

in association with



Peer Review of Health Care Professionals: A Systematic Review of the Literature

Prepared for the Australian Commission on Safety and
Quality in Healthcare

June 2009

Table of contents

INTRODUCTION	3
Types of publications considered for the reviews	5
Types of participants	5
Types of interventions	5
Types of study measures	5
Search strategies for identifying publications.....	6
Search terms for identifying publications	6
Publication selection.....	9
Quality assessment of studies.....	10
Data extraction	11
Data analysis.....	11
RESULTS	12
Studies Identified.....	12
Study Types.....	12
Participants and Settings.....	12
Peer Review Information Sources	13
Number of Peer Reviewers	14
Methodological Quality	14
Categorical analysis	14
DISCUSSION	24
Application of peer review to performance assessment	24
Effectiveness of peer review.....	25
Reliability and Validity of Peer Review Methods.....	26
Features of the peer review process	27
Voluntary peer review methods	29
Peer review for under-performance	30
Limitations of this review and future directions	30
Implications for Future Research.....	31
REFERENCES	32
APPENDIX 1 – SUMMARY TABLES.....	41
TABLE 1 – ASSESSMENT OF THE CLINICIAN’S DOMAINS OF PROFESSIONAL PRACTICE	41
TABLE 2 – ASSESSMENT OF DELIVERY OF CARE IN ACCORDANCE WITH CLINICAL GUIDELINES.....	54
TABLE 3 – ASSESSMENT OF ORGANISATIONAL QUALITY OF CARE.....	61
TABLE 4 – PEER REVIEW AS A REQUIREMENT FOR CONTINUING PROFESSIONAL DEVELOPMENT	69
TABLE 5 – ASSESSMENT OF SIGNIFICANT EVENTS	71
TABLE 6– QUALITY ASSURANCE OF RADIOLOGY / PATHOLOGY PRACTICE.....	91
TABLE 7 – PEER REVIEW FOR THE PURPOSES OF CREDENTIALLING HEALTH CARE PROVIDERS.....	95



TABLE 8 – ASSESSMENT OF SUSPECTED UNDER-PERFORMANCE OF THE HEALTH CARE PROVIDER.....98

INTRODUCTION

Peer review can be defined as the evaluation of the creative work or performance of an individual by other people in the same field (Evans et al. 2004). The aim of peer review is to assist in the formulation of informed judgments about the professional practice of health care professionals, with the ultimate goal of identifying ways to improve and maintain quality of care (Thomas 2002; Bowie et al. 2005).

A peer is a health care professional with equivalent experience in similar organisational environments who also has the knowledge and skills to contribute to the assessment of fellow medical practitioner's clinical and professional performance (Norcini 2003).

Peer review involves collection of information regarding the professional practice of the health care provider. This information is appraised by the professional's peers, often against standards, performance indicators and / or outcome parameters. Professional judgment by a peer is seen as a valid method of making informed judgments on professional practice (Heaton 2000; General Medical Council 2009).

A common element of peer review is performance assessment. Performance assessment can be defined as the evaluation of one or more elements of clinical practice against accepted standards of practice (Wass et al. 2003; Steinbrook 2005). There are multiple domains of professional practice across which performance assessment may be relevant. These include clinical expertise (diagnostic and treatment skills and knowledge), communication (with patients, families, colleagues and in record-keeping), management (of time, resources, within systems and personally), scholarship (including life-long learning, teaching and research) and professionalism (including collaboration, teamwork, honesty, integrity, probity, respect for patients and ethical practice) (McAvoy et al. 2001).

Once information is collected by peer review methods, feedback may then be provided to the health care professional (Spigelman and Swan 2003). Clinical standards are not static and are constantly interpreted and adapted to changing circumstances (Kuhn 2006). In many circumstances, peer review is the only mechanism to judge the professional practice of others working in the same field and in similar settings, because the peer regularly performs similar work and possesses the relevant expertise to evaluate it (Rethans et al. 1991; Evans et al. 2004; Leape and Fromson 2006).

Assessments motivate professionals to learn, provide feedback to the professional regarding how much they have learned and whether they have reached a required standard, and enable the professional to make decisions about future learning needs (Brown and Doshi 2006; Overeem et al. 2007). The individual practitioner may perform self-assessment to gauge performance, or external methods of assessment may be used.

There are many factors to consider in the design and conduct of peer review, including who participates, the information on which conclusions are based, the conduct of peer review processes including the avoidance of bias and the documentation of processes and

outcomes. It is important that policy-makers and educators in health care who are responsible for establishing and operating systems for the assessment of individual health care professionals have good insights into the application of peer review to the assessment of health care professionals, and are familiar with the different peer review methods available (Sibley et al. 1982; Thomson et al. 2001).

Not all medical staff participate in peer review. Among those that do participate, the degree of participation may vary. For example, as part of the Royal Australasian College of Surgeons (RACS) Continuing Professional Development (CPD) Program all surgeons who conduct operative procedures in hospitals, day surgery units or private rooms are required to participate in a surgical audit each year, and to submit the audit for peer review (RACS 2006; RACS 2008). Most other Australian specialty colleges do not currently have a requirement for compulsory peer review activities as part of CPD (Watters et al. 2006).

The aim of this systematic review was to appraise the published literature regarding the review of health care professionals by their peers in order to examine the purposes of peer review across health care settings, describe the peer review methods being used in the peer review of health care professionals, and to assess their effectiveness.

Effectiveness of peer review can be appraised according to a range of theoretical frameworks. For the purposes of this review, the four levels of effectiveness appraised are based on a theoretical framework applied to effectiveness of continuing medical education (Curran and Fleet 2005). These are:

1. Participant satisfaction
2. Learning outcomes
3. Performance improvement (either reported or measured)
4. Patient outcomes

and will be considered in this systematic review.

METHODS

Types of publications considered for the reviews

Publications where clinical peer review was a topic for study were considered for inclusion. Both qualitative and quantitative study types were considered.

Publications reporting methods of peer review (defined for the purposes of coding as any appraisal of healthcare professionals performed by other similar health care professionals), using described methods for appraisal, and including but not limited to written, electronic or verbal methods, were sought.

Publications reported in a language other than English, and publications assessing peer review for research evaluation purposes (including for appraisal of journal manuscripts or research grant applications), were excluded.

Types of participants

Healthcare professionals responsible for patient care were the primary participant type of interest. Publications whose participants were from healthcare-related professions, or who were healthcare professionals not providing patient care (e.g. fulfilling administrative functions) were also considered for inclusion.

Types of interventions

Publications that stated peer review had occurred were excluded unless a description of the peer review methodology was provided.

It was common for peer review methods to be used for a range of purposes and to be applied in conjunction with other performance assessment methods. Publications therefore included the following subcategories:

1. Evaluation of single peer review method for performance assessment purposes;
2. Comparative assessment of different performance assessment peer review methods;
3. Comparative assessment of peer review with non-peer review performance assessment methods; and
4. Evaluation of peer review methods for purposes other than performance assessment.

Types of study measures

The main study measures assessed included the following:

1. Clinical expertise
 2. Communication
-

3. Management
4. Scholarship
5. Professionalism
6. Patient outcomes.

Additional measures assessed included the following:

1. Evidence of quality improvement occurring in clinical practice as a result of peer review activity; and
2. Change in professional's knowledge, attitudes and / or behaviour as a result of the peer review activity occurring.

Search strategies for identifying publications

The search strategies used to identify publications were as follows:

1. Manuscripts were sought for the years 2004 to 2009 in the first instance. However, the literature review was extended to include publications conducted between 1998 and 2009 in order to increase the identification of relevant publications.
2. Searches were conducted in the MEDLINE (1998 – current), PsychINFO (1998 – current), CINAHL (1998 – current), EMBASE (1998 – current), Dare (1998 – current) and Cochrane (1998 – current) databases (including the Cochrane Effective Practice and Organisation of Care – EPOC database) for English language publications, and were supplemented with searches of proprietary search engines (including Google®, Google Scholar® and Yahoo®).
3. Direct analysis of output from known centres of excellence, international, national and state-based government agencies was also conducted.
4. Databases of ongoing trials searched included Current Controlled Trials (www.controlled-trials.com) and The National Research Register (www.update-software.com/National/nrr-frame.html)
5. The reference lists of review publications and of all included publications were searched in order to find other potentially eligible publications.

Potential missing and unpublished publications were going to be sought by contacting experts in the field. This was not necessary.

Search terms for identifying publications

Peer Review



1. Health care economics and organizations// peer review, health care [MeSH term, all sub trees and subheadings included]
2. Health services administration// peer review, health care [MeSH term, all subheadings included]
3. Health care quality, access and evaluation// peer review, health care [MeSH term, all subheadings included]
4. peer review [in abstract or title]
5. (peer adj (assess* or perform* or compl* or apprais* or credential* or re-certif* or certif* or licens*)) [in abstract or title]
6. or/1-5

Clinical studies

1. *randomized-controlled-trial in pt*
 2. *"randomized-controlled-trials"/ all subheadings*
 3. *"random-allocation" in MIME, MJME*
 4. *random* or alloc* or assign**
 5. *(#4 in TI) or (#4 in AB)*
 6. *#1 or #2 or #3 or #5*
 7. *controlled-clinical-trial in pt*
 8. *clinical-trial in pt*
 9. *explode "clinical-trials"/ all subheadings*
 10. *(clin* near trial*)*
 11. *(#10 in TI) or (#10 in AB)*
 12. *"cross-over-studies" in MIME, MJME*
 13. *cross-over near (stud* or trial* or design*)*
 14. *crossover near (stud* or trial* or design*)*
 15. *#7 or #8 or #9 or #11 or #12 or #13 or 14*
 16. *"double-blind-method" in MIME, MJME*
-



17. *"single-blind-method" in MIME, MJME*
 18. *(singl* or doubl* or trebl* or tripl*) near (blind* or mask*)*
 19. *(#18 in TI) or (#18 in AB)*
 20. *#16 or #17 or #19*
 21. *"control"/ all subheadings*
 22. *control* in TI*
 23. *control* in AB*
 24. *#21 or #22 or #23*
 25. *explode "research-design"/ all subheadings*
 26. *TG=comparative-study*
 27. *explode "evaluation-studies"/ all subheadings*
 28. *"follow-up-studies" in MIME, MJME*
 29. *"prospective-studies" in MIME, MJME*
 30. *control* or prospectiv* or volunteer**
 31. *(#30 in TI) or (#30 in AB)*
 32. *#25 or #26 or #27 or #28 or #29 or #31*
 33. *#6 or #15 or #20 or #24 or #32*
 34. *(TG=animal) not ((TG=human) and (TG=animal))*
 35. *#33 not #34*
 36. *"qualitative"/ all subheadings*
 37. *qualitative* in TI*
 38. *qualitative* in AB*
 39. *#36 or #37 or #38*
 40. *"descriptive"/ all subheadings*
 41. *descriptive* in TI*
-

42. *descriptive** in AB
43. #40 or #41 or #42
44. "observational"/ all subheadings
45. *observation** in TI
46. *observation** in AB
47. #40 or #41 or #42

Meta-analyses

1. "meta-analysis" in MIME,MJME
2. *meta-analysis* in PT
3. #1 or #2
4. (*meta anal** or *metaanal**) in TI,AB
5. (*integrativ** *research review** or *research integration*) in TI,AB
6. (*quantitativ* synthes**) in TI,AB
7. (*pooling** or (*pooled analys**) or (*mantel* haenszel**)) in TI,AB
8. (*peto** or *der simonian** or *dersimonian** or *fixed effect** or *random effect**) in TI,AB
9. #4 or #5 or #6 or #7 or #8
10. #3 or #9
11. (TG=animal) not ((TG=human) and (TG=animal))
12. #10 not #11

Additional key words of relevance were not identified during any of the electronic or other searches.

Publication selection

All publications identified in the searches were imported into a bibliographic database (Reference Manager Version 10). Codebooks were created separately for the peer reviewed literature and gray literature searches. Each title and abstract was coded separately in both the bibliographic database and the codebook. The following exclusion criteria were applied to the titles abstracts:

1. The publication was an individual case report
2. The publication was an editorial, letter or opinion piece
3. The publication did not deal with peer review of health care professionals
4. The publication was a duplicate publication
5. The publication was published in a language other than English

Publications were also excluded if the data had since been updated in a subsequent publication.

To determine the publications to be assessed further, two independent reviewers scanned the titles, abstract sections and keywords of records retrieved. Full publications were retrieved for further assessment if the information given suggests that the publication:

1. Included subjects who were health care professionals;
2. Described a peer review method; and
3. Provided information regarding the purpose of the peer review activity.

Abstracts that did not meet inclusion criteria were coded according to reason for rejection. If there was any doubt regarding the details of the publication from the information given in the title and abstract, the full publication was retrieved for clarification. Inter-rater agreement for publication selection was measured using the kappa statistic (Cohen 1960). It was intended that a third party would resolve differences in opinion and that if resolving disagreement was not possible, the authors would be contacted for clarification. This was not required.

Quality assessment of studies

Current guidelines emphasise the need for a structured approach to quality assessment for qualitative studies to be included in reviews (BMJ 2007). Papers were not excluded for reasons of quality, however quality assessment was undertaken to enable publications of poorer quality to be identified and accounted for in data synthesis (Jensen and Allen, 1996).

The quality of reporting of each publication was assessed by quality criteria based on:

1. Type of evidence: Evidence was graded according to three categories of quality – meta-analyses or randomized controlled trials, non-randomised or observational studies, and evidence based on qualitative data or expert opinion.
 2. Directness: Evidence was graded according to the following additional factors – lower rating for studies that were of limited generalisability, because the included population was too narrow or too broad; and for outcomes that were difficult to generalise, such as those only reported as composite outcomes or that were poorly defined.
-

(Guyatt et al. 2008).

Based on these criteria, studies were broadly subdivided into the following three categories:

A - Quality criteria met: low risk of bias.

B - One or more of the quality criteria only partly met: moderate risk of bias.

C - One or more criteria not met: high risk of bias.

Both reviewers independently assigned a quality score to each publication. Differences of opinion were resolved through discussion. Consensus was reached on all studies.

Data extraction

Data that were extracted included the following:

1. General information: peer review publication / non-peer review publication, title, authors, contact address, country, year of publication.
2. Study characteristics: study type, duration, if applicable - randomisation, allocation concealment (and method), blinding (subjects, people administering treatment and outcome assessors).
3. Intervention: details of peer review method, details of comparison interventions if applicable (method, timing).
4. Participants: sampling (random / convenience), exclusion criteria, total number and number in comparison groups, gender, age, profession type, assessment of compliance / relapse, withdrawals / losses to follow-up (reasons / description), subgroups.
5. Results: for outcomes and times of assessment.

A template data extraction form was developed. Two reviewers performed data extraction independently. Differences in data extraction were resolved by consensus.

Data analysis

Data from publications were summarized narratively, by chronicling and ordering the evidence to produce an account of the evidence. This enabled integration of quantitative and qualitative evidence (Guyatt et al. 2008).

Feedback on the narrative and its interpretation was then sought from personnel of the Australian Commission on Safety and Quality in Healthcare. Based on feedback received, both reviewers revised and refined the salient points in the narrative jointly.

Where objective outcomes data were provided, data were extracted and reported for all relevant outcomes.

RESULTS

Studies Identified

The search strategy was performed in February 2009 and identified the following numbers of publications at each stage:

Potentially relevant abstracts identified and screened for retrieval in peer-reviewed literature (n= 1253)

- Abstracts excluded (n=1167)

Potentially relevant publications identified in gray literature search and screened for retrieval (n=743)

- Publications excluded (n=695)

Total publications retrieved for more detailed evaluation (n=134)

- Publications excluded (n=45)
- Potentially appropriate publications to be included in the systematic review (n=89)
- Publications excluded from the systematic review as duplicate publications (n=4)
- Publications included in systematic review, with usable information (n=85)

A total of 85 publications met the inclusion criteria and were included in the review. The kappa statistic for publication selection was 0.77 (95% CI: 0.65 to 0.89). The publications were all published between 1998 and 2008.

Study Types

One included study was a randomized controlled trial, 43 were evaluation studies of a single peer review method, 14 were evaluation studies comparing peer review methods, 10 were descriptive studies of a single peer review method, 6 were qualitative studies and one was a mixed qualitative/quantitative study.

Participants and Settings

Publications related to one or more health care provider type. A total of 75 publications related to peer review activities in medical practitioners. A number of publications reported peer review activities across multiple medical practitioner types.

Where specified or specifically appraised, 12 publications related to GPs, 8 to emergency department medical practitioners, 7 to surgeons, 6 to paediatricians, 5 to radiologists and obstetricians / gynaecologists respectively, 3 to psychiatrists, anaesthetists, pathologists and resident doctors respectively and 1 to ophthalmologists.

There were 18 publications related to peer review activities in nurses, 3 to allied health practitioners and dentists respectively, and 1 to pharmacists.

A total of 33 publications related to peer review in the US, 24 in the UK, 11 in Canada, 5 in The Netherlands and Australia respectively, 2 in Germany and 1 each in India, Finland, Pakistan, Spain, Japan and Zambia (one study was conducted in both Australia and the UK).

Participants were practicing in in-patient settings in 32 publications, in outpatient settings in 28 publications, and across both settings in 24 publications. The range of outpatient settings in which peer review activities were conducted included general practices, general medical clinics, hospital outpatient clinics, ambulatory care settings, and other private practice settings.

Most publications described peer review of an existing workforce within their institution. Where participants were selected, this usually involved obtaining a convenience sample of participants. A number of publications described recruiting their participants from databases of employees working within a facility or area health service, or of health care professional registration databases.

Peer Review Information Sources

Information sources were frequently used in peer review. Some peer review processes used multiple information sources.

- 38 publications described the use of chart audit as a source of data for the peer review process;
 - 24 publications described the use of peer assessment of routinely collected information sources;
 - 11 used routinely collected incident reports, reports of significant event analyses or mortality reports (e.g. maternal mortality report)
 - 7 used registry data sources (e.g. routinely collected trauma registry data)
 - 6 used routine survey-based audit information (e.g. routinely collected information regarding indications for some types of operations)
 - 19 publications described the use of questionnaires provided to key informants as a source of data for the peer review process;
-

- 12 of these were based on multi-source feedback (360 degree evaluation) questionnaire methods
- 13 publications described the use of information collected at interview with the health care professional undergoing peer review;
- 12 publications described the use of information collected during a site visit at the health care professional's place of work;
- 9 publications described the use of information contained within ICD coded datasets (e.g. hospitalization statistics, coded complication rates);
- 3 publications described the use of autopsy data; and
- 2 publications described requiring the health care professional to develop a portfolio or a resume for the peer review process.

Number of Peer Reviewers

The number of peer reviewers involved in the peer review processes described in publications is provided in Appendix 1.

Most publications described the use of multiple peer reviewers to undertake peer review tasks. The commonest number of peer reviewers used was between 2 and 3 health care professionals.

In some studies, a single peer reviewer performed peer review tasks. However, where there were issues identified by the peer reviewer, a secondary peer review process was often instituted involving multiple peer reviewers.

In other studies, a single peer reviewer performed the entire peer review task.

Methodological Quality

The methodological quality of included publications is described in Appendix 1.

One study was categorised as 'A', indicating that quality criteria were met (van Eijk et al. 2001). A total of 49 publications were categorised as 'B', indicating that one or more criteria were partly met, and 35 publications were categorized as 'C' indicating one or more criteria were not met.

Categorical analysis

Descriptive and interpretive categories were identified through independent coding of the results of included publications, determined according to the primary rationale for conducting the peer review process. The major categories that emerged were aligned with the purposes of the peer review activity and were:

1. Assessment of the clinician's domains of professional practice
-



2. Assessment of delivery of care in accordance with clinical guidelines
3. Assessment of organizational quality of care
4. Peer review as a requirement for Continuing Professional Development
5. Assessment of significant events
6. Quality assurance of radiology / pathology practice
7. Peer review for the purposes of credentialing health care providers; and
8. Assessment of suspected under-performance of the health care provider.

These are described below.

A number of publications described more than one rationale for conducting the peer review process. Where the primary rationale was multifold, the publication was coded into multiple categories.

1. Assessment of the clinician's domains of professional practice

There were 24 publications that described the use of peer review methods with the intent of appraising one or more specific domains of professional practice of the health care provider.

There were 5 publications describing the use of peer review to assess a single domain of professional practice. Of these, 1 publication described the use of a survey of peers, 2 described chart audit by peers for the assessment, and 2 described peer review group processes for the assessment of the health professional (Table 1).

There were 19 publications describing the use of peer review for the assessment of multiple domains of professional practice. Of these, 12 publications described the use of a survey of peers for peer review, 1 described chart audit by peers for the assessment, 1 described a peer review group discussion process for the assessment of the health professional, 1 described the use of a site visit by peers to review practice, and 4 used a combination of methods in the peer review process (Table 1).

The dimensions of professional practice assessed across publications were:

- Clinical skills and knowledge (19 publications).
 - Professionalism (11 publications), including:
 - Teamwork;
 - Relationships with colleagues; and
 - Relationships with patients.
-

- Communication (10 publications) including;
 - Patient education; and
 - Record keeping.
- Management (9 publications), including:
 - Use of diagnostic testing;
 - Resource utilization; and
 - Administrative skills.
- Scholarship (6 publications), including:
 - Performance in teaching duties.

Surveys of peers were primarily for the purposes of providing feedback to the individual regarding their performance across domains of practice appraised. However, three of the survey methods described (Medical Council of Ireland 2008; Thomas et al. 1999; Archer et al. 2005) and the site visit method described by Moss et al. (2005) were for the purposes of assessing the competence of the health care provider, and the results of the assessment were used by organisations or professional bodies to inform decisions regarding ongoing registration / accreditation of the provider.

There was some evidence that peer review resulted in improved professional practice in the domains of clinical skills and knowledge, management, professionalism and communication (Hyrkas et al. 2003; Norton et al. 1998; Paige et al. 2008). There was also evidence that clinicians who participate in peer reviewing their colleagues experienced improvement in their own professional practice (Paukert et al. 2003).

The reliability of survey instruments used for assessment of professional practice was appraised in a number of publications (Table 1). Reliability was found to increase with increasing numbers of peer reviewers (Archer et al. 2005; Hall et al. 1999; Lelliott et al. 2008), and with the use of a systematic approach to assessment, using clinical practice guidelines as a point against which assessment was conducted (Campbell et al. 1999; Soroka et al. 2004).

Four publications applied multiple peer review methods to the assessment of professional practice of health care professionals. Results demonstrate that information obtained by chart audit was improved by the addition of a structured interview with the health care professional (Goulet et al. 2007).

2. Assessment of delivery of care in accordance with clinical guidelines

There were 11 publications that described the use of peer review methods to assess the professional care delivered by the health care professional against clinical practice guidelines.

Six publications described peer review of patient medical records to conduct the assessment, 2 used peer review of pathology / radiology requests made by medical practitioners, 2 performed face to face assessments of the health care professional and 1 used a questionnaire completed by the subject of the peer review to inform the peer review process (Table 2).

Chart audit was facilitated by the use of structured assessment instruments by peer reviewers in 5 of the 6 publications describing this method (Table 2). Structured assessment instruments were developed for the purposes of peer review, were derived from clinical practice guidelines and were completed by peer reviewers when reviewing patient charts. In the 6th publication, implicit review (the professional opinion of the peer reviewer) was used to assess the health care provider (Isetts et al. 2003). In this publication, a panel of 12 health care professionals assessed the professional's clinical practice. The outcomes being assessed were well defined (drug-related morbidity and achievement of pharmacotherapeutic goals). Intra-class correlation coefficients of agreement between panel members ranged from 0.73 to 0.85.

Peer review of health care professionals pathology ordering demonstrated substantial variation in the assessment by peers of the appropriateness of requested pathology tests (kappa 0.33 to 0.42). A large number of peer reviewers (approximately 16) were required to achieve sufficient reliability to make peer review of pathology ordering practices a reliable assessment of clinical practice (Bindels et al. 2003). In the other publication of peer review of clinician's pathology ordering practices, a panel of 17 peers conducted peer review of test requests. The findings of this peer review process were successfully used to reduce inappropriate ordering of pathology and radiology tests (Neilson et al. 2004).

Provision of face-to-face feedback to health professionals by peers who reviewed specific aspects of clinical practice was described in hospital and outpatient settings. This method was demonstrated to improve the compliance of clinicians with clinical practice guidelines for medication prescribing in both publications (Sucov et al. 2005; Van Eijk et al. 2001).

The reliability of peer review in assessing clinical practice against clinical guidelines was improved by:

- The use of structured assessment tools (Campbell et al. 1999; Johnstone et al. 1999);
 - The addition of multiple reviewers (Bindels et al. 2003; Hofer et al. 2004); and
-

- When the condition being assessed was one where the evidence base is well-developed (Hofer et al. 2004).

3. Assessment of organizational quality of care

There were 11 publications that described the use of peer review methods to assess the quality of care provided by an organization or by a group of practitioners for an organisation (Table 3).

Nine publications described peer review by multiple methods, including site visits to the health care facility to conduct the assessment, inspection of medical records and interviews with staff. Two publications described the use of peer review of patient medical records to perform the assessment.

Three of the publications regarding site visits described assessment by 'visitatie' – an external peer review process performed in The Netherlands by teams of peers who visit health professionals premises, inspect patient records and interview staff (Hofhuis et al. 2006; Lombarts and Klatzinga 2001; Lombarts and Klatzinga 2003). According to authors, this process increased the awareness of participating health care professionals of ways to improve practice and in 33% of practices surveyed, resulted in actual improvements in professional practice (Hofhuis et al. 2006). The improvement in patient care due to visitatie was improved by also providing practices with management consultancy visits, which assisted practices to implement the recommendations of visitatie (Lombarts and Klatzinga 2003). In the other two publications of peer review by site visits, peer review was subjectively rated as a useful activity but impacts on professional practice were not assessed (Crean et al. 2003; King et al. 2004).

Three publications described the American College of Obstetrics and Gynaecology (ACOG) Voluntary Review of Quality of Care (VRQC) program. In this program, hospitals may voluntarily request the ACOG conduct an external assessment of the facility. This may be for the purposes of evaluating the hospital's performance in obstetrics and gynaecology against accepted standards of care. Site visits, staff interviews, review of clinical records and review of administrative data by a team of peer reviewers inform the assessment. The hospital then receives a written report outlining findings (Gluck and Scarrow 2003; Stumpf 2007; Lichtmacher 2008). Results suggest that organizational systems are more common triggers for review than concerns with individual medical practitioners (Gluck and Scarrow 2003). Specific concerns identified by peer reviewers include problems associated with organizational quality assurance processes, and issues associated with staffing levels (Lichtmacher 2008).

Peer review of organizational quality of care by chart audit was described for surgery, anaesthetics, and medicine (including geriatrics). In both publications, peer reviewers used structured assessment instruments to facilitate the peer review process. Reviews were resource intensive and time-consuming (Pacala et al. 2000; Walshe et al. 2001). Inter-rater reliability was poor in one publication (Pacala et al. 2000). In the other publication, quality of

care was judged to be more deficient by reviewers when adverse events occurred, or when documentation was inadequate (Weingart et al. 2001).

4. Peer review as a requirement for Continuing Professional Development

There were 3 publications that described peer review as a requirement of the formal Continuing Professional Development (CPD) program of the health care professional. Publications were heterogeneous in the purposes of the peer review activities they described and there were few similarities between publications (Table 4).

Bowie et al. (2008) described the peer review of Significant Event Analysis (SEA) reports and clinical audit reports prepared by GPs as part of their ongoing CPD requirements. Assessment of the reports was facilitated by the use of structured review instruments. According to the publication findings, novice peer reviewers (who had received minimal training and had no prior experience in assessment of the reports) were able to reliably provide peer feedback to GPs on the reports.

Greenwood and Williams (2008) described a peer review process to inform the learning needs of professionally isolated rural psychiatrists. Peer review groups for education and CPD were perceived as useful for increasing learning opportunities. Maidment (2006) described the views of dentists toward the use of peer review methods for CPD. Peer review was perceived as acceptable and effective for changing professional knowledge and practice. However, neither publication provided objective measures of effectiveness to support these views.

5. Assessment of significant events¹

There were 25 publications that described the use of peer review methods for the assessment of significant events (Table 5).

Methods used for peer assessment included peer review of written reports of significant events, review of medical records of patients affected by significant events, the use of peer review panels to assess significant events, and formal audit programs established to routinely review potentially significant events when they occurred.

Peer review of written reports of significant events was the subject of 6 publications. Four of these related to peer review of significant events reported by GPs, and two of significant events occurring in hospital settings. The peer review process was facilitated by the use of structured assessment tools in 4 publications. Peer review of significant event reports provided useful feedback regarding contributing factors to the occurrence of incidents (Nuckols et al. 2008). However, where participation in the process was voluntary, participants were highly selective about the significant event reports they submitted for peer

¹ Significant events are defined as untoward sequelae of the therapeutic relationship, including adverse events and complaints received regarding the standard of care provided.

review, due to concerns about confidentiality, litigation and professional embarrassment (Bowie et al. 2005).

The reliability of peer review of written reports of significant events improved with increasing numbers of peer reviewers (Forster et al. 2007). Where unstructured assessment was used, inter-rater reliability varied widely (Forster et al. 2007; Nuckols et al. 2008). Peer review using structured assessment tools achieved moderate inter-rater reliability (McKay et al. 2007). However, evidence suggested that participants may require training in how to write a written report of sufficient quality to enable peer review of the significant event to occur (McKay et al. 2006).

Peer review panels were described for the peer review of significant events in 5 publications. Three described peer review of maternal and perinatal adverse events, one described a peer review panel assessment of trauma deaths (Takanayagi et al. 1998) and one described a peer review panel for investigation of adverse events and the individual nursing staff member's contribution to the event (Diaz 2008). All peer review panels involved the assessment of the significant event by multiple peer reviewers. Four described a structured, formal peer review process and one described an informal peer review panel (Takanayagi et al. 1998). Participation in panels as a peer reviewer was demonstrated to be educational and was a useful process for identifying factors associated with substandard delivery of clinical care (Rankin et al. 2006; Stekelenburg and Van Roosmalen 2002). Peer assessment by panel members was demonstrated to generate higher levels of agreement between members when assessment was made against pre-defined practice standards (Kernaghan and Penney 2006).

There were 5 publications which described the peer review of trauma outcomes by peer review groups. Like peer review panels, the groups were comprised of health care professionals who met to discuss significant trauma events in order to identify and respond to contributing factors. All peer review groups involved the assessment of the trauma outcome by multiple peer reviewers. Four of the peer review groups routinely assessed all trauma deaths that had occurred. One group assessed adverse trauma outcomes more broadly, including the quality of care received by patients who experienced adverse outcomes (Jacobs et al. 2006).

Peer review groups were able to identify factors associated with preventable death, including inadequate pre-hospital care, inappropriate inter-hospital transfer, limited hospital resources, and factors associated with integration of trauma care (Forsythe et al. 2002; Jat et al. 2004; Martin et al. 2007; Shanti et al. 2003). Trauma outcome judgments collected anonymously by peer reviewers were significantly less favourable than those obtained non-anonymously (Jacobs et al. 2006). Peer review methods were more sensitive in identifying preventability of deaths than the use of registry data-derived trauma severity scoring (Shanti et al. 2003). Data regarding the value of autopsy data in enhancing the peer review process for identifying preventable factors contributing to death in trauma patients were mixed (Forsythe et al. 2002; Martin et al. 2007).

Peer review of the medical charts of patients who had experienced significant events was the subject of 2 publications. A structured assessment tool was used to facilitate the peer review process and multiple peer reviewers conducted the assessment of the significant event in both publications. Peer review resulted in the identification of contributing factors to adverse events in both studies. Nurse ratings of adverse events were demonstrated to be consistent with physician ratings (Silver et al. 2007). Reliability of the assessment improved with increased numbers of reviewers (Hofer et al. 2000).

Peer review audit of significant events associated with surgery was the subject of 5 publications. Thompson et al. (2005) described the Scottish Audit of Surgical Mortality and Semmens et al. (2005) described the West Australian Audit of Surgical Mortality. Both are structured audit programs of deaths associated with surgery, designed to provide feedback to participating surgeons regarding factors associated with surgical deaths. Changes in professional practice by surgeons as a result of participation in audit activities were described, including improved prophylaxis for deep vein thrombosis, and improved use of intensive care unit and high dependency unit treatment for post-surgical care (Allen and DeSimone 2002; Semmens et al. 2005; Thompson et al. 2005). Reduced patient mortality was associated with surgical audit activities in two publications (Olcott et al. 2000; Thompson et al. 2005).

6. Quality assurance of radiology / pathology practice

There were 3 publications that described peer review for the purposes of radiology quality assurance and 2 that described peer review for pathology quality assurance (Table 6).

Radiology quality assurance was conducted by peer review of images previously reported by another radiologist, and their corresponding reports. In two publications, multiple radiologists performed peer review of their colleagues (Britton et al. 2001; Jolly et al. 2001) and in one study a single peer reviewer performed the task initially, but then referred the film and report to a committee of radiologists for further assessment if there was disagreement with the first radiologist's findings, and the diagnosis was one which the peer reviewer felt should be made most or all of the time (Borgstede et al. 2004).

Results demonstrated that large numbers of reports needed to be reviewed in order to identify instances of poor practice (Jolly et al. 2001; Borgstede et al. 2004). Peer reviewer consistency of rating their peers was also problematic, with peer reviewers rating certain reports unpredictably independent of training (Britton et al. 2001; Jolly et al. 2001; Borgstede et al. 2004).

Pathology quality assurance was conducted by peer review of pathology specimens previously reported by another pathologist, and their corresponding reports. Multiple peer reviewers performed the assessment in both publications (Bhatia et al. 1998; Zardawi et al. 1998).

Results demonstrated that a large number of specimens and reports needed to be peer reviewed in order to find diagnostic errors and discrepancies. In addition, significant resources were required to undertake peer review of the large volume of cases, and

worsened the turnaround time for pathology reporting (Bhatia et al. 1998). However, the method was successful in identifying errors in both procedural and technical pathology practice.

7. Peer review for the purposes of credentialing health care providers

Peer review for the purposes of credentialing health care providers was the subject of 3 publications (Table 7). Two publications described peer review in medical practitioners, and one in dentists. In two publications, a single peer reviewer performed the initial assessment, which was then referred to a committee for further assessment routinely in one study (Johnstone et al. 1999), and was referred if there was disagreement between peer reviewer and practitioner in the second study (Borgstede et al. 2004). Multiple peer reviewers routinely assessed all materials in the third publication (Maidment et al. 2006).

In the first publication, peer radiologists reviewed images and radiology reports previously reported by another radiologist. The findings of peer reviews were entered onto a database that was interrogated when decisions regarding the re-credentialing of the radiologist were to be made (Borgstede et al. 2004). In a second publication, peer radiation oncologists reviewed each other's patient charts and accompanying films to confirm diagnostic accuracy, adequacy of the doctor's documentation, and adequacy of clinical care. The participant received feedback from the process monthly, which was also provided to the employing organization for use in the credentialing process (Johnstone et al. 1999). The third publication described the assessment of dental practitioners by completion of a portfolio of evidence supporting fitness to practice. This portfolio was assessed by three peers using a structured assessment tool. The results of the assessment were used to inform decisions regarding re-validation of the practitioner (Maidment et al. 2006).

There was no information provided in publications reviewed regarding the acceptability of methods for re-credentialing, comparison of methods with other forms of assessment, or measures of reliability and / or validity of the assessment methods used.

8. Assessment of suspected under-performance² of the health care provider

There were 10 publications that described the assessment of suspected under-performance of health care professionals (Table 8). There were 7 publications that described the use of multiple peer review methods by multiple peer reviewers in their assessment of under-performance, 2 that described the use of chart audit by multiple peer reviewers (Liang 1999; Billett et al. 2005) and 1 that was an audit of the procedural elements of peer review committees in a geographical area (Spigelman and Swan 2003). There was insufficient information provided in the last publication to determine whether peer review committees were also using multiple peer review methods to undertake their assessments as this was not the purpose of the audit.

² Under-performance is defined as clinical practice by a health care provider that does not meet one or more defined minimum standards of clinical practice (Southgate et al. 2001)

All of the 7 publications where multiple peer review methods occurred describe the use of a peer review committee in managing the peer review process. These committees were described as formal structures, often with elected members who fulfilled the role for a specific timeframe. Committees were developed as part of an organization's governance responsibility, and had defined reporting relationships between the committee and one or more organizations.

Chart audit, interviews with relevant staff members, appraisal of administrative datasets (including patient and / or staff complaints, morbidity and mortality data) and inspection of premises were common elements of peer review processes described. The outcomes of peer review described in publications included further education of the staff member (including re-training) (Agee 2007; Lichtmacher 2008), performance management (including supervision requirements) (Southgate et al. 2001; Agee 2007), conditional registration (Southgate et al. 2001) or revocation of rights to practice (Southgate et al. 2001).

In 2 studies, the peer review processes described were part of the UK General Medical Council procedures for the assessment of suspected under-performance of doctors. The governance arrangements associated with these peer review processes were described by Southgate et al. 2001, and were referred to in background information within Hutchinson et al. 2001. The peer review process was developed as part of the General Medical Council's governance responsibilities. In this process peer reviewers produce a report in a standardized format that is referred to a statutory committee of the General Medical Council. The results of peer review, supported by tests of competence, inform decisions of the committee regarding professional registration. Reporting relationships between assessors, the Committee and relevant professional bodies are formal and defined.

In 3 studies, the peer review processes described were part of the American College of Obstetrics and Gynaecology (ACOG) Voluntary Review of Quality of Care (VRQC) program (Gluck et al. 2003; Stumpf 2007; Lichtmacher 2008). Participation in this program by an organization is voluntary. A health care facility makes a formal request to ACOG to perform a VRQC review. The review may be triggered by problems with a specific physician. A team of peer reviewers, trained by ACOG, assess background information, interview relevant hospital staff, inspect the facility, independently audit clinical records using a structured assessment tool, and prepare a written report outlining findings. The ACOG committee review the findings of the assessment process before making recommendations to the requesting hospital. Problems commonly identified by this process include communication issues and poor documentation of clinical reasoning.

Problems identified with the peer review processes described for under-performance were that they can be resource-intensive and time-consuming (Billett et al.2005). Barriers identified to resolving poor performance that may exist within an organization include the unwillingness of some doctors to seek advice, and a protective culture where complaints are not made against doctors by other staff (Hutchinson et al. 2001).

DISCUSSION

This systematic review provides an account of the literature regarding the peer review of health care professionals. The publications identified through the review methods were heterogeneous and described the application of peer review of health care professionals for a variety of different purposes, and using a range of peer review methods.

Application of peer review to performance assessment³

Medical practitioners were the most frequently described health professional type to be the subject of peer review, although the peer review of allied health and nursing professionals was also described. Allied health and nursing peer review processes were generally consistent with those of medical practitioners, with similar purposes for the peer review process and similar methods of peer review used. The frequency of assessment of medical practitioners may be higher as the assessment and maintenance of medical practitioner competence has been given worldwide attention (Southgate et al. 2001b), partly in response to concerns about medical practitioner performance and patient safety (Southgate et al. 2001), as well as increasing demands for accountability to patients and funding agencies (Kohn et al. 1999).

Publications generally described the application of peer review for performance assessment of the health care professional (or group of professionals).

Two broad types of peer review for performance assessment emerged:

- Assessment to ensure providers met minimum standards of practice (i.e. to ensure competence of the provider); and
- Assessment for the purposes of clinical practice improvement.

Assessment against minimum standards of practice was more commonly summative in nature whereas formative, summative, or a mixture of both assessment types were described for assessment conducted for the purposes of clinical practice improvement.

Assessment to ensure providers met minimum standards of practice was more commonly a formal peer review activity, whereas both formal and informal assessments were described for the purposes of clinical practice improvement.

Formal peer review activities all involved peers systematically reviewing aspects of a health care professional's work and normally included documented, structured assessment processes. They also generally included formal linkages to an organization or professional

³ Performance assessment can be defined as the appraisal of the extent to which a health care professional provides health care services consistent with known good practice and resulting in expected patient benefit (Wass et al. 2003).

association, with the findings of the peer review process frequently used to enable that organization or professional association to ensure the ongoing professional performance and / or competence of the health care professional.

Effectiveness of peer review

Publications demonstrated that peer review of health care professionals can be effective across the four domains of participant satisfaction, learning outcomes, performance improvement and patient outcomes (as described in the Introduction section).

Participant satisfaction was the subject of numerous publications (Wendling and Hoekstra 2002; Crean et al. 2003; Bowie et al. 2005; Kelly 2005; Moss et al. 2005; Maidment 2006; Rankin et al. 2006; Greenwood and Williams 2008). Peer review team members, as well as the professionals who were the subject of peer review, reported that the peer review process provided valuable and useful feedback to the health care professional. Participation in the peer review of other health care professionals also provided participants with a learning experience that provoked reflection on their own clinical practice (Rankin et al. 2006).

Learning outcomes were specifically assessed in two publications (Maidment et al. 2006; Greenwood and Williams 2008). According to results, peer review contributed to participants formulating learning objectives to inform their continuing professional development needs. Participation in the peer review process as a peer reviewer had the additional benefit of increasing the awareness of the reviewer of evidence-based guidelines and standards of care (Rankin et al. 2006). Further, writing a serious event analysis report was educational for the participant and provided opportunities to reflect on practice and learn from errors (Bowie et al. 2005).

Publications where performance improvement was assessed found that participants of peer review processes intended to make improvements to their practice. However, feedback and recommendations arising from peer review processes did not always result in improvements occurring, or resulted in improvements being made in the minority of areas identified where improvement was desirable (Norton et al. 1998; Lombarts and Klazinga 2003; Semmens et al. 2005; Hofhuis et al. 2006; Medical Council of Ireland 2008). However, where professionals were supported to implement the recommendations of peer review processes, and where audit processes were implemented to monitor compliance with recommendations, reported performance improvement was higher (Lombarts and Klazinga 2003; Semmens et al. 2005).

Peer review was associated with measurable improvement in performance across a range of settings and with a range of peer review methods. For example:

- Peer review of radiology reporting resulted in measurable improvement in quality of reporting; peer review of radiology and pathology test requests resulted in reduction in unnecessary radiology and pathology testing (Zardawi et al. 1998; Britton et al. 2001; Neilson et al. 2004).
-

- Peer review educational by visits to GPs and pharmacists resulted in altered prescribing behaviour (van Eijk et al. 2001).

There was some evidence demonstrating that peer review activities can improve patient outcomes. In one publication meeting inclusion criteria, a carotid quality assurance peer review team provided feedback to participating surgeons regarding morbidity and mortality for their patients, compared with rates for their institutions. Mortality rate, stroke rate, length of stay, and cost decreased during the 5-year period of the study. This was hypothesized to be as a result of the intervention (Olcott et al. 2000). In another publication, deaths after elective surgery were associated with the commencement of a surgical audit involving peer review (Thompson et al. 2005). However, in a third publication meeting inclusion criteria, surgeons were asked to rate the appropriateness of surgery for abdominal aortic aneurysm. They were unable to consistently identify candidates for whom surgery was appropriate. Consequently, panel decisions correlated poorly with 5 year mortality outcomes of patients (Silverstein and Ballard 1998). More evidence is required before definitive statements regarding the relationship between peer review activities and improved patient outcomes can be made.

Reliability and Validity of Peer Review Methods

The reliability and validity of the peer review methods used by reviewers are important for ensuring that peer review is both accurate and meaningful. The reliability of the peer review method is the degree to which one can depend on the accuracy of the method's results. For example, whether different peer reviewers would award the same results given the same participant's performance (inter-rater reliability) or whether the same participant would score the same result if the assessment were repeated under the same conditions. Reliability is assessed using a range of metrics, including inter-rater reliability, intra-rater reliability and / or generalisability.

The validity of the peer review method concerns the degree to which the peer review method assesses what it is intended to assess. For example, whether the assessment covers the range of parameters it is intended to cover. Validity assessment involves a number of constructs, including content validity, construct validity and criterion validity. There was little published information identified in this review demonstrating the validity of peer review methods. Further research is required to address this evidence gap.

A number of publications described the results of assessments of the reliability and / or validity of varying peer review methods. Although heterogeneity prevents direct comparison of reliability and validity outcome measures between publications per se, several patterns emerged from comparisons between publications.

- For survey-based assessment tools, intra-class correlation coefficients of reliability varied widely (ICC 0.1 to 0.85). Inter-rater reliability (kappa) values also varied widely (between 0.14 and 0.95), depending on the instrument type and number of comparisons being made. A series of publications by Lockyer et al. assessed multi-source feedback survey instrument reliability, including generalisability. With the instrument used, internal reliability was high (Cronbach's alpha >0.95) and generalisability was good
-

(generalisability coefficients between 0.56 and 0.85) (Lockyer and Violato 2004; Lockyer et al. 2005; Lockyer et al. 2006; Violato et al. 2006; Lockyer et al. 2008).

- Reliability and / or validity of panel-based assessments were also assessed. For panel-based assessment, levels of agreement between panels (kappa values) ranged between 0.1 and 0.85, depending on the panel composition, type of peer review being undertaken, and number of comparisons being made (Isetts et al. 2003; Kernaghan and Penney 2006). Panel processes were generally demonstrated to have sufficiently high test reliability to support their application in peer review assessment processes, particularly in the assessment of professional competence and suspected under-performance (Cronbach alpha 0.61 to 0.88) (Southgate et al. 2001).
- Reliability of individual assessments demonstrated that inter-rater reliability was low (kappa 0.33 to 0.42; Bindels et al. 2003). Reliability of single physician peer review, aided by structured implicit review instruments developed to guide the peer review assessment, was also low (ICC 0.16 to 0.26; Hofer et al. 2004).

No publication assessed all aspects of validity. Content validity of two instruments was assessed by appraisal of peer review survey instruments' scoring lists by experts (Hall et al. 1999; Archer et al. 2005). Construct validity was assessed in two publications by assessment of discriminative ability of the instrument using factor analysis (Lockyer and Violato 2004; Archer et al. 2005).

These data suggest that peer review methods that are poorly designed may not provide consistently reliable, accurate information regarding the performance of the health care professional. Careful attention therefore needs to be paid to the design of the peer review process.

Features of the peer review process

There were a number of features of the peer review process that were associated with review that is more rigorous. These included increasing the number of reviewers, basing peer review on evidence-based guidelines, using structured assessment methods, using multiple peer review methods to perform an assessment, and by training reviewers.

1. Increasing the number of reviewers

Peer review can be undertaken by an individual peer or by multiple peers. Depending on the purpose of the peer review activity, multiple peer reviewers may be preferable to reliance on an individual for peer review of one's performance.

The number of peer reviewers participating in the peer review process varied widely across publications. This was because the appropriate number of reviewers depended on the purpose of the peer review activity, the peer review method being applied to the assessment of the health care provider, the properties of the peer review instrument (where one was used), and the domains of clinical performance being assessed.

The sensitivity and specificity of peer review when compared with objective measures was assessed by Takanayagi et al. 1998 and by Forster et al. 2007. Findings demonstrated that the positive predictive value of the peer review also increased with increased numbers of reviewers.

It is therefore not possible to nominate an optimal number of peer reviewers required for the peer review of a health care provider. However, increasing the number of reviewers generally increases the reliability of peer review processes (Hall et al. 1999; Bindels et al. 2003), particularly where the peer review method is a survey (Hall et al. 1999; Thomas et al. 1999; Campbell et al. 2008; Lelliott et al. 2008).

2. Basing peer review assessment on evidence-based guidelines

The reliability of peer review improves when the clinical conditions upon which the review is based are those where the evidence base is well developed (Campbell 1999; Hofer et al. 2004).

When the subject of the peer review process is professional practice for which there is debate regarding what constitutes appropriate practice, or for less common conditions, peer appraisal is more challenging. Further, professional disagreement about the evidence base and its application to the clinical circumstances relevant to the peer review reduces the inter-rater reliability of the peer review process (Silverstein and Ballard 1998).

3. Using structured assessment methods

The reliability of the peer review process was greater with the use of structured assessment methods compared with unstructured processes. The addition of structured assessment tools in particular, such as surveys and checklists, increased the reliability of peer review between assessors in some publications (Margo 2002; Evans et al. 2004; McKay et al. 2007). However, if the structured assessment tool was poorly designed and unreliable it did not improve reliability (Pacala et al. 2000; Hofer et al. 2004; Soroka et al. 2004).

Many peer review activities involved the use of structured assessment tools by reviewers. Structured processes may enable reviewers to be more closely guided in the review process, and enable the assessment of specific domains of competence to be more targeted. It is therefore important that structured assessment tools applied to the peer review of health care professionals are well designed and suited to the purpose of the peer review activity.

4. Using multiple peer review methods to perform an assessment

A number of publications described the use of multiple peer review methods. The use of multiple methods increased the sensitivity of the peer review process in identifying issues relating to the professional practice of the subjects of the review in a number of publications (Takanayagi et al. 1998; Hofer et al. 2000; Shanti et al. 2003; Kernaghan and Penney 2006).

The use of multiple peer review methods was common in formal peer review activities, and where the implications of the peer review process on the individual were greater (for example, in the assessment of under-performance, where the professional's registration status may be influenced by the findings of the assessment).

Multiple peer review methods provide a more holistic assessment of the health professional. Peer review of an individual component of performance may be undertaken using a single peer review method. However, judgments about multiple domains of clinical performance are improved through the use of multiple peer review methods of assessment. Therefore, where more holistic assessment of performance is desired, multiple peer review methods may be preferred.

5. Training the peer reviewers

A number of publications described training for peer reviewers. In some clinical systems, training is provided to peer reviewers to improve the reliability of peer review processes. For example, in the UK General Medical Council's performance procedures, potential clinical assessors are short-listed and interviewed against specific criteria relating to their specialist experience, their experience of assessment, evaluation and management, and their community and public service commitment and activity (McAvoy et al. 2001).

Training is dependent on the peer review method to be applied by the reviewer and is relevant across a number of peer review performance domains. These include but are not limited to training in the use of the formal assessment instruments to be applied; interview techniques; clinical governance; communication; and observational skills training (McAvoy et al. 2001).

The exact relationship between quality of peer review and training requirements of peer reviewers in health care remains to be defined. There were 2 comparisons of trained with untrained assessors (Jolly et al. 2001; Bowie 2008). Findings indicated that, irrespective of 'training', assessors rated peers unpredictably unless the peer review process was facilitated by the use of structured assessment instruments (Bowie 2008).

In another publication, participants who were required to submit a criterion audit of significant event analysis report in standard format for peer review were found to be unable to apply the audit methods and complete a report without training (McKay et al. 2006). Findings suggested that where the peer review methods to be applied require the use of new skills by the subject of the peer review process (for example, the development of a portfolio by the participant), training of participants may also be required.

Voluntary peer review methods

Voluntary systems are attractive as intrinsic motivation of the health care professional to drive quality improvement is generally less resource intensive for administering organisations, and are more consistent with adult learning principles (Overeem et al. 2007).

However, voluntary systems may not produce desired behaviour change (Norton et al. 1998; Lombarts and Klazinga 2003; Semmens et al. 2005; Hofhuis et al. 2006; Medical Council of Ireland 2008).

Further, voluntary peer review processes are feasible but often less reliable than mandatory processes, as they are more prone to modest to poor participation rates by health care professionals (Nelson et al. 2006) and often experience difficulties attracting sufficient peer reviewers (Medical Council of Ireland 2008). Voluntary peer review processes may experience poor participation rates as peer review activities are often time-consuming and / or resource intensive (Pacala et al. 2000; Walshe et al. 2001; Billett and Kendall 2005).

Organisations therefore may need to recognize the limitations of voluntary peer review processes when designing peer review systems and deciding what type(s) of peer review activities they wish to adopt or maintain for specific organizational purposes.

Peer review for under-performance

Where assessment of under-performance was the purpose of the peer review process, the methods used were usually formal and processes for peer review were generally well defined. Multiple peer review methods were commonly employed and multiple peer assessors used. Structured assessment methods for peer review were usually applied, and assessment and training of the peer reviewers frequently undertaken prior to the peer review occurring.

The purposes of the assessment of under-performance may be punitive and / or remedial. The implementation of remediation processes was outside scope of the review. However, publications made reference to end-point assessments, mentor programs, courses in management and communication skills for addressing specific under-performance issues identified through peer review. The use of sanctions, such as restricting scope of practice, specifying supervision requirements for practice, and limiting or removing legislation were also referred to. Regardless of purpose, peer review of under-performance was demonstrated to be a source of stress for both the health care provider being reviewed and for peers participating in the review process.

Limitations of this review and future directions

This review has several limitations. Firstly, it was restricted to English-language publications only. This means that bias due to exclusion of material in languages other than English cannot be ruled out. Secondly, the methodological quality of the studies varied greatly; therefore results should be interpreted with caution. Thirdly, most studies included have been conducted on small volunteer-based samples of participants. This limits the generalisability of the findings and increases the likelihood of selection bias influencing the results of individual studies. Finally, the literature regarding peer review in health care lacks the rigorous application of medical subject headings, compared with other topics in the published clinical literature. This may result in the non-retrieval of some studies.



Implications for Future Research

Although there is a large and growing body of literature regarding the peer review of health care professionals, much research in the area lacks methodologically rigorous appraisal of the peer review activity studied. Future studies in this area may benefit from greater attention being paid to the research methods used for evaluation of peer review activities. In particular, empirical evidence supporting the contribution of peer review to improvement in the clinical practice of health care professionals undergoing assessment is needed. Objectively measured improvement, rather than self-reported improvement would provide greater certainty that peer review contributes meaningfully to the maintenance and improvement of clinical competence in participants.

REFERENCES

1. Agee C. Improving the peer review process: develop a professional review committee for better and quicker results. *Healthcare Executive*. 2007;22(3):72-3.
 2. Allen J, DeSimone K. Valid peer review for surgeons working in small hospitals. *American Journal of Surgery*. 2002;184(1):16-8.
 3. Archer J, Norcini J, Davies J. Use of SPRAT for peer review of paediatricians in training. *British Medical Journal*. 2005;330(7502):1251-3.
 4. Bhatia A, Singh N, Arora V, Gupta K. Prospective peer review in fine needle aspiration cytology. Another step toward quality assurance. *Acta Cytologia*. 1998;42(4):865-8.
 5. Billett J, Kendall N, Old P. An investigation into GPs with high patient mortality rates: a retrospective study. *Journal of Public Health*. 2005;27(3):270-5.
 6. Bindels R, Hasman A, van Wersch J, Pop P, Winkens R. The reliability of assessing the appropriateness of requested diagnostic tests. *Medical Decision Making*. 2003;23(1):31-7.
 7. Borgstede J, Lewis R, Bhargavan M, Sunshine J. RADPEER quality assurance program: a multi-facility study of interpretive disagreement rates. *Journal of the American College of Radiologists*. 2004;1(4):59-65.
 8. Bowie P, McKay J, Dalgetty E, Lough M. A qualitative study of why general practitioners may participate in significant event analysis and educational peer assessment. *Quality and Safety in Health Care*. 2005;14(3):185-9.
 9. Bowie P, McKay J, Murray L, Lough M. Judging the quality of clinical audit by general practitioners: a pilot study comparing the assessments of medical peers and NHS audit specialists. *Journal of Evaluation in Clinical Practice*. 2008;14(6):1038-43.
 10. British Medical Journal, 2007. Clinical Evidence: About GRADE. BMJ Publishing Group. <http://clinicalevidence.bmj.com/cweb/about/index.jsp>
 11. Britton P, McCann J, O'Driscoll D, Hunnam G, Warren R. Interval cancer peer review in East Anglia: implications for monitoring doctors as well as the NHS breast screening program. *Clinical Radiology*. 2001;56(1):44-9.
 12. Brown N and Doshi M. Assessing professional and clinical competence: the way forward. *Advances in Psychiatric Treatment*. 2006;12:81-91.
 13. Campbell J, Richards S, Dickens A, Greco M, Narayanan A, Brearley S. 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*. 2008;17:187-93.
14. Campbell S, Roland M, Shekelle P, Cantrill J, Buetow S, Cragg D. Development of review criteria for assessing the quality of management of stable angina, adult asthma and non-insulin dependent diabetes mellitus in general practice. *Quality in Health Care*. 1999;8(1):6-15.
 15. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37-46.
 16. Crean P, Stokes M, Williamson C, Hatch D. Quality in paediatric anaesthesia: a pilot study of interdepartmental peer review. *Anaesthesia*. 2003;58(6):543-8.
 17. Curran V, Fleet L. A review of evaluation outcomes of web-based continuing medical education. *Medical Education*. 2005;39:561-7.
 18. Diaz L. Nursing peer review: developing a framework for patient safety. *The Journal of Nursing Administration*. 2008;38(11):475-9.
 19. Eaton K, Fleming W, Rich J. A report of an evaluation of the pilot peer review scheme for general dental practitioners working in the general dental services in England. *British Dental Journal*. 1998;184(4):178-82.
 20. Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. *British Medical Journal*. 2004;328:1240-5.
 21. Farin E, Follert P, Gerdes N, Jackel W, Thalau J. Quality assessment in rehabilitation centres: the indicator system 'Quality Profile'. *Disability and Rehabilitation*. 2004;26(18):1096-104.
 22. Forster A, O'Rourke K, Shojania K, van Walraven C. Combining ratings from multiple physician reviewers helped to overcome the uncertainty associated with adverse event classification. *Journal of Clinical Epidemiology*. 2007;60(9):892-901.
 23. Forsythe R, Livingston D, Lavery R, Mosenthal A, Hauser C. Autopsies in trauma do not add to peer review or quality assurance. *Journal of Trauma*. 2002;53(2):321-5.
 24. General Medical Council. UK. Business Plan 2009. Accessed at: http://www.gmc-uk.org/publications/business_plans/index.asp
 25. Gerlach F, Beyer M, Romer A. Quality circles in ambulatory care: state of development and future perspective in Germany. *International Journal of Quality in Health Care*. 1998;10(1):35-42.
-

26. Gluck P, Scarrow P. Peer review in obstetrics and gynaecology by a national medical specialty society. *Joint Commission Journal on Quality and Patient Safety*. 2003;29(2):77-84.
 27. Goulet F, Jacques A, Gagnon R, Racette P, Sieber W. Assessment of family physicians' performance using patient charts: inter-rater reliability and concordance with chart-stimulated recall interview. *Evaluation and the Health Professions*. 2007;30(4):376-92.
 28. Greenwood J, Williams R. Continuing professional development for Australian rural psychiatrists by videoconference. *Australas Psychiatry*. 2008;16(4):273-6.
 29. Guyatt G, Oxman A, Vist G et al. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*. 2008;336:924-6.
 30. Hall W, Violato C, Lewkonja R, Lockyer J, Fidler H, Toews J, et al. Assessment of physician performance in Alberta: the physician achievement review. *Canadian Medical Association Journal*. 1999;161(1):44-5.
 31. Heaton C. External peer review in Europe: an overview from the ExPeRT Project. *International Journal for Quality in Health Care*. 2000;12:177-82.
 32. Hofer T, Asch S, Hayward R, Rubenstein L, Hogan M, Adams J, et al. Profiling quality of care: is there a role for peer review? *BMC Health Services Research*. 2004;4(1):9.
 33. Hofer T, Bernstein S, DeMonner S, Hayward R. Discussion between reviewers does not improve reliability of peer review of hospital quality. *Medical Care*. 2000;38(2):152-61.
 34. Hofhuis H, Van Den Ende C, De Bakker D. Effects of visitation among allied health professionals. *International Journal of Quality in Health Care*. 2006;18(6):397-402.
 35. Hutchinson A, McIntosh A, Williams M, Barbour R. Senior health professional's perceptions of variations in medical practice: a qualitative and quantitative study. *Medical Education*. 2001;35(1):45-51.
 36. Hyrkas K, Koivula M, Lehti K, Paunonen-Ilmonen M. Nurse Managers' conceptions of quality management as promoted by peer supervision. *Journal of Nursing Management*. 2003;11(1):48-58.
 37. Ireland (Medical Council of). *What is peer review?* Dublin; 2008.
 38. Isetts B, Brown L, Schondelmeyer S, Lenarz L. Quality assessment of a collaborative approach for decreasing drug-related morbidity and achieving therapeutic goals. *Archives of Internal Medicine*. 2003;163(15):1813-20.
-

39. Jacobs D, Sarafin J, Huyhn T, Miles W, Sing R, Thomason M. Audience response system technology improves accuracy and reliability of trauma outcome judgments. *Journal of Trauma*. 2006;61(1):135-41.
 40. Jat A, Khan M, Zafar H, Raja A, Hoda Q, Rehmani R, et al. Peer review audit of trauma deaths in a developing country. *Asian Journal of Surgery*. 2004;27(1):58-64.
 41. Jensen M and Allen N. Meta-Synthesis of Qualitative Findings. *Quality in Health Research*. 1996;6:553-60.
 42. Johnstone P, Rohde D, May B, Peng Y, Hulick P. Peer review and performance improvement in a radiation oncology clinic. *Quality Management in Health Care*. 1999;8(1):22-8.
 43. Jolly B, Ayers B, MacDonald M, Armstrong P, Chalmers A, Roberts G, et al. The reproducibility of assessing radiological reporting: studies from the development of the General Medical Council's Performance Procedures. *Medical Education*. 2001;35(1):36-44.
 44. Kelly J. Development and implementation of an external peer-review process. *Journal of Healthcare Management*. 2005;50(2):137-42.
 45. Kernaghan D, Penney G. Do panels vary when assessing intrapartum adverse events? The reproducibility of assessments by hospital risk management groups. *Quality and Safety in Health Care*. 2006;15(5):359-62.
 46. Keyes G, Singer R, Iverson R, McGuire M, Yates J, Gold A, et al. Analysis of outpatient surgery center safety using an internet-based quality improvement and peer review program. *Plastic and Reconstructive Surgery*. 2004;113(6):1760-70.
 47. King L, Radford G, Gair R. Evaluation of peer review of infection control arrangements for healthcare-associated infections in four acute care trusts in the East of England. *Communicable Diseases and Public Health*. 2004;7(4):245-50.
 48. Kohn L, Corrigan J, Donaldson M, eds. *To err is human: building a safer health care system*. Washington, CD: Institute of Medicine, 1999.
 49. Kuhn D. Problem doctors: is there a system-level solution? Comment and response. *Annals of Internal Medicine*. 2006;144:861.
 50. Leape L and Fromson J. Problem doctors: is there a system-level solution? *Annals of Internal Medicine*. 2006;144:107-15.
 51. Lelliott P, Williams R, Mears A, Andiappan M, Owen H, Reading P. Questionnaires for 360-degree assessment of consultant psychiatrists: development and psychometric properties. *British Journal of Psychiatry*. 2008;193(2):156-60.
-

52. Liang B. Clinical assessment of malpractice case scenarios in an anesthesiology department. *Journal of Clinical Anesthesiology*. 1999;11(4):267-79.
 53. Lichtmacher A. Quality Assessment Tools: ACOG voluntary review of quality of care program, peer review reporting system. *Obstetrics and Gynaecology Clinics of North America*. 2008;35:147-62.
 54. Lockyer J, Violato C. An examination of the appropriateness of using a common peer assessment instrument to assess physician skills across specialties. *Academic Medicine*. 2004;79(10):5-8.
 55. Lockyer J, Violato C, Fidler H. The assessment of emergency physicians by a regulatory authority. *Academic Emergency Medicine*. 2006;13(12):1296-303.
 56. Lockyer J, Violato C, Fidler H. Assessment of radiology physicians by a regulatory authority. *Radiology*. 2008;247(3):771-8.
 57. Lockyer J, Violato C, Fidler H. A multi-source feedback program for anaesthesiologists. *General Anaesthesia*. 2005;53(1):33-9.
 58. Lombarts M, Klazinga N. A policy analysis of the introduction and dissemination of external peer review (visitatie) as a means of professional self-regulation amongst medical specialists in The Netherlands in the period 1985-2000. *Health Policy*. 2001;58(3):191-213.
 59. Lombarts M, Klazinga N. Supporting Dutch medical specialists in the implementation of visitatie recommendations: a descriptive evaluation of a 2-year project. *International Journal of Quality in Health Care*. 2003;15(2):119-29.
 60. Maidment Y. A comparison of the perceived effects on Scottish general dental practitioners of peer review and other continuing professional development. *British Dental Journal*. 2006;200(10):581-4.
 61. Maidment Y, Rennie J, Thomas M. Revalidation of general dental practitioners in Scotland: the results of a pilot study. Part 1 - feasibility of operation. *British Dental Journal*. 2006;200(7):399-402.
 62. Margo C. Peer and expert opinion and the reliability of implicit case review. *Ophthalmology*. 2002;109(3):614-8.
 63. Martin B, Fallow W, Palmieri P, Tomas E, Breedlove L. Autopsy data in the peer review process improves outcomes analysis. *Journal of Trauma*. 2007;62(1):69-73.
 64. McAvoy P, McCrorie P, Jolly B et al. Training the assessors for the General Medical Council's Performance Procedures. *Medical Education*. 2001;35(s1):29-35.
-

65. McKay J, Bowie P, Lough M. Variations in the ability of general medical practitioners to apply two methods of clinical audit: a five-year study of assessment by peer review. *Journal of Evaluation in Clinical Practice*. 2006;12(6):622-9.
 66. McKay J, Murphy D, Bowie P, Schmuck M, Lough M, Eva K. Development and testing of an assessment instrument for the formative peer review of significant event analyses. *Quality and Safety in Health Care*. 2007;16(2):150-3.
 67. Moss A, Dugal T, Silke B. Attitudes to peer review as a competence assurance structure - results of a survey of Irish physicians. *Irish Journal of Medical Science*. 2005;174(3):43-6.
 68. Neilson E, Johnson K, Rosenbloom S, Dupont W, Talbert D, Giuse D, et al. The impact of peer management on test-ordering behaviour. *Annals of Internal Medicine*. 2004;141(3):196-204.
 69. Nelson K, Wagner A, Gordon T, Silver R. Preoperative surgical case review program in a community-based, academic medical center: a feasibility study. *Journal of Reproductive Medicine*. 2006;51(5):399-404.
 70. Norcini J. Recertification in the United States. *British Medical Journal*. 1999;319:1183-5.
 71. Norman G, Shannon S, Martin M. The need for needs assessment in continuing medical education. *British Medical Journal*. 2004;328:999-1001.
 72. Norton P, Dunn E, Beckett R, Faulkner D. Long-term follow-up in the Peer Assessment Program for non-specialist physicians in Ontario, Canada. *Joint Commission Journal on Patient Quality and Safety*. 1998;24(6):334-41.
 73. Norton P, Faulkner D. A longitudinal study of performance of physicians' office practices: data from the Peer Assessment Program in Ontario, Canada. *Joint Commission Journal on Patient Quality and Safety*. 1999;25(5):252-8.
 74. Nuckols T, Bell D, Paddock S, Hilborne L. Contributing factors identified by hospital incident report narrative. *Quality and Safety in Health Care*. 2008;17(5):368-72.
 75. Olcott C, Mitchell R, Steinberg G, Zarins C. Institutional peer review can reduce the risk and cost of carotid endarterectomy. *Archives of Surgery*. 2000;135(8):939-42.
 76. Overeem K, Faber M, Arah O et al. Doctor performance assessment in daily practice: does it help doctors or not? *Medical Education*. 2007;41:1039-49.
 77. Pacala J, Kane R, Atherly A, Smith M. Using structured implicit review to assess quality of care in the Program of All-Inclusive Care for the Elderly (PACE). *Journal of the American Geriatric Society*. 2000;48(8):903-10.
-

78. Paige J, Aaron D, Yang T, Howell D, Hilton C, Cohn I, et al. Implementation of a preoperative briefing protocol improves accuracy of teamwork assessment in the operating room. *The American Surgeon*. 2008;74(9):817-23.
 79. Paukert J, Chumley-Jones H, Littlefield J. Do peer chart audits improve residents' performance in providing preventive care? *Academic Medicine*. 2003;78(10):39-41.
 80. Rankin J, Bush J, Bell R, Cresswell P, Renwick M. Impacts of participating in confidential enquiry panels: a qualitative study. *British Journal of Obstetrics and Gynaecology*. 2006;113(4):387-92.
 81. Rethans J, Sturmans F, Drop R et al. Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. *British Medical Journal*. 1991;303:1377-80.
 82. Royal Australasian College of Surgeons. *Surgical Competence and Performance Guide*, Melbourne, Australia. 2006.
 83. Royal Australasian College of Surgeons. *Continuing Professional Development Information Manual 2007-2009*, Melbourne, Australia. 2008.
 84. Sargeant J, Mann K, Ferrier S, Langille D, Muirhead P, Hayes V, et al. Responses of rural family physicians and their colleague and co-worker raters to a multi-source feedback process: a pilot study. *Academic Medicine*. 2003;78(10):42-4.
 85. Saturno P, Palmer R, Gascon J. Physician attitudes, self-estimated performance and actual compliance with locally peer-defined quality evaluation criteria. *International Journal of Quality in Health Care*. 1999;11(6):487-96.
 86. Semmens J, Aitken R, Sanfilippo F, Mukhtar S, Haynes N, Mountain J. The Western Australian Audit of Surgical Mortality: advancing surgical accountability. *Medical Journal of Australia*. 2005;183(10):504-8.
 87. Shanti C, Tyurski J, Rishell K, Wilson R, Lozen Y, Seibert C, et al. Correlation of revised trauma score and injury severity score (TRISS) predicted probability of survival with peer-reviewed determination of trauma deaths. *The American Surgeon*. 2003;69(3):257-60.
 88. Sibley J, Sackett D, Neufeld V et al. A randomized trial of continuing medical education. *New England Journal of Medicine*. 1982;306:511-5.
 89. Silver M, Hougland P, Elder S, Haug J, Pritchett T, Donnelly S, et al. Statewide identification of adverse events using retrospective nurse review: methods and outcomes. *Journal of Nursing Measurement*. 2007;15(3):220-32.
-

90. Silverstein M, Ballard D. Expert panel assessment of appropriateness of abdominal aortic aneurysm surgery: global judgment versus probability examination. *Journal of Health Services Research and Policy*. 1998;3(3):134-40.
 91. Soroka M, Feldman L, Crump T. Quality-of-care review of optometric records: inter-rater reliability. *Journal of healthcare Quality*. 2004;26(5):29-33.
 92. Southgate L, Campbell M, Cox J, Foulkes J, Jolly B, McCrorie P, et al. The General Medical Council's Performance Procedures: the development and implementation of tests of competence with examples from general practice. *Medical Education*. 2001;35(1):20-8.
 93. (b) Southgate L, Hays R, Norcini J et al. Setting performance standards for medical practice: a theoretical framework. *Medical Education* 2001;35:474-81.
 94. Spigelman A, Swan J. Measuring clinical audit and peer review practice in a diverse health care setting. *Australian and New Zealand Journal of Surgery*. 2003;73(12):1041-3.
 95. Steinbrook R. Renewing board certification. *New England Journal of Medicine*. 2005;142:260-73.
 96. Stekelenburg J, van Roosmalen J. The maternal mortality review meeting: experiences from Kalabo District Hospital, Zambia. *Tropical Doctor*. 2002;32(4):219-23.
 97. Stumpf P. Voluntary review of quality of care peer review for patient safety. *Best Practice Research in Clinical Obstetrics and Gynaecology*. 2007;21(4):557-64.
 98. Sucov A, Nathanson A, McCormick J, Proano L, Reinert S, Jay G. Peer review and feedback can modify pain treatment patterns for emergency department patients with fractures. *American Journal of Medical Quality*. 2005;20(3):138-43.
 99. Takayanagi K, Koseki K, Aruga T. Preventable trauma deaths: evaluation by peer review and a guide for quality improvement. *Emergency Medical Study Group for Quality. Clinical Performance and Quality in Health Care*. 1998;6(4):163-7.
 100. Thomas P, Gebo K, Hellmann D. A pilot study of peer review in residency training. *Journal of General Internal Medicine*. 1999;14(9):551-3.
 101. Thomas, D. The Rise and Recession of Medical Peer Review in New South Wales 1856-1994. Doctoral Thesis, University of Sydney, Sydney. 2002.
 102. Thompson A, Ashraf Z, Burton H, Stonebridge P. Mapping changes in surgical mortality over 9 years by peer review audit. *British Journal of Surgery*. 2005;92(11):1449-52.
-

103. Thomson O'Brien M, Freemantle N, Oxman A et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews*. 2001;2:CD003030.
 104. van Eijk M, Avorn J, Porsius A, de Boer A. Reducing prescribing of highly active anticholinergic antidepressants for elderly people: a randomised trial of group versus individual academic detailing. *British Medical Journal*. 2001;322(7287):654-7.
 105. Violato C, Lockyer J, Fidler H. Assessment of paediatricians by a regulatory authority. *Paediatrics*. 2006;117:796-802.
 106. Walshe K, Wallace L, Freeman T, Latham L, Spurgeon P. The external review of quality improvement in health care organisations: a qualitative study. *International Journal of Quality in Health Care*. 2001;13(5):367-74.
 107. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet* 2003;357:945-9.
 108. Watters D, Green A, van Rij A. Guidelines for surgical audit in Australia and New Zealand. *Australian and New Zealand Journal of Surgery* 2006;76:78-83.
 109. Weingart S, Mukamal K, Davis R, Davies D, Palmer R, Cahalane M, et al. Physician-reviewers' perceptions and judgments about quality of care. *International Journal of Quality in Health Care*. 2001;13(5):357-65.
 110. Wendling A, Hoekstra L. Interactive peer review: an innovative resident evaluation tool. *Family Medicine*. 2002;34(10):738-43.
 111. Zardawi I, Bennett G, Jain S, Brown M. Internal quality assurance activities of a surgical pathology department in an Australian teaching hospital. *Journal of Clinical Pathology*. 1998;51(9):695-9.
-

APPENDIX 1 – SUMMARY TABLES

TABLE 1 – ASSESSMENT OF THE CLINICIAN’S DOMAINS OF PROFESSIONAL PRACTICE

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
a) Assessment of a single domain					
<i>i. Survey of peers</i>					
Paige 2008	Evaluation of multi-source feedback in peri-operative setting.	USA. Operating room teamwork performance at a rural community hospital evaluated before and after training and implementation of a preoperative briefing protocol.	After each operative case, every team member completed a multi-source feedback survey assessing teamwork behaviour. The survey contained both self and peer assessments.	Data were assessed for 1 general surgeon and 9 operating room staff. Individuals over-estimated their teamwork behaviours before protocol implementation. Post protocol implementation, self-assessment accuracy and teamwork improved.	C
<i>ii. Chart audit</i>					
Paukert	Evaluation of method for assessment by	USA.	Residents audited charts and scored their peers on 12 clinical	1005 charts were audited.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
2003	peer chart audit.	Peer review of medical residents to assess competence.	preventive services.	Audit of their peers' charts resulted in significant improvement in the resident's own performance.	
Soroka 2004	Evaluation of method for assessment of clinical treatment by medical record review.	USA. Peer review assessment of optometry care provided.	20 examiners assessed medical records using a review instrument developed for assessment of optometry care.	29 records were reviewed. Inter-rater reliability was poor. The instrument was unreliable in measurement of optometry care.	B
<i>iii. Group peer discussion</i>					
Gerlach 1998	Descriptive study of state of development of peer review groups (for educational peer review).	Germany. Survey of the quantitative development of peer review groups in ambulatory care.	Survey of all regional associations of Statutory Health Insurance Physicians.	Peer review group numbers increasing but few peer review groups evaluated.	C
Hyrkas 2003	Qualitative study. Face to face interviews.	Finland. Assessment of the influence of peer review on	9 nurse managers participated in regular group feedback.	Peer review assists nurse managers by providing support and feedback for	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
		nurse managers' quality management practice.		reflection on practice.	
b) Assessment of multiple domains					
<i>i. Survey of peers</i>					
Archer 2005	Evaluation of cross-sectional surveys for assessment of clinical performance.	UK. Hospital-based assessment of paediatric senior house officers and specialist registrars.	Multi-source feedback questionnaire (Sheffield peer review assessment tool – SPRAT) used to evaluate trainee clinical performance. Ratings obtained were from an average of 10 raters, for 24 questions across 5 domains (good clinical care; maintaining good medical practice; teaching, training, assessing and appraising; relationships with patients; and working with colleagues).	112 paediatric staff, assessed using 921 completed questionnaires, participated. Time taken by raters to complete the questionnaire was six minutes. Reliability was obtained for most doctors with four questionnaires completed.	B
Campbell 1999	Descriptive study. Development of peer review panel method for assessing quality	UK. Peer review of medical practitioners (hospital and community) for assessment	Three multi-professional panels (made up of 6 GPs per panel, hospital specialists and practice nurses) rated doctor's clinical performance in managing	The reliability of peer review in assessing clinical performance was improved with the addition of a systematic approach	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	of care provided by medical practitioners.	of quality of management of three common chronic diseases.	<p>specific chronic conditions against criteria derived from clinical practice guidelines.</p> <p>Panel members used evidence-based guidelines to construct review criteria for use in peer assessment of medical practitioners standard of care.</p>	to assessment, against recognized clinical standards.	
Hall 1999	Evaluation of multi-source feedback using structured questionnaires.	<p>Canada.</p> <p>Peer review program providing structured feedback to physicians.</p>	<p>Program assesses 90 attributes across the 6 domains of medical knowledge and skills, attitudes and behaviour, professional responsibilities, practice improvement activities, administrative skills, and personal health.</p> <p>6 medical peers, 6 consulting physicians to whom patients were referred and 6 non-physician co-workers completed structured questionnaires.</p>	Coefficients of variation ranged from 0.64 with 4 peer reviewers to 0.82 with 10 peer reviewers.	B
Ireland (Medical)	Evaluation of multi-source feedback	Ireland.	Convenience sample of GPs recruited.	A total of 62 of 167 eligible GPs completed the study.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Council of) 2008	using structured questionnaires.	Peer review by multi-source feedback to assess professional competence of GPs.	<p>GPs nominated medical colleagues the Medical Council should ask to provide feedback on the GP's professional practice.</p> <p>An electronic questionnaire was then completed by medical colleagues and patients in order to assess the clinical competence, patient interaction, professional self-management, and psychosocial management of patients by the participant.</p> <p>A third party collated responses to questionnaires and provided a confidential report directly to the participating GP.</p>	<p>Two-thirds of GPs reported an intention to make changes to their practice based on their feedback.</p> <p>Participants reported difficulties persuading medical peers to participate as peer reviewers.</p> <p>As the process was voluntary some participants took months to complete the process.</p> <p>Administration of the process was time-consuming.</p>	
Lelliott 2008	Evaluation of multi-source feedback using structured questionnaires.	UK. Peer review program providing structured feedback to psychiatrists.	<p>Questionnaire was administered to doctors, patients and clinical colleagues (peers).</p> <p>The questionnaire comprised 25 items (contextual,</p>	<p>347 participant psychiatrist and 4422 colleague ratings analysed.</p> <p>Ratings from 13 peers were required to achieve a generalisability coefficient</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			performance evaluation, descriptive and free text).	of sufficient size (rho>0.75).	
Lockyer 2004	Evaluation of peer survey using structured questionnaires.	Canada. Assessment of clinical competencies using a common peer review instrument applied across internal medicine, paediatric and psychiatry clinicians.	Survey with 36 items assessed psychiatry (n=101), paediatrics (n=100) and internal medicine (n=103) specialists. An average of 7 to 8 peers competed the survey for each participant.	2306 surveys were analysed. Cronbach's alpha was 0.98. Generalisability coefficient (mean of 7.6 raters) was 0.83. The instrument was able to be applied across specialties.	B
Lockyer 2005	Evaluation of multi-source feedback using structured questionnaires.	Canada. Peer review of performance of anaesthetists.	Surveys with 19 and 29 items were applied to participants by 8 co-workers and 8 medical colleagues respectively. Domains included communication skills, professionalism, collegiality, continuing professional development, and collaboration.	Data were collected for 186 anaesthetists. Instruments had high internal reliability (Cronbach's alpha greater than 0.95) and generalisability coefficients of 0.56 for co-workers and 0.69 for medical colleagues.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Lockyer 2006	Evaluation of multi-source feedback using structured questionnaires.	Canada. Peer review of performance of emergency physicians.	Surveys with 20 and 30 items were applied to participants by 8 co-workers and 8 medical colleagues respectively. Domains included communication skills, professionalism, collegiality, and self-management.	Data were collected for 187 emergency physicians. Instruments had high internal reliability (Cronbach's alpha greater than 0.94) and generalisability coefficients of 0.85 for co-workers and 0.84 for medical colleagues.	B
Lockyer 2008	Evaluation of multi-source feedback using structured questionnaires.	Canada. Peer review of performance of radiology physicians.	Surveys with 38, 29 and 20 items were applied to participants by 8 radiologists, 8 referring physicians and 8 co-workers respectively. Domains included clinical competence, collegiality, professionalism, workplace behaviour, and self-management.	Data were collected from 190 radiologists. Reliability was high (Cronbach's alpha greater than 0.95). Generalisability coefficients were 0.88, 0.79 and 0.87 respectively.	B
Sargeant 2003	Evaluation of multi-source feedback	Canada. Peer review of rural family	Surveys were applied to participants by patients, co-workers and medical	Data were collected for 113 physicians.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	instrument.	physicians.	colleagues. 8 medical colleagues (professional peers) completed the questionnaire. Familiarity of each group of raters with the participant was assessed.	Medical colleagues were significantly more familiar with participants than were consultants.	
Thomas 1999	Evaluation study. Survey method to assess clinical competence.	USA. Peer review of internal medicine interns.	Senior residents and interns completed survey-based evaluations of interns and the end-of-month ward rotations.	177 analyses for 16 interns. Ratings between assessors were highly correlated. Instrument showed high internal consistency.	B
Violato 2006	Evaluation of multi-source feedback using structured questionnaires.	Canada. Peer review of performance of paediatricians.	Surveys with 22 and 38 items were applied to participants by 8 co-workers and 8 medical colleagues respectively. Domains included communication skills, professionalism, collegiality, continuing professional	Data were collected for 100 paediatricians. Instruments had high internal reliability (Cronbach's alpha greater than 0.94) and generalisability coefficients of 0.87 for co-	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			development, and collaboration.	workers and 0.78 for medical colleagues.	
<i>ii. Chart audit</i>					
Farin 2004	<p>Descriptive study.</p> <p>Description of quality assessment program provided.</p>	<p>Germany.</p> <p>Quality assessment in rehabilitation centres, incorporating peer review of rehabilitation care across the domains of anamnesis, diagnosis, therapy, clinical epicrisis, socio-medical epicrisis and after-care.</p>	<p>A random sample of discharge reports from rehabilitation centres were selected for peer review.</p> <p>Discharge reports were forwarded together with the corresponding therapy plan for the client to senior physicians who worked at other rehabilitation centres.</p> <p>Senior physicians evaluated the reports using a structured checklist based on evidence-based guidelines.</p> <p>Senior physicians graded discharge reports and therapy plans accompanying them using the criteria.</p>	<p>Peer review process generated data that enabled assessment of rehabilitation centres across dimensions of quality.</p> <p>Data provided a basis for benchmarking facilities against key quality measures.</p>	C
<i>iii. Site visit</i>					

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Moss 2005	Evaluation of physician's attitudes to peer review assessment.	Ireland. Peer review of physician competence.	Survey of all physicians in the Irish Medical Directory.	Response rate 67%. 87% reported on-site assessment of competence every 5 years was the preferred method. 70% would pay to be assessed.	C
<i>iv. Group peer discussion</i>					
Wending 2002	Descriptive study of interactive peer review method.	USA. Peer review of inpatient family practice residents.	First, second and third year residents evaluated each other in discussion format. There were between 5 and 7 residents on each rotation. The resident who was the subject of the review was provided with verbal feedback by all other residents on the rotation in a group meeting facilitated by a faculty staff member. The faculty staff member collated feedback and prepared a written report for each	Majority rated the peer review process as useful. Peer comments correlated well (qualitative assessment) with resident's own impressions of their performance.	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			resident containing their individual feedback.		
<i>v. Combination of methods</i>					
Eaton 1998	Evaluation study of dentist peer review method.	UK. Peer review of general dental practitioners working in general dental services.	Dentists completed a resume, which was then reviewed by three peer reviewers, who had received 'consistency training'.	Participants reported that they would undertake initiatives to improve patient services. No reliability or validity measures reported. No effectiveness criteria measured.	C
Goulet 2007	Comparative study of peer review methods for assessment of performance.	Canada. Peer review of GPs' performance.	25 GPs selected at random to be reviewed by peers. For each, 25 to 40 patient records randomly selected for evaluation. 8 GPs were trained as peer reviewers. 1 peer performed an assessment of the GP's professional practice by	Chart-stimulated recall interviews provide additional information for assessment of the GP's performance, compared with chart review alone.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>performing a chart audit and interviewing the GP about the cases referred to in the charts.</p> <p>Another peer performed an assessment of the GP's professional practice by performing a chart audit only.</p> <p>Items assessed were record keeping, clinical investigation, diagnostic accuracy and treatment plan.</p> <p>The results of the two peer reviewers were then compared.</p>		
Norton 1998	Evaluation of peer review method for assessment of office-based practices.	Canada. Peer review of quality of care and record keeping of non-specialist physicians.	Randomly selected physician office practices assessed by a physician peer who performed a tour of the premises and review of random selection of 20 to 30 medical records to evaluate record keeping and content of the records.	Follow-up of physician practices demonstrated the peer review process improved practice.	B
Norton 1999	Evaluation of peer review method to	Canada.	A physician peer reviewed the physical facilities, the system of	109 non-specialist physicians re-assessed at	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	determine the quality of care provided by physicians.	Peer review of all physicians who turn 70 years of age were assessed, and a random selection of physicians within specific practice and specialty areas.	<p>record keeping, the content of 20 to 30 medical records, and interviewed the physician being assessed.</p> <p>The assessed physician was then assigned a grade.</p>	<p>least 10 years apart.</p> <p>Performance declined over the time period for most physicians (64%).</p>	

TABLE 2 – ASSESSMENT OF DELIVERY OF CARE IN ACCORDANCE WITH CLINICAL GUIDELINES

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
<i>i. Review of pathology / radiology requests</i>					
Bindels 2003	Evaluation of peer review of pathology testing.	The Netherlands. Peer review of pathology tests ordered by GPs to assess appropriateness of pathology request.	A random selection of 1217 tests from 253 requests made by GPs were assessed for appropriateness by 3 peer reviewers.	Substantial variation was found in assessment of appropriateness of requested pathology tests. Inter-rater kappa values ranged from 0.33 to 0.42. Sixteen reviewers were needed to obtain joint reliability of 0.95.	B
Neilson 2004	Evaluation of peer review of pathology and radiology testing.	USA. Peer review of pathology tests ordered by all hospital clinical staff to assess appropriateness of pathology request.	Hospital Resource Utilization Committee (17 members from across clinical specialties) peer reviewed the ordering habits of providers against practice guidelines. On the basis of results of peer review, the software used by clinicians to order tests was modified in order to change test ordering patterns.	Reduction in radiology and pathology testing was achieved. Results of peer review informed necessary modifications to test ordering software.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
<i>ii. Chart audit</i>					
Campbell 1999	Descriptive study. Development of peer review panel method for assessing quality of treatment of chronic diseases.	UK. Peer review of medical practitioners (hospital and community) for assessment of quality of management of three common chronic diseases.	Three multi-professional panels (made up of 6 GPs per panel, hospital specialists and practice nurses) rated doctor's clinical performance in managing specific chronic conditions against criteria derived from clinical practice guidelines. Panel members used evidence-based guidelines to construct chart review criteria for use in peer assessment of medical practitioners standard of care.	The reliability of peer review in assessing clinical performance was improved with the addition of a systematic approach to assessment, against recognized clinical standards.	B
Hofer 2004	Evaluation of peer review of patient records for assessing treatment of chronic diseases.	USA. Peer review of medical records for assessing quality of care for chronic disease of patients who had received both inpatient and outpatient care.	Structured implicit review instrument developed to guide peer review assessment. Instrument based on evidence based clinical guidelines. Reviewers trained to perform peer review. 3 peer reviewers independently	12 reviewers conducted a total of 496 reviews of 70 patient records. Reliability of single physician review is low (ICC 0.16 to 0.26). Reliability is higher for conditions where the evidence base is well-developed (e.g.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			assessed medical records against items in the instrument.	ICC 0.46 for diabetes and hypertension).	
Isetts 2003	Evaluation of pharmacist peer review method.	USA. Peer review method for assessing the pharmaceutical decision-making of pharmacists.	A panel of 12 pharmacists and physicians assessed the quality of therapeutic determinations made by pharmacists. Measures for assessment included drug-related morbidity and achievement of therapeutic goals by medical record review. An implicit review process was used.	A total of 4779 panel evaluations of clinical decisions were made. ICCs for agreement of panel members ranged from 0.73 to 0.85.	B
Johnstone 1999	Descriptive study of method for peer review of quality of care.	USA. Peer review for performance improvement and physician re-credentialing in a radiation oncology clinic. Program based on criteria established by the American College of	All radiation oncologists attended weekly chart rounds. Radiation oncologists reviewed each other's patient charts at meetings for laboratory and pathology reports, and adequacy of the doctor's documentation. Films were checked by a peer radiation oncologist.	Peer review method commonly identified inconsistent chart documentation. Method used to compare care received with clinical practice standards.	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
		Radiology.	<p>A checklist was then used to assess the patient chart for adequacy of care (against clinical practice guidelines) and a subjective evaluation of overall quality is made.</p> <p>Participants received their feedback from the process monthly.</p> <p>The institution also used findings to inform credentialing processes.</p>		
Saturno 1999	<p>Comparative study.</p> <p>Comparison of self-audit with peer audit.</p>	<p>Spain.</p> <p>Peer review of physicians to assess compliance with guidelines.</p>	<p>50 cases of both upper respiratory tract infection and high cholesterol were randomly selected from patient presentations to a health centre in the previous 6 months.</p> <p>2 trained peer reviewers abstracted information from medical records.</p> <p>Peer reviewers assessed treatment against 13 quality</p>	Physicians rated their own performance as better than their peers.	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>evaluation criteria.</p> <p>Results of self-audit were compared with peer audit.</p>		
Stumpf 2007	<p>Descriptive study.</p> <p>Description of quality assessment program provided.</p>	<p>USA.</p> <p>American College of Obstetricians and Gynaecologists Voluntary Review of Quality of Care program.</p>	<p>3 peer reviewers assess 3 to 5 medical records on-site.</p> <p>Each record is reviewed independently, using a worksheet with explicit objective criteria based on clinical practice guidelines.</p> <p>Departmental and hospital personnel are also interviewed on-site to obtain further information.</p>	<p>The majority of deficiencies were in institutional systems rather than specifically obstetric or gynaecologic clinical problems.</p> <p>The types of deficiencies found appeared to be similar across the spectrum of institutions visited, and not related to the size of the institution or of the department of obstetrics and gynaecology, or the number of deliveries done annually.</p> <p>No reliability or validity measures reported.</p> <p>No effectiveness criteria measured.</p>	C
<i>iii. Face to face assessment / visit to clinical practice</i>					

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Sucov 2005	Evaluation study of individual and group feedback on treatment behaviours.	USA. Peer review used to modify pain treatment provided by emergency department staff to patients with fractures.	The treatment of consecutive patients presenting with fractures was reviewed by peers in the Department of Emergency Medicine. Physicians received individual feedback regarding their clinical practice and feedback regarding the practice of their department as a whole.	1454 patients participated. Treatment rates increased by 30% post intervention.	C
Van Eijk 2001	Randomised controlled trial of practice visit methods.	The Netherlands. Peer review of GPs and pharmacists to reduce prescribing of anti-cholinergic antidepressants.	190 GPs and 37 pharmacists were organised into 21 peer review groups. Participants received 2 educational visits to target their prescribing of anti-cholinergic antidepressants. Visits were provided to participants either individually or to the group as a whole.	A 26% reduction in prescribing was observed in the individual visits intervention arm. A 45% reduction in prescribing was observed in the group intervention arm.	A
<i>i. Survey instrument / questionnaire</i>					

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Nelson 2006	Evaluation of peer review method for preoperative assessment.	USA. Peer review of gynaecologists to assess appropriateness of gynaecologic surgery.	<p>All surgeons voluntarily submitted a 1-page case summary questionnaire for each elective gynaecologic surgical procedure.</p> <p>A committee of departmental peers reviewed each submission for appropriateness of indications and completeness of preoperative evaluation.</p> <p>Cases were evaluated against American College of Obstetricians and Gynaecologists' quality assessment criteria.</p>	<p>Reviewed summaries were categories as either meeting criteria, not meeting criteria but appropriate or not meeting criteria and no criteria available for assessment.</p> <p>86% of procedures were deemed appropriate.</p> <p>Participation was poor as the program was voluntary.</p>	C

TABLE 3 – ASSESSMENT OF ORGANISATIONAL QUALITY OF CARE

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
<i>i. Chart audit</i>					
Pacala 2000	Evaluation of structured implicit review.	USA. Peer review of care provided to patients enrolled in the Program for All-Inclusive Care of the Elderly.	Trained geriatricians and geriatric nurse practitioners peer reviewed medical records using a structured implicit review assessment instrument.	Inter-rater reliability was poor. Reviews were time-consuming.	B
Weingart 2001	Evaluation of method for assessment of quality of care by structured implicit chart review.	USA. Peer review of quality of care by hospital physicians.	Hospitalisation records for 1134 patients were assessed. A single peer reviewer reviewed a patient chart using a structured chart review instrument and provided written open-ended comments about each case. 2 reviewers then re-reviewed the patient chart using the same structured instrument. Reviewers assessed records for	Reviewers questioned quality more frequently in cases with serious or fatal outcomes, technical mishaps and inadequate documentation.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			deficiencies in quality of care in surgery, anaesthetics and medical domains.		
<i>ii. Multiple methods</i>					
King 2004	Evaluation of peer review method for assessing infection control arrangements.	UK. Peer review of medical practitioners' infection control practices.	<p>A peer review team of epidemiologists, a public health physician and a microbiologist were formed to assess infection control practice.</p> <p>The peer review team met with infection control doctors, nurses, laboratory staff and other relevant hospital personnel.</p> <p>Staphylococcus aureus bacteraemia, healthcare associated infections and infection control were discussed.</p> <p>The peer review team reviewed data with hospital staff and offered to provide hospital staff with assistance in improving infection control rates.</p>	<p>Hospital staff who participated reported that the peer review process raised the profile of infection control and healthcare-associated infection within their facilities.</p> <p>Changes in rates of healthcare associated infection occurring during the time period the peer review process was conducted were not reported in the publication.</p>	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>The peer review team then drafted an action plan for the hospital, generated from discussions at the meeting.</p>		
Hofhuis 2006	Evaluation of peer review by 'visitatie'.	<p>The Netherlands.</p> <p>Peer review of allied health professionals in community practice.</p>	<p>Peer review teams visited allied health practices and provided feedback and recommendations for improvement to participants on the basis of site inspection, audit of patient charts and staff interviews.</p> <p>Evaluation questionnaires sent to 151 allied health professionals who had received a visitation from peer reviewers.</p> <p>Questionnaire examined change in awareness of weak and strong aspects of competence and performance, intention to carry out the recommendations, and actual improvements.</p>	Participants increased awareness of strengths more than weaknesses in care delivered. Actual improvements were made on 33% of the aspects.	B
Lombarts 2001	<p>Qualitative study.</p> <p>Policy analysis of</p>	<p>The Netherlands.</p> <p>Peer review as a</p>	Policy analysis.	Visitatie has served as a strategy to protect the	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	introduction and dissemination of external peer review by 'visitatie'.	means of professional self-regulation of medical specialists.		autonomy of physicians.	
Lombarts 2003	Descriptive study of application of recommendations of peer review 'visitatie'.	The Netherlands. Peer review of medical specialists for improving quality of patient care.	Medical specialist practices participated first in visitatie (Dutch external peer review process involving site visits, inspection of medical records and interviews with staff). 'Quality Consultations' carried out in 25 specialist group practices (67 specialists) across surgery, paediatrics and gynaecology by two management consultants. 20 hours of management consultancy in each practice to assist practices to implement recommendations of visitatie (external peer review) process.	Management consultants assisted medical specialty practices to implement the recommendations of the external peer review process.	C
Lichtmacher 2008	Descriptive study of voluntary peer review method for assessment of areas	USA. Peer review of obstetric and	Hospital makes a request to ACOG to perform a VRQC review.	Reasons for request for review commonly include organizational problems, problems with a specific	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	of clinical concern.	gynaecology care, in the American College of Obstetrics and Gynaecology (ACOG) Voluntary Review of Quality of Care (VRQC) program.	<p>Hospital provides patient chart to ACOG and any background documentation required to undertake the review.</p> <p>A team of 3 peer reviewers is selected from a panel of trained reviewers by ACOG.</p> <p>The peer reviewers assess background information, conduct interviews with the relevant hospital leadership group, conduct a tour of the facility and evaluate its resources, perform confidential staff interviews, conduct an objective evaluation of the clinical records against pre-determined criteria, and prepare a written report outlining findings and recommendations.</p>	<p>physician or group of physicians, or desire to evaluate the hospital's obstetric and gynaecology performance against accepted standards of care.</p> <p>Problems commonly identified across each domain include: poor communication, poor documentation of clinical reasoning, lack of effective quality assurance program, and inadequate nurse staffing levels.</p>	
Gluck 2003	<p>Descriptive study.</p> <p>Description of quality assessment program provided.</p>	<p>USA.</p> <p>American College of Obstetricians and Gynaecologists Voluntary Review of</p>	<p>Hospital makes a request to professional body to perform a peer review.</p> <p>Hospital provides patient chart to professional body and any</p>	Data from the first 100 site visits revealed departmental and systems deficiencies were more common triggers for review than clinical	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
		Quality of Care program.	<p>background documentation required to undertake the review.</p> <p>A team of 3 peer reviewers is selected from a panel of trained reviewers.</p> <p>Peer reviewers assess 3 to 5 medical records on-site.</p> <p>Each record is reviewed independently, using a worksheet with explicit objective criteria based on clinical practice guidelines.</p> <p>Departmental and hospital personnel are also interviewed on-site to obtain further information.</p>	concerns.	
Stumpf 2007	<p>Descriptive study.</p> <p>Description of quality assessment program provided.</p>	<p>USA.</p> <p>American College of Obstetricians and Gynaecologists Voluntary Review of Quality of Care</p>	<p>3 peer reviewers assess 3 to 5 medical records on-site.</p> <p>Each record is reviewed independently, using a worksheet with explicit objective criteria based on clinical practice</p>	<p>The majority of deficiencies were in institutional systems rather than specifically obstetric or gynaecologic clinical problems.</p> <p>The types of deficiencies</p>	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
		program.	<p>guidelines.</p> <p>Departmental and hospital personnel are also interviewed on-site to obtain further information.</p>	<p>found appeared to be similar across the spectrum of institutions visited, and not related to the size of the institution or of the department of obstetrics and gynaecology, or the number of deliveries done annually.</p> <p>No reliability or validity measures reported.</p> <p>No effectiveness criteria measured.</p>	
Crean 2003	Evaluation of peer review teams performing hospital site visits.	<p>UK</p> <p>Peer review of paediatric anaesthetic departments.</p>	<p>Peer review visits were conducted by teams which included medical (paediatric anaesthetists) and non-medical members.</p> <p>Peer reviewers collected evidence of sound departmental structure, organization and management, as outlined in the Joint Committee Good Practice Guide for Departments of Anaesthesia.</p>	<p>Peer review team members 'felt the process was useful'.</p> <p>No objective measures of outcomes assessed.</p>	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Walshe 2001	Qualitative study of external peer review method for assessing health care provider organisations.	UK. Peer review within health care provider organizations by clinical governance review teams who inspected premises, interviewed staff, and audited specific aspects of practice.	Face to face and telephone interviews were conducted with senior managers, clinicians and members of a regional clinical governance review team pre and post clinical governance review within their organisations.	163 participants. Preparing for external review was resource-intensive and time-consuming. External review did not lead to major changes in policy, strategy or practice.	B

TABLE 4 – PEER REVIEW AS A REQUIREMENT FOR CONTINUING PROFESSIONAL DEVELOPMENT

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Bowie 2008	Comparative study of accuracy of clinical audit by auditor type.	UK. Peer review by GPs of significant event analysis (SEA) reports and clinical audit reports.	11 trained and experienced GP audit assessors recruited. 10 novice assessors (no prior training or experience) recruited. Both groups used structured instruments to assess SEA reports and criterion audit reports submitted by GPs as part of their continuing professional development requirements (and for GP registrars as part of their summative assessment). 12 SEA reports and 12 clinical audit reports peer reviewed independently by the members of both groups.	The novice peer reviewers gave numerically accurate feedback and their results were comparable with experienced medical peer assessors.	B
Greenwood 2008	Descriptive study. Description of continuing professional development	Australia. Peer review group for education and continuing professional development of isolated	Peer review group formed to reduce professional isolation experienced by isolated psychiatrists and to enable participation in continuing professional development.	106 psychiatry professionals participated in the national sessions. Video-conference attendance at seminars	C

	program provided.	rural psychiatrists and trainees.	6 national interactive video-conference education seminars were provided.	provided learning opportunities for rural psychiatrists.	
Maidment 2006	Comparative study of methods for peer review of dental care.	Scotland. Peer review to inform provision of continuing professional development.	Postal questionnaire to 268 dentists.	Dentists viewed peer review as an acceptable method for continuing professional development. Courses and reading journals were ranked higher than peer review as acceptable and effective for changing knowledge and practice.	C

TABLE 5 – ASSESSMENT OF SIGNIFICANT EVENTS

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
a) General					
<i>i. Peer review of written reports of significant events</i>					
Bowie 2005	Qualitative analysis of focus group transcripts.	UK. GP-based significant event analysis (SEA) and educational peer assessment.	Two independent GP assessors reviewed SEA reports using a standardized instrument, and provide feedback to the participant about improvements, if required.	SEA reports are submitted by GPs for peer assessment resulting in educational rewards being obtained. GPs are highly selective about which SEAs they submit because of concerns about confidentiality, litigation or professional embarrassment.	C
Bowie 2008	Comparative study of accuracy of clinical audit by auditor type.	UK. Peer review by GPs of significant event analysis (SEA) reports and clinical audit reports.	11 trained and experienced GP audit assessors recruited. 10 novice assessors (no prior training or experience) recruited. Both groups used structured instruments to assess SEA	The novice peer reviewers gave numerically accurate feedback and their results were comparable with experienced medical peer assessors.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>reports and criterion audit reports submitted by GPs as part of their continuing professional development requirements (and for GP registrars as part of their summative assessment).</p> <p>12 SEA reports and 12 clinical audit reports peer reviewed independently by the members of both groups.</p>		
McKay 2006	Comparative study of two methods of clinical practice audit.	<p>UK.</p> <p>Voluntary peer review of GPs using clinical audit or significant event analysis (SEA) reports.</p>	<p>2 trained GPs reviewed SEA reports and criterion audit reports using a structured assessment instrument.</p> <p>Participants submitted a criterion audit or significant event analysis in standard formats.</p> <p>The quality of the reports submitted by participating GPs was the factor being assessed.</p>	<p>Of 1002 criterion audit reports, 55% were judged to be satisfactory to enable assessment by peer reviewers.</p> <p>Of 883 significant event analysis reports, 65% were judged as satisfactory to enable assessment by peer reviewers.</p> <p>Many participating GPs were unable to complete a SEA report of criterion audit</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
				report to a standard sufficient to enable peer review to occur.	
McKay 2007	Evaluation of adverse event reporting process using structured questionnaire.	UK. Peer review of GP's quality of their written analysis of a significant event (SEA).	Peer reviewers were trained to conduct peer review of SEA reports. 4 reviewers independently assessed 20 SEA reports using a structured assessment instrument. Inter-rater reliability was assessed.	The assessment instrument provided consistent information on the standard achieved by the significant event analysis report (G>0.7). Inter-rater reliability between the 4 assessors was moderate (G>0.6).	B
Forster 2007	Evaluation of independent peer review for assessing accuracy of ratings by reviewers.	Canada. Peer review of adverse events, to quantify the misclassification rate and evaluate the effect of combining physician adverse event ratings.	Three physicians independently rated adverse event reports. Estimates were used as a base case for four simulations of 10 000 cases rated independently by five reviewers.	Reviewer sensitivity and specificity for adverse event classification were 0.86 and 0.94 respectively. The positive predictive value of peer review increased with increased number of reviewers.	B
Nuckols 2008	Descriptive study of peer review of	USA.	Physician reviewers used implicit judgment to assign	Contributing factors to the occurrence of incidents were	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	incident reports.	Peer review of incident report narratives from an academic and a community hospital to assess factors contributing to the occurrence of the incident.	<p>contributing factor subcategories to each report.</p> <p>Reports were not validated against other data sources.</p> <p>Reviewers were first trained in peer review.</p> <p>Reviewers assigned subcategories of contributing factors they felt contributed to the occurrence of the incident.</p> <p>A single reviewer performed the initial assessment of the incident report.</p> <p>This was followed by a secondary assessment of 10% of the reports by 2 additional peer reviewers who performed their assessments independently.</p>	<p>identified by peer reviewers in 80% of incident-report narratives.</p> <p>Inter-rater reliability was low to moderate, depending on the factors being assessed (kappa = 0.41 – 0.72).</p>	
<i>ii. Peer review panels</i>					
Diaz 2008	Descriptive study of nursing peer	USA.	A peer review panel is convened to assess medical	Author provides description of application of the	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	review method.	A nursing peer review framework was developed. The steps in the peer review process are: referral process for peer review, assigning a facilitator, methods of data collection, presentation of data collected, documentation of peer review and communication of peer review.	<p>errors and the individual staff member's contribution to them.</p> <p>Subject of the peer review process prepares a written statement detailing information relevant to the clinical incident.</p> <p>The chair of the panel convenes the peer review session.</p> <p>The number of nurses providing peer review was not specified in the study.</p>	<p>framework to a case study.</p> <p>Primary focus of framework was to facilitate establishment of a culture of safety, eliminate medical errors, and reduce recurrence of untoward events.</p>	
Kernaghan 2006	Comparative study of methods for peer review of clinical care.	<p>UK.</p> <p>Peer review of quality of care delivered in cases where intrapartum adverse events have occurred.</p>	Case notes for 8 cases of perinatal death and neonatal encephalopathy due to intrapartum events were first assessed by the hospital's risk management group, who completed an assessment based on a standard template.	<p>Agreement among the panel was highest in domains where explicit standards were provided.</p> <p>Where subjective assessment of quality of care was made by panel members (without the use of pre-defined standards),</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>Clinical notes of the cases were then sent to an independent review panel (comprised of a neonatologist, neonatal nurse, obstetrician and midwife) for secondary assessment of quality of care.</p> <p>Case notes were sent to another independent review panel (comprised of same speciality mix as first independent panel) who repeated the assessment.</p> <p>Panels assessed quality of care against pre-defined standards of care (based on clinical guidelines where available) across three domains: admission assessment, recognition of incident, and method and timing of delivery.</p>	<p>agreement between panels was low.</p> <p>Agreement for the overall assessment was also low between all panels (kappa ranged from 0.1 to 0.36)</p>	
Rankin 2006	Qualitative study. Descriptive	UK. Peer review of clinical	Semi-structured interviews with 18 health professionals who had participated in at	Participants valued attendance at panels as a learning experience that	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	analysis of impact of participating in confidential enquiry panels.	practice relating to stillbirths and deaths in infancy.	least one confidential enquiry panel. Purposive sampling technique used.	provoked reflection on their own clinical practice.	
Stekelenburg 2002	Descriptive study of methods used to review maternal mortality.	Zambia. Peer review of clinical care where maternal death has occurred.	A maternal mortality review committee (a multidisciplinary peer review committee comprised of paediatric, obstetric and other hospital staff) peer reviewed patient files where maternal death had occurred. Causes of death, classification, substandard care factors, recommendations and implementation were analysed.	The maternal mortality review process identified substandard care factors and was educational for participating staff.	C
<i>iii. Review of charts</i>					
Hofer 2000	Evaluation of peer review of patient records for assessing	USA. Peer review of medical records to assess quality of	Structured implicit review instrument used to guide peer review assessment.	13 reviewers conducted 741 structured implicit reviews of 95 records.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	reliability of ratings by reviewers.	hospital care.	<p>Physicians independently reviewed medical records of patients who had experienced a severe adverse event related to laboratory abnormalities while in hospital.</p> <p>Degree to which event was caused by medical care, and quality of care leading up to the adverse event assessed.</p> <p>Physicians then worked in pairs to discuss differences of opinion and re-rate the record.</p>	Reliability of assessment increased marginally with discussion between reviewers compared with individual assessment (kappa increase from 0.14 to 0.17 for assessment of quality of care)	
Silver 2007	Comparative study of physician and nurse rating of adverse events.	<p>USA.</p> <p>Adverse event surveillance using chart audit of hospital discharge data was performed.</p>	<p>A physician reviewed case records for hospital discharges and coded the presence of an adverse event where it occurred.</p> <p>A nurse reviewer reviewed the same records and performed coding</p>	<p>1 035 records were independently reviewed by physicians and nurses.</p> <p>Agreement between physicians and nurses was between 68.2% and 78.5%.</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>independently.</p> <p>Reliability of nurse identification of adverse events was assessed through comparison with physician review.</p>		
b) Trauma-specific					
<i>i. Peer review of trauma deaths</i>					
Forsythe 2002	Comparative study of different trauma death data sources.	<p>USA.</p> <p>Peer review of trauma care for cases where patient death occurred.</p>	<p>Autopsy reports for all mortalities at a trauma centre were reviewed by trauma specialists.</p> <p>1 of 6 trauma physicians working at the facility performed a peer review of the patient chart and prepared a written report.</p> <p>The report was discussed at a monthly peer review meeting (a multidisciplinary meeting attended by trauma staff at the facility).</p>	<p>A total of 216 autopsies were reviewed.</p> <p>No errors were identified in autopsy data compared with other sources.</p> <p>Autopsy information added little useful information to peer review of deaths in a mature trauma program.</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>A peer review numeric score was then assigned estimating the preventability of the death.</p> <p>The autopsy results, peer review results, and trauma registry record were then compared to identify strengths and weaknesses of different data sources.</p>		
Jat 2004	Descriptive study of method for peer review of care outcomes.	<p>Pakistan.</p> <p>Peer review method for assessing preventability of death due to trauma.</p>	<p>A trauma peer review committee reviewed trauma deaths within the facility. The members of the committee included general surgeons, a neurosurgeon, an orthopaedic surgeon, an anaesthetist and an emergency physician.</p> <p>Trauma deaths were identified and registered in a computerized trauma registry by a third party (not part of the peer review process).</p> <p>Data including registry information and details of pre-</p>	<p>18 trauma deaths reviewed. The committee judged that 6 were preventable, 7 were potentially preventable and 4 were non-preventable (one excluded as records not available).</p> <p>Inadequate prehospital care, inappropriate interhospital transfer, limited hospital resources, and an absence of integrated and organized trauma care were identified as contributing factors.</p> <p>No reliability or validity</p>	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>hospital, emergency room and definitive care were provided to a peer review committee 1 week prior to the meeting.</p> <p>The committee then collectively reviewed all cases and classified each as preventable, potentially preventable or non-preventable.</p>	<p>measures of the peer review process were reported.</p>	
Martin 2007	Comparative study of different trauma death data sources.	<p>USA.</p> <p>Peer review of trauma death data.</p>	<p>Data from trauma admissions was entered into a trauma registry database.</p> <p>Deaths in trauma patients were identified from the database.</p> <p>Deaths were discussed in a monthly trauma peer review meeting – a multidisciplinary meeting of trauma staff within the facility.</p> <p>The participants in the</p>	<p>A total of 112 autopsies were reviewed.</p> <p>Autopsy data enhanced the peer review process for identifying preventable factors contributing to death in trauma patients experiencing immediate (dead on arrival) and /or late (death > 48 hours after arrival) deaths but not early (deaths between 0 and 48 hours after arrival) deaths.</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>meeting assessed various information sources: mortality type, injury severity scale, trauma and injury severity score-generated probability of survival, and findings at autopsy.</p> <p>From this, the peer review group made a judgment of the preventability of the death.</p>		
Shanti 2003	Comparative study of different trauma death data sources.	<p>USA.</p> <p>Peer review of trauma care for cases where patient death occurred.</p>	<p>Trauma admissions were entered into a trauma registry database.</p> <p>A "Probability of Survival Score" was calculated based on the features of the trauma admission.</p> <p>All deaths were reviewed by a peer review committee, comprised of surgeons involved in trauma care and the county medical examiner. The peer review committee assessed the preventability of</p>	<p>Data for 10 002 patients were reviewed.</p> <p>The peer review process was more sensitive in determining the preventability of death than other data source.</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>death and identified preventable factors that contributed to the death.</p> <p>The committee assessment of preventability of the death was compared with the probability of survival score.</p>		
Takanayagi 1998	Comparative study of different trauma death data sources.	<p>Japan.</p> <p>Peer review of trauma care for cases where patient death occurred.</p>	<p>Trauma deaths were evaluated independently by two expert peer review panels.</p> <p>Panel assessments of preventability of the death was compared with a trauma injury severity score to assess preventability of the death.</p>	<p>3125 patients were reviewed.</p> <p>Peer review was more sensitive in determining preventability of deaths compared with severity scoring.</p> <p>Agreement between panels was moderate (kappa = 0.62)</p>	B
<i>ii. Trauma morbidity</i>					
Jacobs 2006	Comparative study of methods for peer review of clinical	<p>USA.</p> <p>Peer review methods for judging trauma outcomes.</p>	A trauma outcomes committee performed multidisciplinary review of provider-related complications	Anonymously obtained outcome judgments were significantly less favourable than those obtained non-	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	performance in trauma care.		<p>and assessed the quality of care of patients who experienced adverse outcomes.</p> <p>Cases were selected at random for review from the hospital trauma database.</p> <p>The committee has 22 members from across trauma medical specialties.</p> <p>2 reviewers (both committee members) independently peer review the case records (using a standardised template) of cases selected for committee review prior to the meeting.</p> <p>30 months of non-anonymous peer review trauma care judgments were compared with 30 months of anonymous peer review judgments in this study.</p>	anonymously.	

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
c) Surgery-specific					
<i>i. Audit</i>					
Thompson 2005	Descriptive study. Description of quality assessment program provided.	UK. The Scottish Audit of Surgical Mortality. A review of patient data where surgical death had occurred over the 9 years of conduct of the audit was performed.	When a surgical death occurs the surgeon and anaesthetist involved in the care of the patient complete a report according to a structured proforma. The report is reviewed by a colleague from another geographical area and in the same field of practice. Feedback is provided to the surgeon and anaesthetist. If concerns are raised by the review, a second line peer review investigation is conducted by a peer in the same field of practice. The peer review case notes and completes an assessment with a structured assessment	Deaths after elective surgery had declined to 0.27% in the 9 years since the audit commenced. Fewer adverse events contributed to case of death over time ($p < 0.001$). Failure to offer prophylaxis for deep vein thrombosis and failure to use ICU / high dependency units appropriately became less common over the 9 years since the audit commenced.	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>tool.</p> <p>In this study, data from peer review audit, critical event analysis and individual feedback of deaths while in surgical care over 9 years (1994-2002) were examined for trends over time.</p>		
Semmens 2005	Descriptive study of paper-based audit of surgical mortality.	<p>Australia.</p> <p>Peer review of surgeons to assess surgical mortality outcomes.</p>	<p>When a surgical death occurs the surgeon involved in the care of the patient complete a report according to a structured proforma.</p> <p>The report is reviewed by a colleague in the same field of practice.</p> <p>Feedback is provided to the surgeon.</p> <p>If concerns are raised by the review, a second line peer review investigation is conducted by another peer in the same field of practice. The</p>	Changes in practice were noted in some clinical areas targeted by the audit process.	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>peer reviews case notes and completes an assessment with a structured assessment tool.</p> <p>194 surgeons participated after a patient died under their care.</p> <p>The audit process was completed for 1647 reported deaths.</p> <p>Preventability of death and deficiencies of care that contributed to deaths were assessed.</p>		
Allen 2002	Evaluation of survey-based peer review method.	<p>USA.</p> <p>Hospital-based surgeon peer review.</p> <p>Comparison of quality of surgical care between surgical peers.</p>	<p>Participating surgeons completed a proforma for routine surgical cases.</p> <p>Data regarding complication rates, patient education, resource utilization, use of diagnostic testing, and number of days before the patient returned to work were</p>	<p>11761 cases were entered into the database.</p> <p>Strengths and weaknesses of delivered surgical care, and comparison between surgeons, were enabled by this method.</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>collected.</p> <p>Data was entered into a database maintained by independent third party who analysed data and provided comparative feedback to participating surgeons.</p>		
Keyes 2004	Descriptive study of peer review of outpatient care.	<p>USA.</p> <p>Peer review process for assessing safety of surgery performed in ambulatory and outpatient settings was described.</p>	<p>Each surgeon is required to report all unanticipated sequelae and have at least 6 random cases, and the case notes for all patients where unanticipated sequelae occurred, reviewed by an accepted peer review process biannually.</p> <p>Peer review is performed by either a recognised peer review organization or by a physician other than the operating surgeon.</p>	<p>A total of 411 670 procedures, resulting in 2597 reported sequelae, 1378 significant sequelae and 7 deaths were reported.</p> <p>No reliability or validity measures of the peer review process reported.</p> <p>No effectiveness criteria of peer review were measured.</p>	C
Olcott 2000	Evaluation study	USA.	10 surgeons participated.	763 carotid reconstructions	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	of institutional peer review to influence surgical outcomes.	Peer review of case records for all patients undergoing carotid endarterectomy.	<p>A carotid quality assurance team was assembled, comprising registered nurses and the chief of vascular surgery.</p> <p>Patient records were assessed by the review team.</p> <p>Each surgeon received an annual morbidity and mortality rate report and the average morbidity and mortality rates for the institution.</p> <p>Results and trends were reviewed at an annual meeting of all participants.</p> <p>At this time, methods for improving care and controlling costs were also reviewed, and decisions were made regarding their implementation.</p>	<p>were performed over the 5 years of the study.</p> <p>The mortality rate, stroke rate, length of stay, and cost all decreased during the 5-year period.</p>	
<i>iii. Methodological</i>					

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Margo 2002	Evaluation of case review process using structured questionnaire.	USA. Peer review of ophthalmologists to assess physician error and to measure compliance with standard of care.	Case scenarios and questionnaires were mailed to 280 ophthalmologists. Ophthalmologists were asked to assess the case scenarios and provide answers on the questionnaire regarding their assessment of the occurrence of physician error and their views regarding the compliance of the physician in the case scenario with accepted standards of care.	Response rate was 35%. Unstructured implicit case review was not found to be a reliable method for determining physician error or for measuring compliance with standards of care. Kappa coefficients varied widely (0.55 – 0.83).	C
Silverstein 1998	Comparative study of physician rating for surgery with 5 year mortality data.	USA. Peer review assessment of appropriateness of abdominal aortic aneurysm (AAA) surgery.	9 expert panel members rated appropriateness of 120 cases for elective AAA surgery.	Peers disagree about the appropriate indications for elective surgery for AAA (kappa 0.07) Panel decisions correlated poorly with 5 year mortality outcomes (kappa 0.28).	B

TABLE 6– QUALITY ASSURANCE OF RADIOLOGY / PATHOLOGY PRACTICE

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Borgstede 2004	Evaluation study of case material review by peers	USA. Development and testing of a radiology peer review system to assist in re-certification of radiologists.	Peer radiologists reviewed images previously reported by another radiologist, and assessed the previous interpretations on a 4-point scale. Where the reviewing radiologist disagreed with the first radiologist's findings, and the diagnosis was felt to be one which should be made most or all of the time (i.e. an uncomplicated case), the case was referred to a peer review committee for assessment. The findings of peer reviews were entered onto a database to be interrogated when decisions regarding the re-certification (or re-credentialling) of the radiologist were to be made.	The scoring system, the subspecialty-specific sub categorisation of data collected for each imaging modality, and the validation of scoring consistency were issues limiting the utility of this peer review methodology.	C
Britton 2001	Comparative study of case material peer	UK. The classification of	Three peer review strategies for screening for interval breast cancers	The consensus approach achieved the greatest acceptability with radiologists	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	review.	interval cases of breast cancer by peer reviewers obtained through consensus, versus majority or single peer reviewer.	<p>were compared.</p> <p>1:Cases were read by 5 radiology peers and consensus reached before a case could be classified false negative.</p> <p>2:Cases read by 5 peers and if one or more peers classified the case as a false negative then the case was considered false negative.</p> <p>3:Cases read by 5 peers and classified false negative if the majority of radiologists rated it false negative.</p>	and the best outcome.	
Jolly 2001	Evaluation of method for peer review of reporting of results.	Australia and UK. Peer review of consultant radiologist's reports.	<p>Three trained and 6 non-trained peer reviewers (all consultant radiologists) were divided into groups of 3.</p> <p>Each group peer reviewed a selection of consultant radiologist's reports and the corresponding films.</p> <p>The group members independently reviewed the materials then compared their findings.</p>	<p>Large numbers of reports needed to be reviewed to identify examples of poor performance.</p> <p>Certain peer reviewers, irrespective of training and experience, rated certain reports unpredictably.</p> <p>Large samples of reports and multiple peer reviewers were</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
				required to ensure reliability of the peer review process.	
Bhatia 1998	Evaluation study of case material review by peers.	India. Pathology-based assessment of fine needle aspiration cytology (FNAC) reporting by peers.	Nine pathologists performed reporting and peer review of pathology reports for FNAC to confirm reporting accuracy.	A total of 4836 consecutive cases were subject to peer review, identifying 66 diagnostic errors and discrepancies. The method improved accuracy of diagnosis but significantly increased turnaround time for pathology reporting.	B
Zardawi 1998	Evaluation of peer review meetings for pathology quality assurance.	Australia. Peer review of surgical pathology.	Peer review meetings were attended by all the pathologists and registrars and by representatives of the scientific and technical staff. The peer review group examined all aspects of the case including patient demographics, typing errors, coding, adequacy of clinical history, technical quality and labelling of the histological slides, diagrams, macroscopic description, key to blocks, microscopic reports, and minor and major discordant diagnoses.	Peer review of 3530 surgical pathology cases occurred (10% of the pathology workload). Peer review identified errors in procedural and technical surgical pathology practice.	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			Turn around time (time from specimen collection to report completion), adequacy of specimen sampling, and use of special stains were also assessed.		

TABLE 7 – PEER REVIEW FOR THE PURPOSES OF CREDENTIALLING HEALTH CARE PROVIDERS

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
Borgstede 2004	Evaluation study of case material review by peers	USA. Development and testing of a radiology peer review system to assist in re-certification of radiologists.	<p>Peer radiologists reviewed images previously reported by another radiologist, and assessed the previous interpretations on a 4-point scale.</p> <p>Where the reviewing radiologist disagreed with the first radiologist's findings, and the diagnosis was felt to be one which should be made most or all of the time (i.e. an uncomplicated case), the case was referred to a peer review committee for assessment.</p> <p>The findings of peer reviews were entered onto a database to be interrogated when decisions regarding the re-certification (or re-credentialling) of the radiologist were to be made.</p>	<p>The scoring system, the subspecialty-specific sub categorisation of data collected for each imaging modality, and the validation of scoring consistency were issues limiting the utility of this peer review methodology.</p>	C
Johnstone 1999	Descriptive study of method for peer review of	USA. Peer review for	All radiation oncologists attended weekly chart rounds.	Peer review method commonly identifies inconsistent chart	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	quality of care.	<p>performance improvement and physician re-credentialing in a radiation oncology clinic.</p> <p>Program based on criteria established by the American College of Radiology.</p>	<p>Radiation oncologists reviewed each other's patient charts at meetings for laboratory and pathology reports, and adequacy of the doctor's documentation.</p> <p>Films are checked by a peer radiation oncologist.</p> <p>A checklist is then used to assess the patient chart for adequacy of care (against clinical practice guidelines) and a subjective evaluation of overall quality is made.</p> <p>Participants receive their feedback from the process monthly.</p> <p>The institution uses findings to inform credentialing processes.</p>	<p>documentation.</p> <p>Method used to compare care received with clinical practice standards.</p>	
Maidment et al. 2006	Evaluation of peer review method for dental practitioners.	<p>Scotland.</p> <p>Peer review of dental practitioners to assess fitness to practice and assist in revalidation process.</p>	<p>10 general dental practitioners completed portfolios of evidence of fitness to practice.</p> <p>Portfolios were assessed by a panel of three experts, against purpose-designed assessment tool.</p>	<p>Eight portfolios were assessed as sufficient for revalidation purposes. Two dentists were required to make supplementary submissions of evidence before revalidation recommended by</p>	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
				the panel. No reliability or validity measures reported. No effectiveness criteria measured.	

TABLE 8 – ASSESSMENT OF SUSPECTED UNDER-PERFORMANCE OF THE HEALTH CARE PROVIDER

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
<i>i. Multiple methods for peer review described</i>					
Agee 2007	Descriptive study of peer review group method to assess and manage issues related to physician performance.	USA. Hospital-based physician peer review.	<p>A Professional Review Committee conducted peer review. This was a peer review group composed of Chief Medical Officer of the facility, and 4 active physicians each with a 3 year term of appointment.</p> <p>Physician-related issues including clinical skills, conduct / behavioural complaints and documentation issues were assessed and managed by the Committee.</p> <p>Patient documentation, interviews with practitioners and review of administrative data performed.</p>	The Committee completed reviews of 112 cases. One third of reviews resulted in some level of action (education and / or performance management-related).	C
Kelly 2005	Descriptive study of external peer review of under-performance of	USA. Peer review of under-performance of doctors	<p>External peer review process for a group of 15 hospitals.</p> <p>A governing committee and</p>	83 charts from 14 different hospitals had been reviewed at time of publication.	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	doctors across specialties.	where a facility determines that external peer review is appropriate.	<p>review panel of physicians (70 members) was formed in the geographical area.</p> <p>Requests for review were received from hospitals, who provided the committee with 2 copies of the patient medical chart.</p> <p>Two reviewers were selected from the panel to assess the de-identified patient medical chart.</p> <p>The two reviewers extracted information from records according to a questionnaire – synopsis of the case, assignment of a severity index, specific routine questions, and comments.</p> <p>The reports from both reviewers were then peer reviewed and discussed by the governing committee.</p> <p>Inconsistencies between reviewers (if they occurred) were</p>	<p>The most commonly reviewed specialty was obstetrics.</p> <p>Reviewers and participants favourably viewed the peer review process.</p>	

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			<p>addressed by the governing committee members.</p> <p>A final reports was then sent back to the requesting hospital.</p>		
Lichtmacher 2008	Descriptive study of voluntary peer review method for assessment of areas of clinical concern.	<p>USA.</p> <p>Peer review of obstetric and gynaecology care, in the American College of Obstetrics and Gynaecology (ACOG) Voluntary Review of Quality of Care (VRQC) program.</p>	<p>Hospital makes a request to ACOG to perform a VRQC review.</p> <p>Hospital provides patient chart to ACOG and any background documentation required to undertake the review.</p> <p>A team of 3 peer reviewers is selected from a panel of trained reviewers by ACOG.</p> <p>The peer reviewers assess background information, conduct interviews with the relevant hospital leadership group, conduct a tour of the facility and evaluate its resources, perform confidential staff interviews, conduct an objective evaluation of the clinical records against pre-</p>	<p>Reasons for request for review commonly include organizational problems, problems with a specific physician or group of physicians, or desire to evaluate the hospital's obstetric and gynaecology performance against accepted standards of care.</p> <p>Problems commonly identified across each domain include: poor communication, poor documentation of clinical reasoning, lack of effective quality assurance program, and inadequate nurse staffing levels.</p>	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			determined criteria, and prepare a written report outlining findings and recommendations.		
Gluck 2003	<p>Descriptive study.</p> <p>Description of quality assessment program provided.</p>	<p>USA.</p> <p>American College of Obstetricians and Gynaecologists Voluntary Review of Quality of Care program.</p>	<p>Hospital makes a request to professional body to perform a peer review.</p> <p>Hospital provides patient chart to professional body and any background documentation required to undertake the review.</p> <p>A team of 3 peer reviewers is selected from a panel of trained reviewers.</p> <p>Peer reviewers assess 3 to 5 medical records on-site.</p> <p>Each record is reviewed independently, using a worksheet with explicit objective criteria based on clinical practice guidelines.</p> <p>Departmental and hospital personnel are also interviewed</p>	<p>Data from the first 100 site visits revealed departmental and systems deficiencies were more common triggers for review than clinical concerns.</p>	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
			on-site to obtain further information.		
Stumpf 2007	Descriptive study. Description of quality assessment program provided.	USA. American College of Obstetricians and Gynaecologists Voluntary Review of Quality of Care program.	3 peer reviewers assess 3 to 5 medical records on-site. Each record is reviewed independently, using a worksheet with explicit objective criteria based on clinical practice guidelines. Departmental and hospital personnel are also interviewed on-site to obtain further information.	The majority of deficiencies were in institutional systems rather than specifically obstetric or gynaecologic clinical problems. The types of deficiencies found appeared to be similar across the spectrum of institutions visited, and not related to the size of the institution or of the department of obstetrics and gynaecology, or the number of deliveries done annually. No reliability or validity measures reported. No effectiveness criteria measured.	C
Hutchinson 2001	Mixed methods study (face to face and	UK. Peer review methods for	Study assessed senior doctors' views of the General Medical Council's regulations on the	Barriers to resolving poor performance include the unwillingness of some	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	telephone interviews followed by survey)	assessing poor performance of doctors.	<p>handling of poor medical performance, including peer review methods used.</p> <p>16 interviews held face to face, followed by 28 telephone interviews and a survey of 457 physicians.</p>	<p>doctors to seek advice and the protective culture which prevents complaints being made against doctors.</p> <p>Senior consultants do not consider poor consultation skills as being of the same significance as poor technical skills.</p>	
Southgate 2001	Descriptive study of methods used to assess medical practitioner competence.	<p>UK.</p> <p>Peer review methods used to assess under-performance of GPs by written knowledge examination, observation of clinical practice with simulated + / - real patients, and demonstration of practical skills in a structured clinical examination.</p>	<p>Peer review of GP competence is conducted by: written examination, peer observation of a 10 patient simulated surgery, and an organised structured clinical examination (OSCE) conducted using peer GPs.</p> <p>30 volunteers of good standing, 27 practitioners referred to the General Medical Council for assessment, and 4 practitioners who were the focus of concern over their performance were assessed using the above method.</p>	<p>Knowledge, consulting skills and practical skills were assessed.</p> <p>Test reliability was high (Cronbach's alpha 0.61 to 0.88).</p>	B

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
<i>ii. Peer review by audit of patient documentation</i>					
Billett 2005	Evaluation of peer review of data sources for assessment of suspected under-performance.	UK. General practice-based assessment of GPs with higher than average patient mortality rates.	Retrospective analysis of routine mortality and patient registration data. A crude patient death rate was first calculated for each GP in the geographical region. Next, 2 GPs reviewed the clinical records of patients for whom a medical certificate of Cause of Death had been issued. The 2 GPs independent extracted data from the clinical record of the deceased patients using structured forms. Outcome measures included the reviewers' concordance with GP decisions to issue a death certificate.	A total of 257 deaths occurred. Case notes for 153 deceased patients were assessed. The method of investigation was time-consuming, costly, and of limited feasibility for primary-care organizations to introduce.	B
Liang 1999	Evaluation of accuracy of assessment of clinical cases by multiple peers	USA. Peer review of anaesthetic-related malpractice clinical	Participating anaesthetists independently reviewed clinical case notes for 12 malpractice case scenarios to predict the jury verdict (prospective assessment)	Agreement between anaesthetists was poor. Anaesthetists' predictions of jury verdicts were not	C

Study	Study type	Country and peer-review type	Methods	Summary findings	Quality score
	where malpractice is suspected.	scenarios.	and to assess whether the clinician agreed with the jury's verdict (retrospective assessment).	accurate.	
<i>iii. Assessment of frequency of formal processes of peer review committees</i>					
Spigelman 2003	Descriptive study to assess frequency of peer review activities occurring in an area health service.	Australia. Peer review activities for continuing performance assessment and management.	Two surveys, conducted over two consecutive years. Assessment of peer review committees' use of clinical indicators, management of issues arising, and whether meetings are multidisciplinary, chaired and points for action minuted and followed up.	105 and 109 units provided data over the two years respectively. Increases in formal arrangements for peer review committee proceedings were observed over time, including engagement of management, use of a designated chairperson, multidisciplinary participation, systems issues identified and addressed, and minutes of meetings kept.	B