass, V. (1995) Improving oral examinations: selecting, training and monitoring examiners for the MRCGP, *British Medical Journal*, 311, pp. 931-935.

Waltz, C.W. & Bausell, R.B. (1981) *Nursing Research: Design statistics and computer analysis,* p. 71. Philadelphia: F.A. Davis.

Wass, V. Wakeford, R., Neighbour, R. & van der Vleuten, C.P.M. (2003) Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component, *Medical Education*, 37(2), pp. 126-131.

Weiner, E.A. & Stewart, B.J. (1984) *Assessing Individuals,* Boston, Brown.

Westcott, R., Sweeny, G. & Stead, J. (2000) Significant Event Audit in Practice: a preliminary study, *Family Practice*,17(2), pp.173-179.

Wilson, T., Pringle, M., & Sheikh, A. (2001) Promoting safety in primary care – research, action and leadership are required, *British Medical Journal*, 323, pp. 583-584.

Wilson, T. & Sheikh, A. (2002) Enhancing public safety in primary care, *British Medical Journal*, 324, pp. 584-587.

Wood, D. (2007) *Formative assessment*, in *Understanding Medical Education*, ed. Swanick, T., Edinburgh, Association for the Study of Medical Education.

World Health Organisation, (2005) WHO draft guidelines for adverse event reporting and learning systems, *World Alliance For Patient Safety* pp.8-9. Geneva, Switzerland, WHO.

Wu, A.W., Provonost, P. & Morlock, L. (2002) ICU incident reporting systems. *Journal of Critical Care*, 17, pp. 86-94.

Zhang, J., Patel, V., Johnson, T. & Shortliffe, E. (2004) A cognitive taxonomy of medical errors. *Journal of Biomedical Informatics,* 37, pp.193-204.

APPENDICES

Appendix 1: A copy of the postal questionnaire sent to GPs in Greater Glasgow

Greater Glasgow Primary Care Trust
West of Scotland Deanery

Significant Event Analysis in Primary Care: Questionnaire Survey of General Practitioners

Definition of a Significant Event:

"Any event thought by any member of the primary care team to be significant in terms of patient care or the conduct of the practice"

(Pringle *et al.,* 1995)

Part A – About You

Please **tick (✓)** appropriate responses:

1.) How long have you been a general practice principal?

☐ **0-5** years ☐ **6-10** years ☐ **11-15** years ☐ **15-20** years ☐ **>20** years

2.) What is your practice commitment? ☐ **Full-time** ☐ **¾ time** ☐ **½ time**

3) Please indicate if any of the following apply to you?

Please **tick (✓)** appropriate responses:

☐ GP Trainer ☐ GP Deputy Trainer

☐ Undergraduate Tutor ☐ RCGP Member

4.) Please indicate if any of the following apply to your practice?

☐ Training practice ☐ Non training practice

☐ Retaining practice ☐ Undergraduate teaching practice

☐ RCGP practice Accreditation Award ☐ RCP Quality Practice Award

5.) Approximately how many patients does your practice have?

☐ **<2000** ☐ **2000-4999** ☐ **5000-9000** ☐ **>9000**

6.) How many partners are in your practice?

☐ **Single-handed** ☐ **2-3** ☐ **4-6** ☐ **7-9** ☐ **10 or more**

7) Are you: ☐ **Male** ☐ **Female**

8) Which age group do you belong to?

☐ **25-34** years ☐ **35-44** years ☐ **45-54** years ☐ **55-64** years ☐ **> 64** years

Part B – Beliefs about SEA and incident reporting (edited section)

Please indicate whether you agree or disagree with the following statements about significant events and incident reporting: Please **tick (✓)** appropriate responses:

STATEMENT	Strongly Agree (1)	Agree (2)	Disagree (3)	Strongly Disagree (4)
The reporting of adverse events that happen in primary care should be mandatory	☐	☐	☐	☐
I would be willing to take part in a local anonymised incident reporting system	☐	☐	☐	☐
If a mandatory incident reporting system was introduced, I would be selective in what I reported	☐	☐	☐	☐
Significant events are often not acted upon in my practice	☐	☐	☐	☐

Part C – Barriers towards Significant Event Analysis (Edited section)

Please indicate whether you agree or disagree with the following statement about potential barriers towards doing significant event analysis:

Please **tick (✓)** appropriate responses:

	Strongly Agree (1)	Agree (2)	Disagree (3)	Strongly Disagree (4)	Don't Know (5)
Determining when an event is 'significant' can be difficult	☐	☐	☐	☐	☐

Part D – Definition of a Significant Event

We have given a definition of 'significant event' on the front page. If you do not agree with this definition, please explain why:

Part E – Comments

Please add any comments below:

THANK YOU VERY MUCH FOR YOUR TIME & ASSISTANCE
Please return the completed questionnaire using the pre-paid envelope provided.

Appendix 2: A copy of the initial covering letter sent as part of the questionnaire survey of GPs in Greater Glasgow

Dear Dr

Significant Event Analysis in General Practice

As you may know, significant event analysis (SEA) is increasingly being used as a way of improving patient care and safety. However, very little is known about the extent to which it is practised in primary care. We are looking to gather some basic information in this area, and would be very grateful if you would help us by taking part in this postal survey even if you have little or no experience of this topic.

We appreciate that you are often inundated with similar requests, but we believe it is very important to gain some understanding of GPs' knowledge and experiences of everyday significant events. Without your assistance we may be unable to do this properly and so your individual contribution to this survey is greatly valued.

To assist us, we would be pleased if you would complete the enclosed short questionnaire and return it to us using the pre-paid envelope provided by 5th December.

Please be assured that any information you provide will be treated in the strictest confidence. We will send all GPs a brief summary of the main findings shortly after completion of the survey.

Thank you very much in anticipation for your time and assistance.

Yours sincerely

Dr John McKay
Associate Adviser

Mr Paul Bowie
Associate Adviser

Appendix 3: A copy of the questionnaire reply sheet and significant event scenarios sent to selected educational groupings of GPs in Greater Glasgow and Clyde in 2005/6.

Significant Event Scenario 'A'

1). Please indicate if you think this particular significant event could realistically happen in general practice?

YES ❑ NO ❑

---

2). Using the scale below where **1**=Not at all significant and **7**=Extremely significant, please rate how "significant" you believe the scenario to be. (Please circle one number only)

1	2	3	4	5	6	7
Not at all significant						*Extremely significant*

---

3). Please indicate whether if the event happened to you (or your practice) how likely you (or the practice) would be to undertake a formal significant event analysis. (Please circle one number only)

1	2	3	4	5	6	7
Definitely not						*Definitely*

---

4). If you were to undertake a formal analysis of this particular significant event, which of the following individuals would you involve? (please tick all that apply)

☐ **1.** Not applicable	☐ **5.** Practice Manager	☐ **9.** Health Visitor
☐ **2.** Myself	☐ **6.** Reception/Admin Staff	☐ **10.** District Nurse
☐ **3.** GP colleagues	☐ **7.** Practice Nurse	☐ **11.** Pharmacist
☐ **4.** Other Health Professional	☐ **8.** Patient/Relatives	☐ **12.** Others

---

In the near future a confidential NHS reporting and learning system should be in place to allow "significant events" to be notified. The purpose of this system will be to allow relevant events to be reported locally so that others in general practice can be made aware of them and can learn from them.

5a). With this in mind, please indicate whether you think others in general practice may benefit from knowing about this particular significant event. (Please tick)

YES ❑ NO ❑

5b). If you were involved in this particular event, would you personally be prepared to officially report it if required to do so as part of a confidential system? (Please tick)

YES ❑ NO ❑

5c). If Yes, which of the following bodies would you be prepared to report this particular significant event to? (Please tick all that apply)

☐ **1.** Local Medical Committee	☐ **2.** Local Primary Care Organisation	☐ **3.** NHS Education for Scotland (Postgraduate General Practice)
☐ **4**. Community Health Partnership (formerly LHCC)	☐ **5.** Regional NHS Health Board	☐ **6.** National NHS Reporting System (i.e. National Patient Safety Agency)
☐ **7.** National Reporting System (Independent of the NHS)	☐ **8.** Other	

THANK YOU VERY MUCH FOR YOUR TIME & ASSISTANCE

Significant Event Scenarios.

Code	Significant Event Scenario
A	A patient was on long-term warfarin therapy as anti-thrombotic prophylaxis (having had insertion of a prosthetic aortic valve). For several years the patient's INR was being monitored on a monthly basis in the surgery. When going through the routine hospital letters, the patient's doctor received notification that the patient had been discharged from hospital having suffered an embolic CVA. The doctor, on reviewing the patient's INRs, had noticed that the most recent INR reading was 2.2 and that the preceding month to this his INR had been 1.8. His target INR however was 3.5. No alteration had been made in the patient's dosage of warfarin to try and bring the INR near the target level.
B	A patient handed in a form for a Blue Badge "disabled parking" for completion by the doctor. The receptionist attached the forms to the patient's notes and put them in the doctor's tray for his attention. When the doctor checked his tray the patient's notes were present but there was no Blue Badge form. It had become detached from the paper clip. The doctor asked the receptionist why the case notes were in his tray. He was told the reason and it was surmised that the form had become detached. After half an hour of searching the receptionist finally retrieved the form in amongst doctor's other paperwork.
C	A parent arrived at reception to collect a prescription for her son. The prescription, although generated, had not been signed and the "duty doctor" was asked to do so by the reception staff. The duty GP noticed that the telephone advice slip which the practice used for all telephone calls, was still attached to the child's file. The details on the advice slip did not seem to readily match with the treatment prescribed and the duty doctor then checked the patient's case notes and computer entry. Unfortunately there was no entry regarding a consultation that morning either in person or over the telephone; however, a prescription for Penicillin V syrup had been printed in the child's name. The duty doctor therefore spoke to the child's mother and a history was obtained and recorded in the files. A prescription for paracetamol was done and the printed prescription for penicillin was destroyed. In addition there were three previous entries for penicillin V with no data regarding its indication in the case notes.
D	A female patient attended the surgery with breakthrough bleeding and abdominal pain. Amongst other investigations a cervical swab was taken which returned as positive for chlamydia. She was prescribed azithromycin and advised that her partner should be treated (partner was not a patient at the practice). She re-consulted 4 months later with similar symptoms, which again gave a positive chlamydia result. On questioning, she was asked if her partner had been treated after the first attack. She replied that he had not, as he had no symptoms and thought it unnecessary. She was given further treatment and advised her that her partner must contact his GP or sexual health clinic for treatment.
E	A doctor noticed that a female patient had been making regular monthly appointments to see her. However during consultations the topic frequently changed to her ill husband and her concerns about him. The consultations were lasting 30 minutes resulting in the doctor running late. This was stressing the doctor and also causing frustration to the patients left waiting to be seen.
F	A 32 year old female patient attended the surgery complaining of anxiety symptoms. The doctor decided to prescribe some Inderal to help with physical symptoms. He did not notice in the patient's case notes that she had a previous history of asthma, although she was not currently taking any medication. The patient became dyspnoeic and wheezy overnight and called the emergency services. She was taken to hospital where a diagnosis of beta-blocker induced asthma was made. She was given appropriate treatment and was released the following day.
G	A GP received a phone call on his mobile during his lunch break from his wife. The phone call was to tell him that she could not pick up their children from the private nursery at 5pm as she was working late. The GP would therefore have to pick up their children after his surgery finished at 6pm. The nursery had been informed of this by his wife and were aware that the children would be picked up by their father nearer 6.30pm despite the fact that the nursery officially closed at 6pm. This meant that one of the nursery staff had to wait behind with the two children for half an hour after closing time and they made it clear that the nursery was not pleased with the situation.

Appendix 4: A copy of the initial covering letter sent to GPs in Greater Glasgow and Clyde as part of the questionnaire study in 2005/6.

JM/JM

Telephone : 0141 223 1462
Fax : 0141 223 1480

Enquiries to Dr John McKay
Email : john.mckay@nes.scot.nhs.uk

13th January 2005

«Name__Address»

Dear Dr «Name»

SIGNIFICANT EVENTS IN GENERAL PRACTICE

I am studying the potential to grade the seriousness of significant events in general practice and also whether different groups of GPs would be prepared to share or report these events if necessary. There is limited information available in this particular area and hopefully the findings will help inform the understanding of risk and safety in general practice. I am asking you to take part in this study, as your views as a *Trainer* are particularly important.

As a practising GP I am fully aware of the time pressures you are under, but would be extremely grateful if you could participate in this important study by simply reading the attached significant event scenarios and answering the associated questions for each scenario. Once complete please return all the paperwork using the pre-paid envelope provided. Please be assured that any information you provide will be treated in the strictest confidence.

If you wish to discuss any aspect of the study, please feel free to contact me on any of the above. Please note it is my intention to send all participants a one-page summary of the aggregated results of the study once the data have been collated and analysed.

Thank you very much in anticipation for your help.

Yours sincerely

Dr John McKay
General Practitioner/Associate Adviser

Appendix 5: Letter to request consent from GP authors of SEA report for participation in study on the classification of significant events.

Information and Consent Form: Classification of significant events submitted for peer review by general medical practitioners.

Dear Dr

Thank you for submitting your significant event analysis report for peer review. I hope you find the enclosed feedback helpful.

As part of the development of significant event analysis in general practice we would like to invite you to take part in the above study. Your involvement is limited to giving consent to analyse the content of your sea report(s). You can do this by signing the attached consent form and returning it in the prepaid envelope after you have read the information below.

The study has three aims:

- To classify the type and range of significant events submitted by GPs for educational peer review.
- To identify factors determined by GPs as contributing to the occurrence of significant events.
- To increase the understanding of the types of change, which practices, have implemented to minimise the chance of recurrence.

The study would involve two Associate Advisers independently reading the content of your significant event analysis to classify the type of event and identify any changes (if appropriate), which you have put in place to minimise recurrence. No information regarding your identity is available to either Associate Adviser, there is no identifiable patient data, and no written information contained within the report will be disseminated to anyone else or published elsewhere.

With your help we hope to be able to develop a system for classifying the types of significant event submitted for peer review and also develop a greater understanding of the reasons why events happen and gain insight into the types of changes implemented as a result of event analyses.

We hope that you would agree that the above educational research is important and worthwhile and that you would be willing to participate in this. Please note it is entirely up to yourself to decide whether or not to take part and that you are free to withhold your consent at any time without giving a reason. If you do not take part in the study or withdraw from the study, this will not affect your anonymised peer review feedback in any way.

The results of the study will be posted on the NES website. If you are happy to take part I would be grateful if you could sign the consent form overleaf and return it in the pre-paid envelope.

Thank you in anticipation of your help.

Yours sincerely

Dr John McKay
Associate Adviser

Appendix 6: Data collection sheet for the content analysis of SEA reports

1. SEA Submission Details:

IDN: ☐ Registrar ☐ Principal ☐ Title: ☐

2. Peer Assessment Outcome:

Assessors' Scores: **A1:** Sat Uns ☐ **A2:** Sat Uns ☐ **A3:** Sat Uns ☐ **A4:** Sat Uns ☐

3. Significant Event Details:

What happened: Code **1** ☐ **2** ☐ **3** ☐ **4** ☐

Description:

	Circle:
Interface sig. event:	Y
Secondary care:	Y
Out-of-Hours:	Y
Psychiatric care:	Y
Social work department:	Y
Community pharmacy:	Y
Other:	Y

Type of event: Negative: ☐ Positive: ☐ Purely reflective: ☐ Other: ☐ ……

4. NPSA Scale:

Grade: Low ☐ Mod ☐ Sev. ☐ Death ☐ N/A ☐ **Harm Prevented:** Yes ☐ No ☐ N/A ☐

5. Reason(s) for Event? **YES** ☐ **NO** ☐ **N/A** ☐

Code 1:	☐	
Code 2:	☐	
Code 3:	☐	
Code 4:	☐	

Contributory factors (circle): ***a***. *Patient factor/**b**. Individual staff/**c**. Task/**d**. Communication/**e**. Team & Social/**f**. Education & Training/**g**. Working Condition & Environment Event/**h**. Equipment & Resources/**i**. Other*

6. Learning from Event? **YES** ☐ **NO** ☐

Code 1:	☐	
Code 2:	☐	
Code 3:	☐	

Team involvement (circle): GP/Partners/PM/PN/A&C Support/HV/DN/DN/Pharm/SWD/Hosp/NH/RH/Other

7. Action Taken? **YES** ☐ **NO** ☐ **N/A** ☐

Code 1:		
Code 2:		
Code 3:		
Code 4:		

Team involvement (circle): GP/Partners/PM/PN/A&C Support/HV/DN/DN/Pharm/SWD/Hosp/NH/RH/Other

Appendix 7: Content Validity Exercise: Cover letter and guidance notes to SEA Experts.

JM/JM

Telephone : 0141 223 1462
Fax : 0141 223 1480

Enquiries to Dr John McKay
Email : john.mckay@nes.scot.nhs.uk

17 May 2004

«Name__Address»

Dear «Name»

Proposed Peer Assessment Schedule for Significant Event Analysis

I am conducting research into a proposed assessment schedule for the external peer review of significant event analyses undertaken by general practitioners. The purpose of the assessment schedule is to provide formative educational feedback on improving the analysis to the submitting GP, where deemed necessary by GP peers.

The intended research aims to improve and update an existing assessment schedule which is currently used in our deanery to provide feedback to GPs submitting SEA reports under the former PGEA system [1,2,3]

As a leading expert in this area I should be very grateful, therefore, if you would kindly assist my research by reading both the enclosed guidance together with Document A, and then completing Document B and returning it in the stamped addressed envelope provided as soon as possible.

I would hope at some stage in the future to publish this work and will, of course, acknowledge your contribution.

If you have any queries then please contact me at the e-mail address above.

Thank you very much in anticipation for your help.

Yours sincerely

Dr John McKay
GP/Associate Adviser

1. McKay J, Bowie P, Lough M. Evaluating significant event analysis: implementing change is a measure of success. *Education for Primary Care 2003* **14**(1): 34-38
2. Bowie P, McKay J, Lough M. Peer assessment of significant event analyses: being a trainer confers an advantage. *Education for Primary Care 2003* **14**(3): 338-344
3. Lough M. The development of integrated audit for the training of registrars in general practice. MD Thesis, 2003, University of Glasgow

GUIDANCE NOTE

Dear Colleague,

Proposed Peer Assessment Schedule for Significant Event Analysis

The relevance of SEA and peer review has increased considerably in the past five years and both are now established as important components of clinical governance.

In an effort to address the difficulties in providing constructive feedback to general practitioners and their teams, we wish to review all those measurable items that may be important in determining if the analysis of a significant event is satisfactory or not. It is necessary, therefore, that any proposed items to be included as part of this process are subject to a formal assessment of validity and reliability.

The first step in this process is to ensure that the proposed assessment tool has content validity i.e. the extent to which the tool contains the appropriate items to be measured. To help us do this, we would be pleased if you would review the proposed SEA assessment tool, Document A, which is attached (*Document A is Figure 4,Chapter 9*). The tool contains 10 items, each with a 7-point rating scale. Please read this carefully, as this is our initial attempt at developing the items, which we feel should be included in a new assessment tool for SEA. **You are not required to actually complete or write any comments on Document A.**

After you have read Document A, please turn to Document B – the content validity index (CVI) which is also attached (*Document B is Figure 5, Chapter 9*). The CVI is used to measure your own level of agreement for each proposed item in the assessment schedule. Please complete Document B by rating each proposed item using the 4-point scale, the key for which is outlined at the top of the page. IF you score any item as 1, 2, or 3, please add any comments you may have about the relevance of this item for use in an SEA assessment schedule.

At the end of Document B, there is space for you to add any items, which you feel should be included in an assessment schedule for SEA.

Please return Document B using the prepaid envelope provided.

Thank you very much in anticipation for help.

Appendix 8: Information and consent form for focus group participation (Chapter 11)

Title of Study: Assessors views and experiences of assessing and giving feedback on SEA reports using a proposed new peer review instrument

Dear Doctor xxxx,

We would like you to take part in a study. Before you decide to do so it is important that you understand why this research is being done and what it will involve. We would be grateful if you would take the time to read the following information carefully. If there is anything that is not clear to you or if you would like more information, please contact the researcher at the address given overleaf.

What is the purpose of the study?

As you know, the Audit Development Group has been developing the proposed new SEA peer review instrument over the past year. We are currently in the process of evaluating the instrument. As an assessor who has participated in the use of this new instrument we are interested in your experience of using the new assessment instrument and your thoughts on the comparative merits and drawbacks of the new and old instrument. We would also like to explore your ideas on the further development of the instrument.

Why have I been chosen?

We have selected you as a member of the Audit Development Group.

Do I have to take part?

Taking part in the research is entirely voluntarily. It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and asked to sign a consent form. If you do decide to take part you are still free to withdraw at any time without giving a reason.

What will happen to me if I take part?

The interview will last between 60 and 75 minutes. You will be interviewed as part of a focus group of between six and eight people on only one occasion. The researcher will introduce topics for you to discuss with him/her and with your agreement we will tape record the discussion. This will be transcribed anonymously. Once this material is typed up you will be given a copy to see if it is accurate to gain your agreement to study the transcript. When the transcript has been analysed the researchers will write back to you with details of the main themes of the discussion identified and asked if you agree with this. We hope to hold the interviews at 2 Central Quay in Glasgow.

What is the risk of taking part?

The research team do not think there is any serious risk to taking part.

What are the possible benefits of taking part?

The researchers hope to gain greater insight into the benefits and drawbacks of using the new instrument and how to improve the quality of feedback given to our fellow GPs using this model of peer review.

Will I be reimbursed?

To cover the cost of locum reimbursement for your time will make a payment of £100. Any travelling expenses will be reimbursed.

Will my taking part in this study be kept confidential?

All information, which is collected during the course of research, will be kept strictly confidential. Any information about you, which you give, will be anonymised so that you cannot be recognised.

What will happen to the results of the research?

The research team will analyse the comments and thoughts produced by the focus group interview and look for emerging themes. We will produce a report for the PCO/Partnership Steering Group explaining what these are. The research team hope to publish these findings in a journal to allow other researchers to learn from our experiences. All quotes and experiences will be anonymised so that anyone reading any one report will not be able to identify you.

Who is organising the funding?

NHS Education (NES) for Scotland is funding the research.

Who has reviewed the study?

The study design has been reviewed and approved by the Research & Development Department of NHS Education for Scotland. The study has been reviewed by the Chairman of the Multi-Centre Research Ethics Committee for Scotland who has decided that the study does not require ethical approval.

Contact for further information: Dr John McKay, NHS Education for Scotland, 2CQ, 89 Hydepark Street, Glasgow G3 8BW

NHS Education for Scotland

CONSENT FORM

Title of Study: **Reviewers views and experiences of assessing and giving feedback on SEA reports using a proposed new peer review instrument.**

Name of Researcher: Dr John McKay

1. I confirm that I have read and understand the information and consent sheet for the above study and have had the opportunity to ask questions. ❑

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, or legal rights being affected ❑

3. I agree to take part in this study. ❑

4. I agree to a group interview being audio-taped and transcribed ❑

……	……..	……..
Name of Participant	**Date**	**Signature**

……	……..	……..
Name of Researcher	**Date**	**Signature**

Appendix 9: Initial topic guide for focus group study (Chapter 11)

a) Describe your experiences of a) assessing and b) giving feedback on SEA reports using the new feedback instrument?

b) How does the new instrument compare with the existing (old) instrument?

c) In your opinions, how could the new instrument be improved further?

d) Describe your experiences of the initial and ongoing training you received for using the new instrument?

e) How do you as assessors perceive your own feedback skills?

f) Do you have any other comments or issues on giving feedback using the new instrument which you wish to raise which has not yet been brought up in the group?

Appendix 10a: Letter to MDDUS on SEA submitted for Peer Review

JM/JM

Telephone : 0141 223 1462
Fax : 0141 223 1480

Enquiries to Dr John McKay
Email : john.mckay@nes.scot.nhs.uk

8th November 2004

Dr J Rodger
Medical Adviser
MDDUS
Mackintosh House
120 Blythswood Street
GLASGOW
G2 4EA

Dear Jim,

Re Significant Event Analyses Submitted for Peer Review.

Thank you very much for taking the trouble to meet with Murray and myself on the 25th October.

I would like to offer a summary of the issues which we discussed and any action which we would be, required to take.

As you know the current system involves individual general practitioners submitting an analysis of a "significant event" to the department in a structured format. The event is anonymised by myself and then sent to two peer assessors who then provide educational feedback using a specific marking schedule where this is deemed necessary. The feedback is collated by myself and a letter sent to the submitting doctor detailing the peer assessors' comments.

We know from over 700 event analyses submitted that only a small minority describe episodes of severe patient harm.

One of our concerns was regarding what action we would be required to take in the situation where a practitioner does submit an event which has unintentionally resulted in patient harm? An example we described was of a submission where a practitioner had identified and analysed an event where control of a patient's INR had been sub-optimal and the patient subsequently suffered a CVA.

My understanding now is that in an isolated case such as this, where the practitioner has submitted their analysis for educational peer review, if he or she has demonstrated insight into the event and made appropriate change where required, then we need do nothing further than offer back to the practitioner the assessors comments and suggestions. In a situation where the

peer assessors feel that the analysis is not satisfactory, then our duty is to inform the practitioner of any comments as to why this is, but that it is not our responsibility to ascertain that the practitioner has taken any additional action which had been suggested by the assessors.

If, however, there are multiple event analyses submitted by the practitioner and several of these did give cause for concern in terms of patient safety or practitioner behaviour, then this may represent a performance issue for the practitioner. In this situation the best course of action would be for myself to pass this to Dr Lough and Professor Murray to refer to the local primary care organisation to deal with.

With regard to a significant event where there would be immediate cause for concern regarding patient safety, then we would follow the guidance in Good Medical Practice paragraphs 26 and 27 on dealing with problems in professional practice.

I would be grateful if you would acknowledge whether this is a representative summary of our discussion on the issues raised.

Kind regards.

Yours sincerely

Dr John McKay
Associate Adviser (Audit)

Appendix 10b: Reply letter from MDDUS

MDDUS
EXPERIENCE IN PRACTICE

THE MEDICAL AND DENTAL DEFENCE UNION OF SCOTLAND

Dr Jim Rodger
BSc MB ChB BA MBA FRCGP FRCP(Ed) DMJ

Please quote our reference on all correspondence
Our Ref: L130 JR/SO
Your Ref: JM/JM

14 January 2005

Dr John McKay
Associate Adviser
NHS Education for Scotland
3rd Floor, 2 Central Quay
89 Hydepark Street
GLASGOW
G3 8BW

Dear John

Re: Significant Event Analyses

Thank you very much for your letter of 10 January. I do apologise – you had indeed written to me in November but I had mislaid your letter on my desk!

Your summary of discussions is very clear and in fact is clearer than I remember in our discussions.

I would have no hesitation in suggesting that that is a very appropriate way to deal with such matters. I would be pleased to hear of any instances where you have had to apply the advice to difficult cases.

Once again, I do apologise for the delay. It was nice to meet with you and Murray.

Kind regards.

Yours sincerely

S. O'Byrne

Dr J Rodger
MEDICAL ADVISER

REGISTERED OFFICE
MACKINTOSH HOUSE
120 BLYTHSWOOD STREET
GLASGOW G2 4EA

TELEPHONE 0141 221 5858 FACSIMILE 0141 228 1208
E.MAIL: INFO@MDDUS.COM INTERNET: WWW.MDDUS.COM
CHIEF EXECUTIVE & SECRETARY: PROFESSOR GORDON C A DICKSON MLitt PhD FCII FIRM
REGISTERED IN SCOTLAND No. 5093

14/01/05McKay

Appendix 11: Example of Prescribing Error SEA Report and Educational Feedback Provided

Title: Prescribing Error
Date of Significant Event:
Date of Event Analysis:
Lead Investigator:
Permission to use for Educational Purposes: **Yes**

1. What happened?

(Describe what actually happened in detail. Consider, for instance, how it happened, where it happened, who was involved and what the impact or potential impact was on the patient, the team, organisation and/or others).

I was the on-call GP for the practice. A member of staff asked me to sign a repeat prescription for a patient unknown to me. As the patient had run out of tablets I was asked to sign the prescription as he was waiting at the reception desk. The script was for Amitriptyline but the dose appeared to be incorrect so I asked for the patient's notes to confirm what the consultant psychiatrist had requested the patient be commenced on. It was then that I noticed that the hand written request had asked for Amisulpiride to be commenced. The patient had a history of psychosis. This was confirmed by checking the consultant's dictated letter. I therefore changed the prescription to the correct dose of Amisulpiride and explained the change to the patient, who was still clinically stable. He accepted the apology after an explanation. However, it does not alter the fact that this patient had been taking the wrong medication for 2-months the potential result that there could have been a recurrence of his psychosis and all that that may have entailed.

2. Why did it happen?

(Describe the main and underlying reasons – both positive and negative – contributing to why the event happened. Consider, for instance, the professionalism of the team, the lack of a system or a failing in a system, lack of knowledge or the complexity and uncertainty associated with the event).

On investigation it transpired:

- A member of staff had misread the medication requested on the hand written note, and had therefore typed the wrong medication into the computer for the acute prescription.
- The script had been presented to the GP without the hand written request from the hospital. It had been a busy time in the practice and he had signed the script assuming it was the correct medication.
- On review of the hand written hospital request by staff involved it could be seen how the mistake had been made due to the poor quality of the doctor's handwriting.

3. What has been learned?

(Demonstrate that reflection and learning have taken place on an individual or team basis and that relevant team members have been involved in the analysis of the event. Consider, for instance: a lack of education & training; the need to follow systems or procedures; the vital importance of team working or effective communication).

- Unfortunately it is a normal expectation for many the handwriting from many doctors to be poor, resulting in poor communication and the potential for serious errors to occur as a result. Caution must always be exercised when reading and interpreting had-written scripts.
- It was made clear to me and the practice team that errors in prescribing can so easily occur if work pressure exists and handwriting is so poor that it can be misinterpreted, particularly by non-clinical staff.
- Safety-nets within the practice structure are needed to prevent this happening again.

4. What has been changed?

(Outline the action(s) agreed and implemented, where this is relevant or feasible. Consider, for instance: if a protocol has been amended, updated or introduced; how was this done and who was involved; how will this change be monitored. It is also good practice to attach any documentary evidence of change e.g. a letter of apology to a patient or a new protocol).

In view of the error a practice meeting was arranged to discuss the significant event. The meeting included members from all the different teams in the practice, and was conducted in a non-confrontational manner. It was made clear how the error had occurred following discussion with the team members, as described above.

Following discussion and team agreement the following changes were introduced to the prescribing procedure within the practice, which the practice manager would lead on:

1. Hand-written requests from the hospital were to be collected by the patient 48-hours after being handed in to reception, unless urgent.
2. All hand written hospital requests were to be presented to the patient's GP, who was then to write the prescription.
3. Staff involved in prescribing were to change their work environment to a quieter room, away from distractions.
4. It was decided that all GPs should sign their prescriptions in their rooms, again away from any distractions.

How can this be prevented from happening again?

It was decided to review the situation with staff at a practice meeting within the next quarter to ensure that the changes had been successfully implemented, and that no similar errors had occurred.

PEER REVIEW FEEDBACK

Dear Colleague

SEA Report – Prescribing Error

Thank you very much for submitting your SEA report for educational peer review. I now have the feedback from both reviewers on your audit of this significant event and have summarised this below for your consideration:

- Both reviewers were in full agreement that this was a very important significant event which was worthy of analysis.
- A good description of the event was provided which was clear, concise and easy to follow. The reviewers did comment that it was unclear if the patient was given a 1-month or 2-month supply of the drug initially. If the former was the case then it is possible that the error may in fact have occurred twice before being noticed.
- Both reviewers thought more detail could have been provided in explaining why the event had occurred. For example, providing a clearer picture of the normal system for dealing with hand written hospital outpatient prescriptions would have been helpful. Also, the reviewers thought it unusual for a non-clinician (it would have been helpful to know the occupation at this stage) to be given the responsibility of interpreting a hand-written request, the information from which is then put on the repeat prescription system – a point which you commendably explore in the learning and reflections section of the report.
- In terms of insight demonstrated as a result of the analysis, both reviewers commented that the team had reflected well on the event and identified appropriate learning needs. However, one did raise the continued risk associated with non-clinicians adding/altering prescriptions to the system.
- The actions agreed and implemented by the team were considered by the reviewers to be helpful in terms of reducing the chances of this type of event recurring in future. However, they also made a number of further points for you to consider:

    - It would have been very useful if you had informed and discussed the event with the hospital specialist because of their duty of care to the patient and also to bring the handwriting situation to their attention.
    - How will the new system hold up if the GP is on holiday?
    - What if the patient refuses to wait for 48 hours (or 2 working days?) or if the prescription is considered urgent - what is the practice system in this instance?
    - What system is in place to stop the computer operator inadvertently adding the wrong drug to the repeat prescribing list (i.e. to pick up human error)?

You may have already thought about and addressed some, or all, of the points made by the reviewers but simply not included them in your report. If this is the case then we hope that you will find the comments reinforce the effectiveness of your SEA. If this is not the case, we hope that you will find the comments useful for strengthening your learning on this SEA and for supporting you in preparing future SEA submissions.

Thank you once again for submitting your SEA report for peer review
Yours sincerely

Appendix 12: Information and Consent Form for Semi-structured Interview of General Practitioners (Chapter 12)

Title of Study: Participants views on the experiences of undertaking significant event analysis and submitting an event analysis report for peer review.

Dear Dr XXX

We would like to invite you to take part in a research study. Before you decide, it is obviously important that you understand why the research is being done and what it will involve. We would be grateful if you would take time to read the following information carefully. If there is anything that is not clear to you or if you would like more information, please contact the researchers at the address given overleaf.

Thank you for reading this.

What is the purpose of this study?

Peer review of general practitioners significant event analyses is available in the west of Scotland as part of continuing professional development. We are currently in the process of evaluating a new formative review instrument to enhance the feedback, which is given to participants. As someone who has participated in the system of peer review, and had feedback based on reviewers' comments using the new instrument we would be interested in your experience of undertaking a significant event analysis and your thoughts and opinions on the peer review feedback you received.

Why have I been chosen?

We have selected nine general practitioners who have submitted SEA reports for peer review in the west of Scotland. The GPs have been chosen to represent those who received different levels of feedback as determined by the score attributed to their SEA report using the new formative review instrument.

Do I have to take part?

Taking part in the research is entirely voluntarily. It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to

keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part?

The interview will last between 45 and 75 minutes. You will only be interviewed once. The researcher will introduce topics for you to discuss with him/her and with your agreement we will tape record the discussion. This will be transcribed anonymously. Once this material is typed up you will be given a copy to see if it is accurate and to gain your agreement to study the transcript. When the transcript has been analysed by the researchers we will write back to you with details of the main themes of the discussion identified and ask if you agree with this. We hope to hold the interviews either here at 2 Central Quay in Glasgow, or at your practice, whichever is most convenient for you.

What is the risk of taking part?

The research team does not think there is any serious risk to taking part. It may be however, that discussing a particular significant event may cause you distress in which case we would provide you with details of counselling available from the Occupational Health Unit.

What are the possible benefits of taking part?

The researchers hope to gain greater insight into the sea technique and improve the quality of feedback given to participants in this model of peer review. We think that learning about your experiences of this system will lead to further improvement in the system and enhance the educational benefit to yourself and any other GPs submitting SEA reports for peer review in the future.

Will I be reimbursed?

To cover the cost of locum reimbursement for your time we will make a payment of £75. Any travelling expenses will also be reimbursed.

Will my taking part in this study be kept confidential?

All information, which is collated during the course of the research, will be kept strictly confidential. Any information about you, which you give, will be anonymised so that you cannot be recognised.

What will happen to the results of the research study?

The research team will analyse all the comments and thoughts produced by the interviews and look for emerging themes. We will produce a report for the PCO/Partnership Steering Group explaining what these are. The research team hope to publish the findings in a journal to allow other researchers in the field to learn from your experiences. All quotes and experiences will be anonymised so that anyone reading any report will not be able to identify you.

Who is organising and funding the research?

NHS Education (NES) for Scotland is funding the research.

Who has reviewed the study?

The study design has been reviewed and approved by the Research & Development Department of NHS Education for Scotland. The study has been reviewed by the Chairman of the Multi- Centre Research Ethics Committee for Scotland who has decided that the study does not require ethical approval.

Contact for further information:

Dr John McKay
NHS Education for Scotland
Third Floor
2 Central Quay
89 Hydepark Street
Glasgow G3 8BW
0141 223 1456

If you wish to take part in this study please fill in your details below and return this in the stamped address envelope provided.

✂ ----------

Title of Study: General practitioners' views on their experiences of analysing significant events and submitting these for educational peer review.

I am willing to take part in the above study:

Name: Dr **Phone Number**:

Address or practice stamp:

If you reply that you are happy to take part we will contact you to arrange a convenient time and venue. We enclose overleaf a copy of the consent form that we will ask you to sign prior to the interview.

Thank you in anticipation for taking part in this study.

CONSENT FORM

<u>Title of Study</u>: Participants views on the experiences of undertaking significant event analysis and submitting an event analysis report for peer review.

Name of Researchers: Dr John McKay

1. I confirm that I have read and understand the information sheet dated (xxx) for the above study and have had the opportunity to ask questions. ❑
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, or legal rights being affected ❑
3. I agree to take part in this study. ❑
4. I agree to a group interview being audio-taped and transcribed ❑

……	……..	……..
Name of Participant	**Date**	**Signature**

……	……..	……..
Name of Researcher	**Date**	**Signature**

1 for participant, 1 for researcher

Appendix 13: One-to-one Semi-structured Interviews - Initial Topic Guide

1. What factors made you decide to analyse this event rather than others, which have taken place in your practice?
2. What influenced you to write up this significant event analysis?
3. Why did you decide to submit this event analysis for peer review?
4. Describe your experience of the process of submission;
5. In your opinion, how could the process of submission be improved upon?
6. What was your experience of the peer review feedback?
7. How could this be improved upon?
9. Did you experience any educational benefit from participation?
10. Did you use the feedback in your appraisal and if so how?