



THE UNIVERSITY OF  
NEW SOUTH WALES



CENTRE FOR CLINICAL GOVERNANCE RESEARCH

# Peer review in medicine: a comprehensive review of the literature



**The Centre for Clinical Governance Research in Health,  
University of New South Wales**

*The Centre for Clinical Governance Research in Health undertakes strategic research, evaluations and research-based projects of national and international standing with a core interest to investigate health sector issues of policy, culture, systems, governance and leadership.*

First produced 2009 by the Centre for Clinical Governance Research in Health, Faculty of Medicine, University of New South Wales, Sydney, NSW 2052.

© Travaglia J, Debono D. 2009

This report is copyright. Apart from fair dealing for the purpose of private study, research, criticism or review, as permitted under the Copyright Act, 1968, no part of this publication may be reproduced by any process without the written permission of the copyright owners and the publisher.

National Library of Australia

Cataloguing-in-Publication data:

Title: Peer review in medicine: a comprehensive review of the literature

1. Literature review method.

2. I. Travaglia J, II. Debono, D. V. University of New South Wales. Centre for Clinical Governance Research in Health.

Centre for Clinical Governance Research, University of New South Wales, Sydney Australia

<http://clingov.med.unsw.edu.au>

## ***Peer review in medicine: a comprehensive review of the literature***

### **Duration of project**

February to April 2009

### **Search period**

1950 to April 2009

### **Key words searched**

- Peer
- Review
- Appraisal
- Assessment
- Evaluation
- Multi-source
- 360 degree
- Doctor
- Physician (various disciplines identified)

### **Databases searched**

- *Medline from 1950*
- *Embase from 1980*
- *CINAHL from 1982*

### **Criteria applied**

- Peer review
- Peer appraisal
- Peer assessment
- Peer evaluation
- Doctor or physician (various disciplines identified)

### **Contact details**

Contact: Jo Travaglia

Email: [j.travaglia@unsw.edu.au](mailto:j.travaglia@unsw.edu.au)

Telephone: 02 9385 2594

Centre for Clinical Governance Research in Health  
Faculty of Medicine  
University of New South Wales  
Sydney NSW 2052  
Australia

## CONTENTS

<b>1. Introduction</b> .....	<b>4</b>
<b>2. Background</b> .....	<b>4</b>
2.1 Definition of peer review .....	5
<b>3. Method</b> .....	<b>8</b>
3.1 Overview of method and research question .....	8
3.2 Review process.....	8
3.3 Search strategies .....	10
3.3.1 Search of databases .....	10
3.3.2 Hand search of journals .....	10
3.3.3 Search of grey literature.....	10
3.3.4 Snowball and citation tracking .....	10
3.4 Search findings .....	11
3.5 Analysis.....	13
3.5.1 Triangulated reviewer analysis .....	13
3.5.2 Concept analysis.....	13
<b>4. Findings and discussion</b> .....	<b>14</b>
4.1 Overview of concepts emerging from the peer review literature.....	14
4.2 Thematic analysis of peer review literature.....	19
4.2.1 The purpose of peer review .....	19
4.2.2 The peer review method .....	21
4.2.3 The role of reviewers .....	22
4.2.4 Peer review instruments and tools.....	23
4.2.5 The feedback process.....	26
4.2.6 Barriers to peer review and strategies to increase participation .....	28
4.3 The evidence base for peer review.....	29
4.4 Limitations of peer review .....	31
<b>5. Conclusion</b> .....	<b>34</b>
<b>6. References</b> .....	<b>35</b>
<b>7. Appendices</b> .....	<b>50</b>
Appendix A: Evidence sheet .....	51
Appendix B: Examples of peer review instruments.....	52
Appendix C: Advice given by HcAT to assessors providing feedback.....	57
Appendix D: Evidence base for selected validated peer review instruments.....	60
Appendix E: Bibliography and abstracts .....	87

## 1. INTRODUCTION

The Centre for Clinical Governance Research (CCGR) was asked by the Statewide Quality Branch in March 2009 to identify, review and synthesise evidence on a range of topics intended to support the *Understanding clinical practice toolkit*. This review is a comprehensive analysis of the literature on the use of peer review with and by doctors. Following reviews address issues of: clinical audit; morbidity and mortality reviews; case review; limited adverse occurrence screening; clinical indicators; and complaints and patient satisfaction.

The review uses the protocol for the rapid assessment, conceptualisation, and timely concise analysis of the literature [PRACTICAL],<sup>a</sup> developed by the CCGR. PRACTICAL emerged from CCGR's research in the fields of clinical governance, patient safety, interprofessionalism and accreditation amongst other areas.

In this review we present the results of a comprehensive review of the literature on *peer review in medicine*. The literature was identified using a combination of data searching, hand searching of journals and snowball technique. At the end of the review we provide abstracts and citations, arranged alphabetically by author, for the articles identified using the outlined search strategy.

## 2. BACKGROUND

Peer review is one of numerous performance measures and quality and safety strategies which can be undertaken alone, or as part of a wider framework, such as performance appraisal,<sup>1</sup> certification, validation, recertification or revalidation,<sup>1</sup> continuous professional development,<sup>2 3</sup> quality improvement and or clinical governance.<sup>4 5</sup> A relatively new phenomenon in healthcare,<sup>3 6</sup> interest in peer review as a way of evaluating doctors' performance has grown in recent years. This has been partly in response to high profile patient safety inquiries,<sup>7 8</sup> but also as a result of the growing recognition of the changing role of, and demands on, doctors.<sup>9 10</sup> Peer review has also come to be seen as a useful tool in the formalisation and systematisation of requirements for annual assessments of doctors, which in the past had been undertaken in "... *an ad hoc basis with wide variation in practice both between and within categories.*"<sup>11: 1089</sup>

In practice, certain types and elements of peer review, notably morbidity and mortality reviews (considered in a subsequent monograph in this series),<sup>11</sup> and clinical supervision,<sup>12</sup> have traditionally formed part of a medical practitioner's development program. Work on peer review measures at least in the undergraduate classroom has been undertaken in the US since the mid 1970s.<sup>13</sup> <sup>14</sup> The earliest reference to peer review identified through this literature review was a 1978 paper in the Australian and New Zealand Journal of Medicine reporting on a study in Royal Adelaide Hospital which had used peer review to

---

<sup>a</sup> Jeffrey Braithwaite had the idea of labelling the Centre's mode for reviewing literature 'PRACTICAL'. This monograph was written by Joanne Travaglia, Jeffrey Braithwaite and Deborah Debono

compare the quality of care in a generalist and specialist unit in Adelaide.<sup>15</sup> More recently, the Queensland Government Inquiry into Bundaberg Hospital<sup>16</sup> was particularly critical about the lack of peer review processes in the case of Jayant Patel. Ironically it had been an internal peer review within the Kaiser healthcare system in Portland, Oregon (Patel's previous workplace) that had resulted in restrictions being placed on his practice, forcing him to obtain a second opinion in those areas where he continued to be allowed to practice, and eventually led to his move to Australia.<sup>17</sup> The then head of the Australian Health Workforce Advisory Committee (AMWAC), Dr Jeannette Young, was quoted as saying that the "... *shortfalls in Dr Patel's clinical skills would have been identified and corrected if he had been a party to a peer review, where doctors monitor and correct each other's shortcomings.*"<sup>18</sup>

## 2.1 Definition(s) of peer review

There are a number of difficulties associated with the definition of peer review. The first is the term itself: peer review in the medical and allied health, most often refers most often to the peer-reviewing of manuscripts for publication or the awarding of grants. The second is that even when citations use the term "peer review" (and associated terms, as discussed below) to denote the appraisal of one colleague by another, they do not necessarily define exactly what is meant. The following quotation contains several of the most common strands within the literature, and provides insights into the elements of peer review:

*"Peer review is an organised effort whereby people critically appraise, systematically assess, monitor, make judgements, determine their strengths and weaknesses and review the quality of their practice, to provide evidence to use as the basis of recommendations by obtaining the opinion of their peers.*

*Peer review also offers staff an opportunity to both give and receive support, and to network across the site involved. A peer is a colleague of the same profession, grade or setting."*<sup>19: 428</sup>

This definition highlights the: *context* of peer review (organisational, as opposed to an individual act of self improvement); *purpose* of peer review (to critically and systematically appraise, assess and monitor); *focus* of peer review (strengths, weaknesses and quality); *outcomes* (evidence, and recommendations for approval); and *participants* (colleague of same profession, grade or setting: although not necessarily all three).

Several terms are either used interchangeably for, or in association with peer review. These include:

- peer assessment
- peer evaluation
- peer appraisal

- colleague review
- peer rating
- medical peer review
- multi-source assessment
- 360 degree feedback.

Each of these terms utilise the same elements: the evaluation of a doctor by colleagues. Variations include the type of colleague: other doctors, health professionals, co-workers, patients and the number of reviewers. Multi-source assessment refers to the use of more than one reviewer (without specifying which type of reviewer) while 360 degree feedback refers specifically to the use of a combination of self, peer and reviewer assessment. 360 degree feedback is often referred to in the literature as a type of multi-source feedback.

### 3. METHOD

#### 3.1 Overview of method and research question

We undertook a search of terms associated with peer review by medical practitioners using several databases, hand searches of key journals, using the snowball method, and via a search of the grey literature on websites associated with medical peer review. A preliminary search produced findings very similar to those of Evans, Elwyn and Edwards (2004), that is that peer review “...is an emerging area that does not have an extensive literature and is likely to be poorly indexed on electronic databases.”<sup>2:1240</sup> Like Evans et al we found that searching for “peer” and “review” identified too large a volume of irrelevant references (false positives) largely relating to the peer review of literature, and so searched for alternative proxy combinations. The original list of search terms used in this review is presented in Table 1.

**Table 1: List of proxy search terms for peer review of medical practitioners**

Search terms
1. peer
2. review
3. doctor\$
4. physician\$
5. appraisal
6. evaluation
7. assess\$
8. exp hospital physician/ or exp medical staff/ or exp resident/ or exp medical personnel/ or exp coroner/ or exp medical expert/ or exp medical specialist/ or resident/ or exp medical student/ or physician/ or exp anesthetist/ or exp cardiologist/ or exp emergency physician/ or exp female physician/ or exp general practitioner/ or exp gerontologist/ or exp nephrologist/ or exp occupational physician/ or exp pathologist/ or exp pediatrician/ or exp psychiatrist/ or exp radiologist/ or exp surgeon.
9. Multi-source
10.360 degree
11. Feedback

#### 3.2 Review process

We utilised a five phase review process, as outlined in Figure 1. Phase one involved establishing the review parameters, as required by Statewide Quality Branch. The search was limited to peer review for medical practitioners and was to consider instruments as well as evidence for the effectiveness of the tool.

Phase two was the search itself, which involved identifying literature and resources associated with peer review from four different, but overlapping sources: databases; key journals; grey literature; and through snowball technique and citation tracking.

Phase three was the screening of the literature. This involved removing any

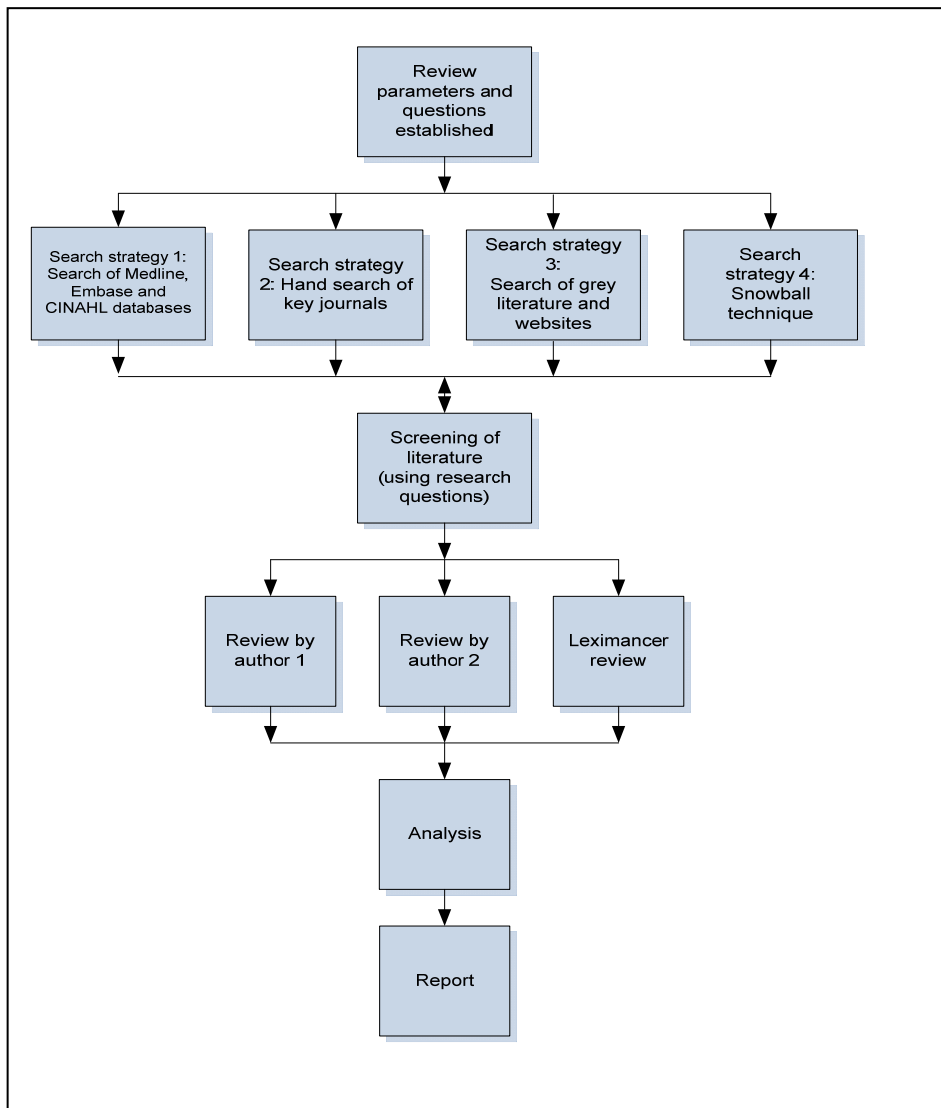


extraneous, inappropriate or incomplete references. As this is a targeted review, only directly relevant references were included.

Phase four was the review of the literature. Research articles were noted, as were articles which provided examples of peer review tools or instruments. The remaining articles were examined by two reviewers, and the abstracts subjected to data-mining in order to identify the key concepts.

Phases five and six were the analysis of findings, and the writing of this report. As mentioned previously, this is a relatively new area, with limited direct research evidence. As a result, a significant amount of data for this review were derived from the grey literature, and in particular from governmental and non-profit organisations established to co-ordinate peer review.

Figure 1: Review process



### **3.3 Search strategies**

#### **3.3.1 Search of databases**

The first level of our search strategy was to use the terms indicated in Table 1 to interrogate three databases: Medline, EMBASE (medicine) and CINAHL (nursing and allied health). In the first analysis we limited our results to references those relating to human subjects, and those pertaining to medical practitioners (doctors, physicians and various specialties).

#### **3.3.2 Hand search of journals**

We then hand-searched key journals for similar terms relating to peer review. The journals searched included:

- Quality and Safety in Health Care
- International Journal of Quality in Health Care
- Medical Education
- Medical Teacher
- Journal of Evaluation of Clinical Practice
- Academic Medicine
- American Journal of Medical Quality.

#### **3.3.3 Search of grey literature**

Our third strategy was to examine the grey literature. This proved particularly useful in the identification of peer review instruments. Amongst the websites reviewed were:

- The Australian Health Workforce Taskforce
- Department of Human Services (Victoria)
- Departments of Health in each state and capital Territory in Australia
- Agency for Healthcare Research and Quality (United States)
- National Health Service (United Kingdom)
- Physician Assessment Review website (Canada)
- Healthcare Assessment and Training (United Kingdom).

#### **3.3.4 Snowball technique and citation tracking**

Our final strategy was to “snowball”. That is, to follow up on any additionally, previously un-identified references in the bibliographies or reference list of articles

reviewed, or as listed on websites.

### 3.4 Search Findings

We present our search findings in Table 2. These include all findings from our database searches up to, and including, the removal of duplicates. Once the references were identified they were downloaded into Endnote X2, a citation manager.

**Table 2: Search findings for selected databases**

	Medline 1950 to April Week 1 2009	Medline in process & other non- indexed citations April 10, 2009	EMBASE	CINAHL 1981 – 2009 (Includes pre- CINAHL)	Total
1. peer and review and physician\$	1620	22	1084	326	3052
2. peer and review and doctor\$	257	4	340	65	666
3. peer and appraisal and physician\$	34	0	26	11	71
4. peer and appraisal and doctor\$	11	1	15	6	33
5. peer and evaluation and physician\$	351	10	193	131	685
6. peer and evaluation and doctor\$	78	1	63	46	188
7. peer and assess\$ and physician\$	553	17	382	118	1070
8. peer and assess\$ and doctor\$	116	5	134	42	297
9. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8	1456	38	1456	465	3415
10. peer and review	14879	5	13018	3485	31387
11. peer and appraisal	356	466	165	204	1191
12. peer and evaluation	3495	13	1837	1974	7319
13. peer and assess\$	6292	136	4622	1852	12902
14. 9 or 10 or 11 or 12	19650	328	15675	5564	41217
15. exp hospital physician/ or exp medical staff/ or exp resident/ or	166654	708	183113	50899	401374

	Medline 1950 to April Week 1 2009	Medline in process & other non- indexed citations April 10, 2009	EMBASE	CINAHL 1981 – 2009 (Includes pre- CINAHL)	Total
exp medical personnel/ or exp coroner/ or exp medical expert/ or exp medical specialist/ or resident/ or exp medical student/ or physician/ or exp anesthetist/ or exp cardiologist/ or exp emergency physician/ or exp female physician/ or exp general practitioner/ or exp gerontologist/ or exp nephrologist/ or exp occupational physician/ or exp pathologist/ or exp pediatrician/ or exp psychiatrist/ or exp radiologist/ or exp surgeon/					
16. 13 and 14	1367	0	1700	257	3324
17. 9 or 16	2781	38	2488	516	5823
18. Multi-source feedback	16	4	14	12	46
19. 360 degree feedback	23	0	7	15	45
20. 360 degree appraisal	2	0	3	2	7
21. 360 degree evaluation	16	1	11	6	34
22. 360 degree assessment	24	2	11	7	44
23. 17 or 18 or 19 or 20 or 21 or 22	2857	45	2531	557	5990
24. Total minus duplicates					4815

While literature base on commentaries about peer review is substantial, the evidence base, particularly in the context of clinical practice, is comparatively

small.<sup>19</sup> On examination of the findings, and in line with the purpose of the review, we removed: citations relating to the peer review of literature; citations relating to the peer review of documents; citations that were incomplete and could not be verified (i.e. were missing information such as author or source, and which could not be identified elsewhere); citations to legal issues or cases United States which were outdated or had no direct bearing on the issue in Australia were removed; citations which addressed peer support or counselling rather than peer review; and citations relating to peer support of patients (such as in self management of chronic diseases).

### **3.5 Analysis**

#### **3.5.1 Triangulated reviewer analysis**

Once the preliminary screening and review was completed, thematic content analysis of the citations was undertaken. As the data was analysed, studies of major peer review instruments were extracted and are presented separately in Appendix D.

Two independent reviewers were used to analyse the evidence. A discussion of the similarities or differences of categorisation of the literature was undertaken until agreement was reached. This step, along with the data mining of the literature, reduces the amount of subjective bias in the analysis of evidence.

#### **3.5.2 Concept analysis**

The citations and abstracts relating to peer review were mined using Leximancer, a computerised content analysis tool. A conceptual map which summarises the key concepts in the literature (Figure 2), and a ranked list of concepts from emerging from the citations (Table 3) are produced in the next section of this document.

## 4. FINDINGS AND DISCUSSION

Peer review can take place on an individual level (that is colleague to colleague) or utilise multiple reviewers. Some forms of peer review involve self-assessment components, while others do not. Peer review processes can involve structured instruments in a formal, or simply require colleagues to provide an 'impression' of their peer's clinical ability, or a combination of both.

Peer review activities of various kinds are currently conducted as part of undergraduate, immediate postgraduate, or continuing professional assessment and development in Australia, and around the world, notably Australia, Canada, New Zealand, the United States, the United Kingdom. Reviews are undertaken by individual doctors for their appraisal purposes, through professional bodies and colleges, educational institutions, healthcare services or systems. They are conducted using observation and or interview, self assessment, 'paper tools' and document analysis, or increasingly, online. Peer review processes and instruments draw on research into feedback, performance monitoring and management, audit, and adult education in both medical and wider context, and as such may not be labelled 'peer review', but rather feedback, monitoring, assessment or evaluation.

As an activity, peer review can be relatively short (taking an hour or less in terms of self assessments), or as in the case of overseas trained medical practitioners in Australia, can last for a year.<sup>20</sup> Peer review can be used to monitor, support and develop a doctor's existing abilities, or identify and address discrepancies or shortfalls in doctor(s)' performance.

In the following sections we will explore these, and additional aspects of peer review in medicine. In the next section, 3.1, we present the findings from our analysis of the peer review literature we present the outcomes of the data-mining of citations and abstracts. In section 3.2 we discuss our findings based on our thematic content analysis of the literature. In section 3.3 we consider the evidence base, and in section 3.4 the known limitations, of peer review.

### 4.1 Overview of concepts emerging from the peer review literature

All of the abstracts identified in the literature review (minus duplicates) were subjected to a data mining process, using the Leximancer software program. This was undertaken for two purposes. Firstly, it provides an objective mapping of the key concepts within the literature, and therefore the published field. Secondly, it allows for triangulation with the traditional content analysis method used by the reviewers, in their analysis of the literature.

The concept map of peer review is presented in Figure 2. The key themes in the literature are depicted by circles. These include: *care, clinical, peer, patient and patients, study, education, management, training, students, significant, included, cases, rate* and *surgery*. The themes speak to the use of peer review by

educational bodies and health services in response to public demand for monitoring of quality and safety of care, and professional requirements for performance development.

Within each theme individual concepts are identified. Concepts are not the equivalent of terms, that is to say, the concept *patient* may include *patients*, *consumers*, *clients*, *users*, as well as the word “patient”. In this way, Leximancer provides an analysis of the underlying argument and structure of the literature, rather than a simple word count. Concepts are depicted by the small dots within circles. Location is significant: concepts that are closer together have a stronger association in the literature.

Figure 2: Concept map of key concepts relating to peer review in medicine

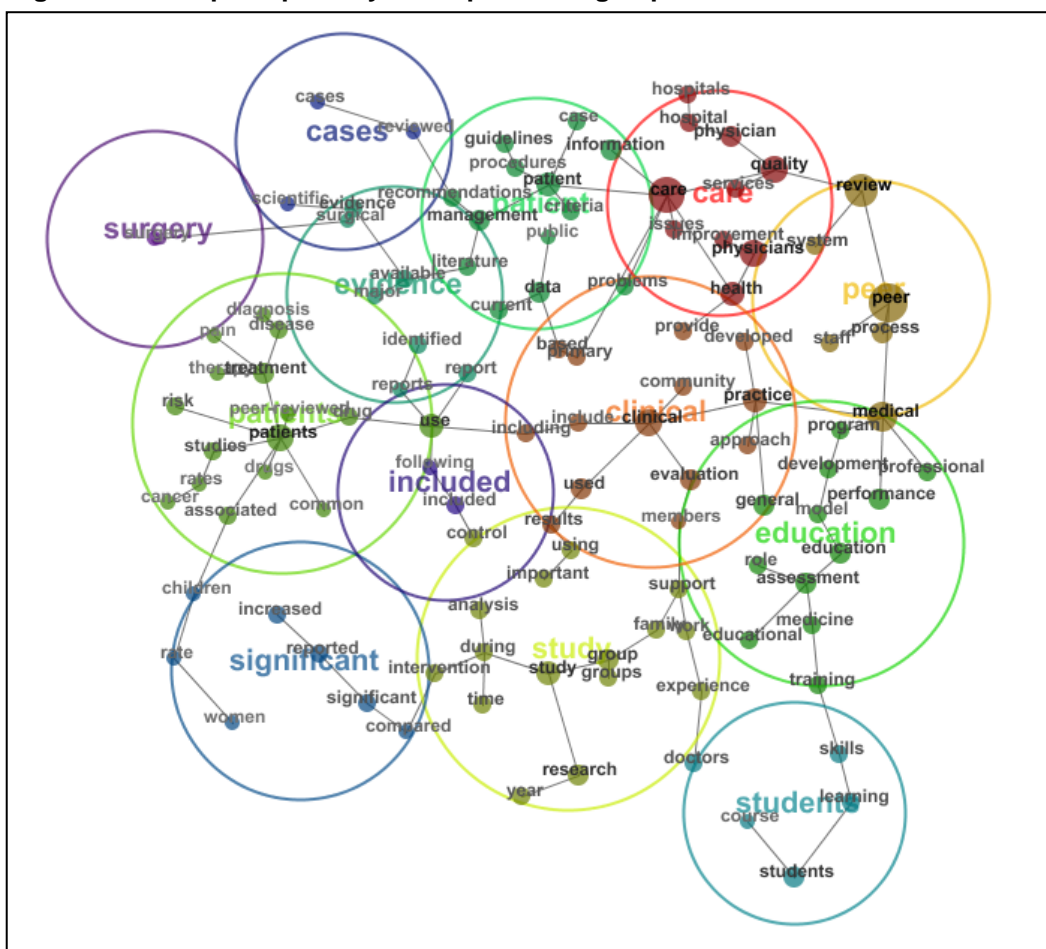


Table 3 below provides a ranked list of these concepts. The list provides insights into the relationships of concepts with each other, and the overall importance of concepts in the literature. As is clearly presented on the map, the central concerns of the literature are the impact of peer review on the *quality* of *care*, through its use as a method of *performance assessment* and *development*. The close ties in the literature between research into peer review at undergraduate level as a teaching tool, and as a method of monitoring for professional development is

evident both in the concept map and in the ranked list of concepts presented in Table 3, below. Relevance indicates the strength of association between a concept and the other concepts on the list: *peer*, for example, has a direct association with 100% of all other concepts (as would be expected), but the concept of *quality* is only associated with 51% of other concepts, indicating that the literature is concerned with other issues as well as quality, for example *process, education, recommendations* and *models*.

**Table 3: Ranked map of key concepts relating to peer review in medicine**

Concept	Count	Relevance
peer	2888	100%
review	2565	89%
care	2314	80%
medical	2000	69%
patients	1604	56%
physicians	1531	53%
quality	1465	51%
clinical	1405	49%
practice	1354	47%
health	1236	43%
study	1068	37%
patient	1039	36%
use	1023	35%
research	957	33%
physician	906	31%
students	886	31%
data	854	30%
process	791	27%
group	767	27%
education	766	27%
treatment	750	26%
assessment	745	26%
management	744	26%
performance	736	25%
information	731	25%
evaluation	713	25%
used	713	25%
general	703	24%
program	652	23%
system	606	21%
development	592	20%
guidelines	581	20%
training	576	20%
groups	575	20%



Concept	Count	Relevance
hospital	573	20%
results	562	19%
support	539	19%
evidence	531	18%
professional	522	18%
educational	518	18%
primary	514	18%
using	514	18%
skills	507	18%
medicine	506	18%
learning	503	17%
based	501	17%
time	501	17%
studies	500	17%
provide	493	17%
improvement	489	17%
important	477	17%
significant	475	16%
problems	474	16%
analysis	466	16%
intervention	453	16%
during	447	15%
staff	444	15%
doctors	441	15%
role	439	15%
year	436	15%
experience	431	15%
risk	430	15%
developed	421	15%
issues	415	14%
services	413	14%
approach	406	14%
including	404	14%
case	400	14%
control	397	14%
criteria	393	14%
hospitals	390	14%
reported	390	14%
available	385	13%
recommendations	384	13%
work	377	13%
increased	377	13%

Concept	Count	Relevance
literature	374	13%
associated	373	13%
family	368	13%
community	362	13%
model	355	12%
current	351	12%
report	347	12%
identified	344	12%
cases	342	12%
included	340	12%
procedures	338	12%
disease	337	12%
compared	331	11%
include	329	11%
drug	316	11%
surgical	309	11%
reports	308	11%
rate	294	10%
course	282	10%
members	280	10%
major	277	10%
rates	275	10%
common	269	9%
reviewed	264	9%
children	256	9%
therapy	248	9%
scientific	248	9%
public	237	8%
following	235	8%
cancer	232	8%
women	232	8%
diagnosis	230	8%
peer-reviewed	226	8%
surgery	211	7%
drugs	195	7%
pain	185	6%

The map and ranked list of concepts highlight the key issues in the peer review literature: the study and validation of the peer review process; the methods used; its use in educational and clinical settings; and the impact of peer review on the quality of care in general, on particularly strategies such as the use of guidelines and on patients.

## 4.2 Thematic analysis of peer review literature

Once we had reviewed the findings from the data mining of the literature, and had undertaken our preliminary review of citations, we established five key organising themes. These are presented in Table 4, below.

**Table 4: Categories identified in the literature on peer review in medicine**

Category
1. Purpose
2. Method(s)
3. Role of reviewers
4. Instruments and tools
5. Feedback
6. Barriers to, and improvement strategies for, peer review

### 4.2.1 The purpose of peer review

Peer review is intended to “offer individuals insight into the way others perceive their performance based on their workplace behaviour and provide an opportunity to reflect on one’s own conduct.”<sup>3:23</sup> From the perspective of professional bodies, peer review has been described as one of the approaches aimed at “achieving and delivering optimal quality of care ... [which] ... requires continual self-examination by the profession, particularly with regard to technical, interpretive, and communicative skills.”<sup>3:23</sup> Central to this process is the gathering of empirical data on the quality of, sources of variability and underlying causes of error, in professional practice.<sup>21</sup> It is said to address the assessment of a range of skills and attributes such as communication, teamwork and interprofessional skills that cannot easily be inferred by traditional review methods like examinations.<sup>22 23</sup> It is equally useful in identifying the difference between what Miller calls cognition or competence (knows how) and behaviour or performance (does).<sup>1 24</sup> It is also increasingly being used to determine if a practitioner or student is displaying “professionalism”.<sup>25 26 27 28</sup>

Peer review methods allow doctors to compare their self-perceptions of their skills and abilities with that of their colleagues. It can be used as a part of a standard performance assessment or review particularly early in the doctors’ career, including at intern and registrar level,<sup>29</sup> as an element in the certification or recertification (or validation and revalidation) process,<sup>30 31</sup> to ensure the continuing competence of clinicians,<sup>32</sup> as part of larger quality improvement and educational strategies,<sup>21 33</sup> including supporting the development of and evaluating adherence to, evidence based clinical guidelines.<sup>21 34-43</sup> It has also developed as a way of identifying and or addressing doctors experiencing difficulty, including those who deviate from standard procedures.<sup>44</sup> In the United States (US) there have been suggestions that peer review could be used as an alternative method to peer witnesses in assessing the standard of care in malpractice cases.<sup>45</sup> Although it

can and is used as a way of identifying and controlling under-performing practitioners,<sup>6 44 46</sup> peer review is currently most often employed (outside of Medical Board procedures<sup>47-49</sup> and for the management of what has recently been term “disruptive physicians”<sup>50 51</sup>) as a type of formative assessment, that is, one which contributes to the individual’s professional development, rather than as a summative (pass-fail) type grade.<sup>52 53 54</sup> It is increasingly used as a way of teaching and assessing students in medical programmes.<sup>55 56 57</sup>

Peer review is used by other professions, including nursing,<sup>58 59</sup> physicians assistants,<sup>57</sup> radiographers,<sup>10</sup> hospital administrators<sup>60</sup> and in workforce planning.<sup>61</sup> There is an increasing interprofessional component in peer review (that is, with members of other and non-health professions acting as reviewers for doctors and visa versa), an issue which will be taken up in the next section.

Over and above its use as a collegial performance and appraisal tool, peer review has been used by health systems as a way of controlling workforce planning,<sup>61</sup> finances and quality of care. In the US, it has been used as a control mechanism for regulating charges,<sup>62</sup> and as a way of monitoring hospital use and quality of care provided to (US) Medicare recipients,<sup>63-65</sup> including the control of hospital utilisation<sup>66 67</sup> and the reduction of “*inappropriate and unnecessary [hospital] admissions.*”<sup>68</sup> (Utilization and Quality Control) Peer Review Organizations (PROs)<sup>69 70</sup> were charged in the early 1980s with providing this mechanism. In 2002 they were transformed into Quality Improvement Organisations (QIOs), which were given expanded powers to review nursing homes, home health services, and doctors' offices as well as a responsibility for providing public education.<sup>71</sup>

PROs used a team of physician and nurse peer reviewer to review a random sample of hospital records, and using local clinical criteria, make judgements about the medical necessity of the inpatient services and length of stay, and the quality of care delivered.<sup>68 72-74</sup> There is some evidence for high inter-rater reliability using the PRO method.<sup>75</sup> While PROs were actively engaged in quality improvement strategies,<sup>76-78</sup> including gathering data about practice patterns and suggesting educational strategies,<sup>78</sup> because of their monitoring function, which included the power to stop clinicians from practicing and to cut funding to services,<sup>62 79</sup> they were described as “*adversarial and punitive.*”<sup>80</sup> As a result, peer review is a much more contested concept in the US, and one which has generated significant literature relating its legal ramifications, including contestations of the accuracy of the findings and reporting of PROs<sup>63 68 72 81-83</sup> and legal challenges by doctors and other clinicians, although there is some preliminary evidence that seems to suggest that these have reduced in the case of QIOs.<sup>80</sup> In addition to these issues, the US literature is also concerned with legal issues such as the confidentiality of peer review documentation and process,<sup>84 84-87</sup> and the role of doctors as expert witnesses in legal cases.<sup>88-90</sup>

Professional bodies utilise peer review in two distinct ways. At one level, Medical Boards, who have the power to bar a clinician from practicing, utilise what is essentially a peer review process. The work of these Boards is, however, outside

of the scope of this toolkit. In addition to such Boards, professional bodies use peer review as a method: of establishing and reviewing guidelines for practice;<sup>91</sup> for determining parameters for the quality and safety of specific procedures and care in general;<sup>92</sup> and as an mechanism for ensuring ethical practice.<sup>93 94</sup> Some professional bodies, such as the Royal Australian and New Zealand College Radiologists (RANZCR), for example, have their own audit tool, which is updated regularly.<sup>95</sup>

Peer review is has traditionally been used in some countries (Belgium, France, the Netherlands) as part of the credentialing process.<sup>96</sup> In the United Kingdom, where there have been suggestions that hospitals be allowed to assess doctors for re-licensure, an early study has shown that peer review is a feasible method for credentialing and re-credentialing doctors. The study noted, however, that the use of peer reviewer as a licensing tool will require health systems to ensure participant confidentiality, as well as the development of new systems and supports for patient feedback, and new instruments for non-clinical specialities.<sup>97</sup>

#### 4.2.2 The peer review method

Peer review can involve a combination of any one of several elements, depending on the review's purpose, process and participants. The process can be driven by one or more of the following: a standardised, tool-driven (questionnaire, case analysis) assessment of the physician;<sup>98</sup> a particular issue or intervention (e.g. the take up of evidence based guidelines for diabetes);<sup>58</sup> through a peer comparison of clinician data (number of complications, days spent in hospital by patients, mortality);<sup>99 100</sup> significant event analysis;<sup>98 101 102</sup> observation (including video-taping of consultations);<sup>26 98 101 103-105</sup> through record, case note or chart review;<sup>106 101 107 108 109 110</sup> objective structured clinical examination;<sup>98</sup> practice visits;<sup>111-115</sup> and or patient feedback or assessment.<sup>29 101</sup> Ongoing peer review/peer support groups have been used extensively in psychiatry,<sup>116 117</sup> but are gaining popularity in areas such as general practice.<sup>118-122</sup> The peer review process generally includes one or more of the following:<sup>1 3 106 123-125</sup>

- Identification of the doctor's strengths and weaknesses by the doctor themselves and their peers
- A comparison of these strengths and weaknesses with an 'average' colleague in their peer group
- Identification of areas which require development
- Creation of an agreed upon development plan to address these areas.

Reviewers may be asked to focus on very specific issues, or to provide a broad "impression" of their colleagues' abilities.<sup>126</sup> Assessment of very specific technical and or non-technical aspects of doctor's performance have included, for example, a doctors ability to undertake structured tasks such as writing of outpatient consultation letters,<sup>101 127 128</sup> or their decision making processes.<sup>15</sup> Reviewers can be asked, or ask, to examine evidence such as case notes and patient records.<sup>106</sup>

Comparative measures have included: patient education; use of guidelines; complication rates; use of diagnostics; resource utilization; and length of stay.<sup>106</sup>

37

It is important to note that while 'peer review' remains the broad term for this process of colleague based feedback, newer nomenclature, interpretation and tools are emerging, largely imported from the broader fields of human resources, organisational and business studies, and organisational psychology. Notable amongst these is the use of 360<sup>0</sup> feedback<sup>25</sup> and multi-source feedback (MSF) or multi-rater feedback (MRF).<sup>129 130 131</sup> These terms were added to the original search for this toolkit only after the original search was conducted: in reviewing the literature it became obvious that these terms, while not supplanting peer review, were a more current variant to this field.

#### 4.2.3 The role of reviewers

In most peer reviews the doctor being assessed is asked to nominate a number of people who will act as (from four to 11 in the literature) reviewers.<sup>3 29 132 133</sup> The doctor then asks the individuals to participate. Reviewers can include medical peers, senior medical practitioners, managers, non-medical co-workers, and or patients.<sup>29 131</sup> There has been some debate as to whether reviewers should be limited to colleagues of the doctor under review, that is internal members of the practice or service, or whether trained external reviewers should be used<sup>31 134</sup> in order to limit the conscious or subconscious influence of existing relationships on the assessment process. Some services have established 'standing' peer review committees "*... of physicians who meet bi-weekly to discuss concerns including clinical issues, conduct/behavioral complaints, documentation issues, individual data and any related physician concerns involving patient care.*"<sup>125: 52</sup>

Most peer review processes now require or advise that the person to be reviewed nominate individuals from a range of professional backgrounds. Whether patients are able to be reviewers depends on the organising body. In the case of the Canadian Physician Achievement Review (PAR), patients can be assessors, whereas with the mini-PAT (mini-Peer Assessment Tool) they are generally not. However, the UK team which developed mini-PAT (and the full Sheffield Peer Assessment Review Tool – SPRAT) have recently been developed and validated a specific patient review tool, the Sheffield Patient Assessment Tool (SHEFFPAT) which enables patients to provide specific feedback about doctors' performance. One variation on the direct involvement of patients in peer review has been the use of consumer feedback panels or forums to provide feedback to the profession as a whole.<sup>135</sup> This form of patient involvement in quality improvement will be discussed in greater length in a related monograph.

Inconsistencies in reviewers' assessments have also been noted. Research in the US, where Peer Review Organisations (PROs) used teams of nurse – doctor reviewers, found evidence of discrepancies between explicit and implicit review and definitions of complications and quality between doctors and nurses.<sup>136</sup>

Once reviewers have agreed to participate they are either sent an assessment form or given access to one on the internet, along with detailed explanatory notes from the organising body. The person under review is asked to complete a self-assessment during the same review period. Once all assessments have been completed, the reviewers' assessments are compared to that of the doctor involved, and the doctor's peer review to that of their doctor's peer group. The results are then sent back to the doctor and or their supervisor.<sup>3 132 137</sup>

#### 4.2.4 Peer review instruments and tools

Davies and Howell (2004) argue that in order for peer review to be effective as an assessment process (and be able to withstand, for example, court challenges), the process and its associated tools need to be reliable, valid, feasible, cost effective and have an educational impact.<sup>1</sup> These concepts provide an effective framework (Table 5) for considering the key issues associated with the peer review process and its related tools, and the issues will be taken up in turn in the next sections.

**Table 5: Characteristics of effective assessment instruments (modified from Davies and Howell (2004))**

Characteristic	Issues
Reliability	<ul style="list-style-type: none"> <li>• Is the instrument robust enough to minimise variations between reviewers (inter-reviewer reliability)?</li> <li>• Is it robust enough to minimise variations within reviewers (intra-reviewer reliability)?</li> <li>• Is the instrument itself reliable, that is, are its findings reproducible (test-retest reliability)?</li> <li>• Is the nature of the case/problem presented to the doctor (case specificity) generalisable?</li> <li>• Is the peer review process robust enough in terms of number of assessments and reviewers to overcome reviewer variation?</li> </ul>
Validity	<ul style="list-style-type: none"> <li>• How well does the instrument measure what it says it measures?</li> <li>• Does the instrument have construct validity? That is, does the tool measure or correlate with ability being reviewed? Do the questions intended to find out about the doctor's ability to communicate with their patients, for example, actually measure that skill?</li> <li>• Does the instrument have content validity? That is, does it include all aspects of the ability being reviewed? Do the questions cover all relevant aspect of communicative ability with the context under review?</li> <li>• Does the instrument have criterion validity? That is, is the instrument an accurate measure of outcomes in the 'real' world?</li> </ul>
Feasibility	<ul style="list-style-type: none"> <li>• Can the instrument be used in different settings, and with different scales of participants?</li> </ul>
Educational impact	<ul style="list-style-type: none"> <li>• Is the instrument capable of generating feedback that can inform professional development?</li> </ul>

Despite the breadth of the field there are relatively few validated peer review instruments for medical practitioners (as opposed, for example, to general feedback instruments).<sup>1 138</sup> The most highly recognised medical peer review (as opposed to feedback or remedial) tools are the Canadian Physician Achievement Review (PAR)<sup>137</sup> and the UK SPRAT<sup>2 139</sup> and their variants. Validated tools have also been created by radiologists in the US, Australia and New Zealand<sup>140-142</sup> in obstetrics and gynaecology,<sup>143</sup> and by UK's General Medical Council, amongst others.<sup>133</sup>

SPRAT was originally developed as a way of assessing good clinical performance as outlined in Good Medical Practice and by the Royal College of Paediatrics and Child Health. The questionnaire was tested in two pilot studies in the Sheffield Children's Hospital, and then modified. The core SPRAT assesses the doctor within five domains of good medical practice: 1. good clinical care; 2. maintaining good medical practice; 3. teaching and training, assessing and appraising; 4. relationships with patients; and 5. working with colleagues.<sup>139</sup> Participants are rated from 1 (very poor) to 6 (very good) in 24 areas (Box 1). Free text space is provided in the questionnaire for observations and examples.

**Box 1: SPRAT questionnaire<sup>139</sup>**

1. Ability to diagnose patient problems	13. Commitment to learning
2. Ability to formulate appropriate management plans	14. Willingness and effectiveness when teaching/training colleagues
3. Ability to manage complex patients	15. Ability to give feedback (private, honest and supportive)
4. Awareness of their own limitations	16. Communication with patients
5. Ability to respond to psychosocial aspects of illness	17. Communication with carers and/or family
6. Appropriate utilisation of resources, eg ordering investigations	18. Respect for patients and their right to confidentiality
7. Ability to assess risks and benefits when treating patients	19. Verbal communication with colleagues
8. Ability to co-ordinate patient care	20. Written communication with colleagues
9. Technical skills (appropriate to current practice)	21. Ability to recognise and value the contribution of others
10. Ability to apply up to date/evidence based medicine	22. Accessibility
11. Ability to manage time effectively/prioritise	23. Leadership skills
12. Ability to deal with stress	24. Management skills

The UK National Health Service (NHS) has established a Healthcare Assessment and Training (HcAT) organisation to administer SPRAT and a wide variety of associated assessment tools. Although SPRAT was originally developed for use by paediatric consultants, it has been modified and widened to cover a range of specialist areas. HcAT currently provides a series of modified versions of SPRAT and associated tools. These include:<sup>144</sup>

- With the Royal College of Paediatrics and Child Health (RCPCH): PAEDmini-CEX; ePAEDCbD (online paediatric case based discussion); eSPRAT (online SPRAT); DOPS (Directly Observed Procedural Skills);



and SAIL (Sheffield Assessment Instrument for Letters)

- As part of the Foundation Program (a two-year generic training programme which forms the bridge between medical school and specialist/general practice training): mini-ePAT; CbD (Case based Discussion); DOPS; and min-CEX
- With the Royal College of General Practitioners (RCGPs): GP-SPRAT; SELF-SPRAT and SHEFFPAT
- With the Royal College of Pathology (RCPATH): PATH-SPRAT; ePATH-SPRAT.<sup>144</sup>

An example of an HcAT tool and procedure is provided in Appendix B. In addition, the UK's General Medical Council (GMC) has developed its own patient and colleague questionnaires which have recently been trialed.<sup>133 145</sup> Once reviews are completed, feedback is provided to the doctor through educational supervisors in the case of trainees, or through HcAT in the case of practitioners.

The Canadian PAR is distributed through the College of Physicians & Surgeons of Alberta. PAR is the general name for a series of tools which undertaken once every five years by each physician in Alberta, irrespective of specialty. PAR involves a set of questionnaires completed by 25 patients, eight physician colleagues and eight non-physician healthcare co-workers. Like SPRAT, PAR covers five physician attributes: 1.clinical knowledge and skills; 2.communication skills; 3.psychosocial management; 4.office management; and 5.collegiality.

PAR includes instruments for a range of practitioners. For general practitioners; surgeons; anesthesiologists; medical specialists (psychiatrists, pediatricians and other medical specialists) and episodic care providers each doctor is provided with patient, medical colleague, co-worker and self assessment questionnaires. For doctors working lab medicine; and diagnostic imaging, questionnaires are given to referring physicians rather than patients, with the remaining three the same as other doctors. Results are reviewed by members of the Physician Performance Committee (PPC), a nine-member Council-appointed group. Doctors who achieve the top or bottom tenth percentile, or who cannot recruit the required number of reviewers are offered a 'practice visit' from their peers (using standardized audit instruments) which will provide the doctor involved with a detailed feedback report.<sup>146</sup>

SPRAT, min-PAT, SAIL and SHEFFPAT type tools (that is traditional peer review and feedback tools) tend to be used as periodic review tools. Tools such as CEX, mini-CEX, chart stimulated recall (CbD in the HcAT parlance) are often used to undertake a more detailed assessment if an area of difficulty is identified, so that a more comprehensive assessment can be obtained.<sup>1 146</sup> Mini-CEX (along with observation and other measures) has also been shown to be an effective measure of both a high and low level of professionalism.<sup>26</sup>

#### 4.2.5 The feedback process

Peer review is primarily and ultimately about the provision of feedback to doctors. Four factors influence feedback: the ability to accept feedback, particularly if it is negative; the tools used (as discussed in the previous section); and the quality of the feedback; and the skills of the person providing the feedback.

Fear of scrutiny and criticism is a common concern with any form of review process.<sup>124</sup> Difficulties with peer review feedback have been attributed at least in part to the doctor's ability to accept negative feedback from colleagues. This was particularly difficult to reconcile when it was inconsistent with the doctor's self perception of their abilities, but could be accepted if a process of facilitated reflection was undertaken.<sup>147 148</sup> Personal, confidential feedback was found to be more effective than generalised feedback for junior doctors,<sup>149</sup> an approach supported by the general literature on 360° feedback.<sup>150</sup>

The tools used to provide feedback vary. The most comprehensive paper (as opposed to face to face) model identified was that of HcAT. Their feedback forms include information on:

- Overall assessor mean score achieved by the clinician for all assessments undertaken
- Overall assessor mean score (generally from 1 to 6) achieved by the clinician's cohort
- Total number of assessments completed by clinician
- Information on the scale of the assessment tool (generally from 1 to 6)
- The mean assessor score per question
- How many assessors responded per question
- Minimum and maximum scores given by any assessor for this question
- Cohort mean score  $\pm$  2 standard deviations
- Clinical problem category
- Comments made on "especially good areas" of the assessment
- Comments made on suggestions for development
- Agreed actions
- Complexity of case (low/average/high), the assessor mean of each assessment and the month it was completed.<sup>151</sup>

A study of General Practitioners in Training (GPiTs) found significant differences between the GPiT's preferred mode of feedback and that provided by their trainers, with the trainees preferring feedback which included a large number of reflective remarks.<sup>152</sup> Non-specific feedback has been shown to provide little

impetus for change in clinician behaviour,<sup>153</sup> while a combination of peer review discussion with feedback was seen as having a greater impact on improving clinicians' adherence to clinical guidelines than feedback alone.<sup>37</sup> These findings are supported by Murphy et al (2009), who as discussed earlier, found multi-source feedback from peers and patient feedback were found to be the most reliable and feasible forms of feedback for general practitioners.<sup>101</sup>

These findings points to the need for adequate training of those providing the feedback, as well as those undertaking the review.<sup>154</sup> Such training is thought to assist the person providing feedback to couch the results in a positive manner, while still providing enough specific details to support the individuals' desire and ability to improve their practice, including developing a performance plan.<sup>155 156</sup> Appendix C provides details of the advice given to SPRAT assessors by HcAT on how to provide feedback.

A recent development in this field is the use of web-based systems for peer review analysis and feedback in the United Kingdom (discussed later in this section) and the United States. Electronic monitoring allows various aspects of the clinicians practice to be compared to their peers, especially if they are working in small or remote hospitals. Online peer review can work in two ways. In the first, commonly used in the UK, the website is used as a point of information, for the lodging and collection of data, and for the provision of statistical feedback, but the data itself is gathered in the traditional way, via observation of the clinician by their peers. In the US variants, the review is restricted to data entered by the clinician themselves. In the first case this was then compared, over the same time period, to the results from a group of peers (surgeons) in the same region. Comparisons were made on: complication rates; patient education; resource utilization; use of diagnostic testing and number of days before the patient returned to work.<sup>99</sup> In a second case, an automated on-demand feedback system collects clinical data stored in a repository as part of the doctor order entry program, and then compares the doctor's usage patterns to relative to a series of set templates. This information is then fed back to the doctors themselves.<sup>157</sup> Other internet based programs allow doctors to undertake assessment surveys, on specific topics such as knowledge of radiation exposure.<sup>158</sup> Research has been conducted into the most appropriate type of questions for online assessments.<sup>159</sup> Online systems are also being used to provide peer feedback to students.<sup>160</sup> An early study of the use of internet based MSF by Lockyear et al (2006) found that for the group of doctors they studied (international medical graduate practicing in Canada), traditional paper based survey methods produced better results than internet or phone survey methods. This may, however, be a function of the population group studied.<sup>161</sup>

Other forms of electronic support for peer review are also being used. In Western Australia, rural psychiatrists have established a video-conferencing peer review group to provide continuing professional development to doctors who are geographically isolated.<sup>162</sup>

#### 4.2.6 Barriers to peer review and strategies for improvement

A number of barriers to or challenges for the peer review process have been identified by researchers and practitioners. These include: lack of meaningful outcome data for peer comparison; time pressures making data collection and analysis difficult; confusion of peer review with revalidation and personal development plans; lack of experience in review procedures; fear of criticism from, and negative evaluations of, colleagues; and negative attitudes of doctors and peers towards peer review.<sup>163 164-166 167</sup>

Several improvements strategies, such as reviewer and feedback training have been identified previously as methods for improving the validity and acceptability of peer review. Wood et al reviewed the literature on MSF and found ten characteristics which could contribute to effective MSF (and peer review). These were:<sup>168</sup>

- The development of a positive review culture and organisational climate
- Clarity of purpose about the review
- Clear descriptions of desired behaviours (competencies)
- Scoring of only a few items at any one time
- Use of simple and fit for purpose scales
- Use of six to ten reviewers
- Comparison of reviewer findings to clinicians' self assessment
- Train those providing feedback
- Involve the individuals being reviewed in the development and assessment process
- Incorporate a developmental element for both reviewers and clinicians, to ensure they are able to address areas which require strengthening.

While the issue of addressing the findings of peer reviews is outside the scope of this monograph, it is important to consider what such strategies might involve. Williams (2007) notes that underperformance in physicians is understudied, even though estimates of what she terms "dyscompetence" in doctors range from 0.6% to 50% depending on the assessment method (with an average of 6% to 12%). In the US [as in Australia], identification of under-performing physicians is either through legal cases or peer review. Underperformance is generally addressed through continuing medical education. In Canada, more specialised programs, involving individual assessments of clinician knowledge and performance, specifically and individually designed educational programs, long-term commitments to education and learning, and ongoing assessments of performance.<sup>169 170</sup>

### 4.3 Evidence base for peer review

The relatively limited research base for peer review means that there is limited evidence for its effectiveness or otherwise. Some authors inferred evidence for peer review in medicine based on findings from other industries or associated techniques. Abdulla (2008) for example, cites studies that showed that 360 degree feedback increased motivation, and subsequently productivity in staff, but highlights evidence from other studies which showed limited or reduced performance, and even anger, after feedback. These difficulties were attributed to the approach and processes used, however, rather than the provision of feedback per se.<sup>3 171</sup> This difficulty in assessing peer review is echoed by Norcini (2003) who argues that *“given the broad range of ways peer evaluators can be used and the sizeable number of competencies they can be asked to judge, generalisations are difficult to derive and this form of assessment can be good or bad depending on how it is carried out.”*<sup>126: 539</sup> Recent studies have shown that 360<sup>0</sup> feedback has been effective in improving medical educators’ own ability to assess and provide feedback to residents’ on their professionalism,<sup>25</sup> in evaluating team based interventions, and even in enhancing teamwork effectiveness.<sup>172</sup> Where 360<sup>0</sup> feedback occurs in group settings, however, at least one study has shown that the results were less accurate, and resulted in greater contrast effects and more positive “halos” than for individual reviewers.<sup>173</sup>

A number of studies have shown that peer review in various forms is a reliable indicator of a range of clinical abilities. Ramsey et al (1993) showed that there was a strong correlation between the peer ratings of a professional’s medical knowledge and their medical board examination scores (between 0.5 – 0.6) but a low correlation between the rating of a professional’s communication and interprofessional relationship abilities and their examination scores (<0.15). In other words, they found that examination scores are not good indicators of professional ‘humanistic’<sup>174</sup> abilities<sup>175</sup> and that peer review might provide an alternative assessment process. In a follow up study, Ramsey et al went further, concluding that peer review was *“feasible to obtain reliable, multidimensional peer evaluations of physicians practicing in diverse clinical settings.”*<sup>176: 364</sup>

Other studies have shown that peer review of various forms is accurate in areas such as intensive care, surgery,<sup>177</sup> radiology,<sup>140</sup> emergency medicine,<sup>178</sup> paediatrics,<sup>139</sup> obstetrics and gynaecology,<sup>143</sup> pre-hospital trauma care<sup>179</sup> and oncology.<sup>180</sup> These studies, and other studies of peer review, are difficult to compare, however, as there are few common evaluation instruments, and the specific aspects evaluated differ. Ramsey (1996), for example, provided reviewers with a nine-point Likert scale, with which they assessed their peers on 11 cognitive and non-cognitive categories, including areas such as integrity and the psycho-social elements of care.<sup>176</sup> SPRAT, the only validated UK peer review tool, uses a six point scale from 1 (‘very poor’) to 6 (‘very good’). SPRAT combines the Likert scale with open-text responses and observations, allowing the reviewers to provide both structured and un-structured responses. The findings from Ramsey’s tool are compared to the American Board Exams, those from SPRAT to the UK’s

General Medical Council's standards of good medical practice.<sup>139</sup> The first study to evaluate workplace assessment tools found that the tools reviewed (mini-CEX, DOPS and MSF) showed that each of these methods were able to produce reliable scores across medical specialities.<sup>181</sup> Assessments of the individual elements or approaches to peer review also differ. A study of chart review, for example, showed that chart review was only moderately reliable in the assessment of general practitioners.<sup>109</sup>

The UK's Foundation Assessment Program (aimed at trainees i.e. registrars) has been in operation for a year, and has only recently been reviewed. The review uses case-focused assessments and multi-source feedback. The study showed "some evidence to support [it's] validity".<sup>107</sup> Another recent (2009) UK study by Murphy *et al* compared general practitioners' view of criterion audit, multi-source feedback from clinical and non-clinical colleagues, patient feedback (the CARE Measure), referral letters, significant event analysis, and video analysis of consultations. The study compared the performance of 171 GP registrars on each tool, in order to assess the tools' reliability and feasibility (factoring in the number of raters and assessments required, and utilising decision studies to determine the number of observations required to achieve high reliability for high-stake assessments for each instrument). The researchers also obtained feedback on the tools from the participating GPs via a questionnaire, which they analysed using descriptive statistics.<sup>101</sup> Of the six tools, multi-source feedback from colleagues and patient feedback on consultations were the most reliable and feasible methods (reliability co-efficient of 0.8 was obtained). The remaining four methods - criterion audit, referral letters, significant event analysis, and video analysis of consultations, were determined to have low feasibility, as they required ten or more assessors, per GP, to achieve reliable assessments.

Positive evaluations (that is, doctors felt that peer review was a useful tool in their development) have been identified in gynaecology,<sup>182</sup> general practice,<sup>183</sup> and general surgery. Hall (1999) found that two thirds of the doctors who had participated in their peer review were either considering or had implemented changes in their practice,<sup>137</sup> and other research has shown that peer review has resulted in either perceived or measured positive changes in: test ordering behaviour in interns;<sup>184</sup> adherence to clinical guidelines;<sup>37 185</sup> and management of acute respiratory infections.<sup>33</sup>

Studies of MRF review found that it is "feasible" in the assessment of communication skills (including identification of areas of strength and those needing improvement), and is useful in identifying areas of self over and under appraisal.<sup>130</sup> Violato *et al* (2008) found that there was evidence for the construct validity of MSF instruments used in general practice, whether the reviewer was a colleague, co-worker (i.e. non medical colleague) or patient.<sup>131</sup>

## 4.4 Limitations of peer review

There are several limitations to peer review processes and tools. The relatively recent introduction of formal peer review into medicine<sup>6</sup> has meant that, unlike other industries,<sup>3</sup> health services and professional bodies are in different states of readiness for the implementation of peer review, and its potential outcomes.

Furthermore the limited number of validated tools means that issues such as reliability of assessment procedures and instruments and the issue of equivalence, as well as more general issues of the impact of interpersonal relationships, the reviewer's degree of comfort with, and knowledge of a relatively new form of assessment, and their awareness of the stakes involved have been identified as concerns with existing tools and procedures.<sup>126</sup> SPRAT, the best known UK peer assessment tool, for example, has only been validated in one field study at the time of writing, although associated tools such as PATH-SPRAT (Histopathology-Sheffield Peer Review Assessment Tool) have also undergone validation testing with positive results.<sup>144</sup>

The choice, number and use of reviewers is another significant concern in the literature. In his critique of mini-PAT<sup>3</sup> Abdulla's (2003) notes concerns about the number of reviewers required to reduce measurement error and or the need to use different statistical sampling methods more suited to a small sample size.<sup>3</sup>

Concerns have also been raised about the: doctor's ability to choose appropriate reviewers (although this was discounted by Ramsey<sup>175 186</sup>), that is selection bias, including choosing reviewers at the 'right' level (i.e. not simply those that are considered to be at the same level as the individual under review)<sup>187</sup>; the effects of doctor-reviewer familiarity on assessments;<sup>155 166</sup> and the effect of professional backgrounds (doctors rating doctors, versus nurses rating doctors for example).<sup>188</sup> Group peer reviews (as in 360<sup>0</sup> feedback) have been found to result in increased "halo" effects.<sup>173</sup>

Consistency between self, colleague, patient and other assessments remains an issue with some forms of peer review, but direct comparison is impossible given differences in tools. Studies have shown that students' self and peer assessments differ significantly from their tutor assessments,<sup>189</sup> although other studies have shown that peer assessments was slightly more accurate than self-assessments.<sup>190 191</sup> Psychiatrists' colleague and patient assessments (of the psychiatrists' abilities to work with colleagues and relate to patients) were found to correlate with each other, while their self-assessment did not.<sup>192</sup> This finding is supported by Violato et al (2006) who showed that high ranking doctors rated themselves lower than their peer's assessment of them, and low ranking doctors, higher than their peers' assessment.<sup>193</sup> There is also some evidence that female (students) rate themselves lower than their peers on some measures, while men consistently rate themselves at the same level as their peers.<sup>194</sup>

Patients' reviews were found by Campbell et al (2008) to be highly skewed towards favourable impressions of doctors' performance, but this effect was very

similar to that of colleague reviewers, and had a high internal consistency. They concluded that patients were able to provide useful complementary perspectives of doctors' performance.<sup>133</sup>

There appears to be no clear evidence as to the effectiveness of peer reviewers based on their professions. One Danish study found that of 22 objectives chosen for assessment of internal medicine residents, 15 could reliably be assessed by doctors, 7 by nurses and none by non-medical personnel (in this case secretaries).<sup>29</sup> An US study, however, showed no significant differences between professions' reviews of resident physicians. Reviewers included therapists, nurses, social workers, case managers, and psychologists.<sup>195</sup> As with all peer review, the differences in outcomes may be attributable to the tools used (in the second case a purpose designed tool for the use of non-medical colleagues was developed) and or the areas of competence reviewed. Van Den Hombergh, Grol et al (1999) found that practice visits and feedback from non-physician observers were more appreciated than those by peers,<sup>113</sup> but that change was more evident after mutual practice visits and feedback than after a visit and feedback by a non-physician observer.<sup>114</sup>

A number of authors have noted that many of the issues associated with peer reviewers, such as halo effect and central tendency, may be addressed through the adequate preparation and education of reviewers.<sup>3 155 188</sup> This is the approach which has been undertaken by the HcAT, who provides both written and online advice to reviewers.

Legally, disclosure of findings from peer reviews are a significant concern in the United States.<sup>196</sup> In the UK, one concern has been the ability of assessments to be robust enough to withstand legal challenges from the doctors involved.<sup>1</sup> In Canada, the confidentiality of PAR is guaranteed under the Medical Professional Act, and information from PAR cannot be used in either disciplinary or legal proceedings, but in the case where a) the public is at immediate risk of harm; b) there is a serious breach of ethics; or c) there is an unreasonable failure to comply with the requirement of the PAR process, physicians can be referred for formal investigation.<sup>146</sup>



The underlying problems in the development of the evidence base for peer review are issues of participation, definition, and consistency. As one Scottish study succinctly put it: “participation in 360-degree appraisal was scant, and no consensus on the use of outcome measures was reached ... More work needs to be done on the appraisal process for it to gain the full confidence of the profession.”<sup>197</sup> This sentiment is supported by what remains the most comprehensive study of peer review tools to date, Evans (2004), who concluded that “*The instruments developed to date for physicians to evaluate characteristics of colleagues need further assessment of validity before their widespread use is merited.*”<sup>2: 328</sup> While more validated tools are currently available, differences in the scope, focus, methods, and approaches to peer review make comparisons difficult, although internal validity and acceptability has been proven in several cases. One important question remains however, there is no direct evidence as yet that peer review is either an indicator of, or actually improves, patient outcomes.<sup>3 158</sup>

## 5. CONCLUSION

Peer review in its various forms is gaining greater recognition and use amongst medical professionals, as clinicians, professional bodies and services seek to find ways to monitor, and improve the quality of medical care. While the principle of peer review is well embedded in medical culture, acceptance of the practice is dependent on a variety of cultural, interpersonal, organisational, technical and procedural issues. There are an increasing number of validated tools, but differences in the methods used, and competencies assessed by such tools makes comparisons across disciplines and systems difficult. Nonetheless, evidence from other industries, along with rigorous studies of existing tools and methods should provide a stronger evidence base in the near future.

## 6. REFERENCES

1. Davies H, Howells R. How to assess your specialist registrar. *Arch Dis Child* 2004;89(12):1089-93.
2. Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. *British Medical Journal* 2004;328(7450):1240-43.
3. Abdulla A. A critical analysis of mini peer assessment tool (mini-PAT). *Journal of the Royal Society of Medicine* 2008;101(1):22-26.
4. Braithwaite J, Travaglia JF. An overview of clinical governance policies, practices and initiatives. *Australian Health Review* 2008;32(1):10-22.
5. Veltman LL. Getting to havarti: moving toward patient safety in obstetrics. *Obstet Gynecol* 2007;110(5):1146-50.
6. Southgate L, Cox J, David T, Hatch D, Howes A, Johnson N, et al. The General Medical Council's Performance Procedures: Peer review of performance in the workplace. *Medical Education, Supplement* 2001;35(1):9-19.
7. Hindle D, Braithwaite J, Iedema R, Travaglia J. *Patient safety: a comparative analysis of eight inquiries in six countries*. Sydney: Centre for Clinical Governance Research in Health, University of NSW and Clinical Excellence Commission, 2006.
8. Hindle D, Braithwaite J, Iedema R. *Patient safety research: a review of the technical literature*. Sydney: Centre for Clinical Governance Research, UNSW, 2005.
9. Royal College of Physicians. *Doctors in society: medical professionalism in a changing world*. London: Royal College of Physicians of London, 2005. <http://www.rcplondon.ac.uk/pubs/books/docinsoc/docinsoc.pdf> (accessed 21 April 2009).
10. Alderson CJ, Hogg P. Advanced radiographic practice - The legal aspects. *Radiography* 2003;9(4):305-14.
11. Berenholtz SM, Hartsell TL, Pronovost PJ. Learning From Defects to Enhance Morbidity and Mortality Conferences. *American Journal of Medical Quality* 2009:1062860609332370.
12. Rose DM, Boyce PM. Clinical supervision and its effects on the quality of practice of qualified practitioners. *Australasian Psychiatry* 1999;7(1):11-13.
13. Linn BS, Arostegui M, Zeppa R. Performance rating scale for peer and self assessment. *British Journal of Medical Education* 1975;9(2):98-101.
14. Linn BS, Arostegui M, Zeppa R. Peer and self assessment in undergraduate surgery. *Journal of Surgical Research* 1976;21(6):453-6.
15. Douglas RM, Blood A. Evaluation and peer review of the role of specialist and general medical units in a teaching hospital. *Australian & New Zealand Journal of Medicine* 1978;8(3):337-43.
16. Davies G. *Queensland Public Hospitals Commission of Inquiry*. Brisbane: Queensland Government, 2005.

17. Thomas H. *Sick to Death: A Manipulative Surgeon and a Healthy System in Crisis-a Disaster Waiting to Happen*. Sydney: Allen & Unwin, 2007.
18. Australian Broadcasting Corporation. *New health chief contacts Bundaberg support group*. Sydney: Australian Broadcasting Corporation, . <http://www.abc.net.au/news/stories/2005/07/26/1422540.htm> (accessed 21 April 2009).
19. Rout A, Roberts P. Peer review in Nursing and Midwifery: a literature review. *Journal of Clinical Nursing* 2008;17(4):427-42.
20. Australian Competition & Consumer Commission Australian Health Workforce Officials' Committee. *Report to Australian Health Ministers: Review of Australian Specialist Medical Colleges*. Canberra: Commonwealth of Australia, 2005. <http://www.nhwt.gov.au/documents/Publications/2005/Review%20of%20Australia%20specialist%20medical%20colleges.pdf> (accessed
21. Alpert HR, Hillman BJ. Quality and variability in diagnostic radiology. *Journal of the American College of Radiology* 2004;1(2):127-32.
22. Whitehouse A, Hassell A, Bullock A, Wood L, Wall D. 360 degree assessment (multisource feedback) of UK trainee doctors: field testing of team assessment of behaviours (TAB). *Medical Teacher* 2007;29(2-3):171-6.
23. Joshi R, Ling FW, Jaeger J. Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills. *Academic Medicine* 2004;79(5):458-63.
24. Miller GE. The assessment of clinical skills/competence/performance. *Academic Medicine* 1990;65(9 suppl):S63-7.
25. Stark R, Korenstein D, Karani R. Impact of a 360-degree professionalism assessment on faculty comfort and skills in feedback delivery. *Journal of General Internal Medicine* 2008;23(7):969-72.
26. Reed DA, West CP, Mueller PS, Ficalora RD, Engstler GJ, Beckman TJ. Behaviors of highly professional resident physicians. *JAMA - Journal of the American Medical Association* 2008;300(11):1326-33.
27. Maker VK, Donnelly MB. Surgical Resident Peer Evaluations-What Have We Learned. *Journal of Surgical Education* 2008;65(1):8-16.
28. Arnold L. Assessing professional behavior: Yesterday, today, and tomorrow. *Academic Medicine* 2002;77(6):502-15.
29. Allerup P, Aspegren K, Ejlersen E, Jorgensen G, Malchow-Moller A, Moller MK, et al. Use of 360-degree assessment of residents in internal medicine in a Danish setting: a feasibility study. *Medical Teacher* 2007;29(2-3):166-70.
30. Ouellet R. Revalidation of Canadian physicians. *Canadian Medical Association Journal* 2009;180(5):540.
31. Adams R, Jelley D, Illing J, Van Zwanenberg T, Walker C. Will GP appraisal undermine practice professional development plans? *Clinician in Management* 2004;12(4):187-90.
32. Anonymous. Continuing competence of physicians. *JAMA* 1971;217(11):1537-41.

33. Reyes-Morales H, Flores-Hernandez S, Tome-Sandoval P, Perez-Cuevas R. A multifaceted education intervention for improving family physicians' case management. *Family Medicine* 2009;41(4):277-84.
34. Adams HP, Del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the early management of adults with ischemic stroke: A guideline from the American heart association/American stroke association stroke council, clinical cardiology council, cardiovascular radiology and intervention council, and the atherosclerotic peripheral vascular disease and quality of care outcomes in research interdisciplinary working groups. *Stroke* 2007;38(5):1655-711.
35. Adams JA, Kaplan RA, Starling SP, Mehta NH, Finkel MA, Botash AS, et al. Guidelines for Medical Care of Children Who May Have Been Sexually Abused. *Journal of Pediatric and Adolescent Gynecology* 2007;20(3):163-72.
36. Albert US, Koller M, Lorenz W, Doherty J, Schulz KD, Wagner U, et al. A concept for the implementation and evaluation of the guideline "Early Detection of Breast Cancer in Germany". *Zeitschrift für Ärztliche Fortbildung und Qualitätssicherung* 2004;98(5):347-59.
37. Amin FA, Harrison S, Heywood P. Testing interventions for improving adherence to diabetic guideline: A randomised controlled study in Bahrain. *Journal of the Bahrain Medical Society* 2004;16(4):156-65.
38. Anonymous. A multicentre audit of hospital referral for radiological investigation in England and Wales. Royal College of Radiologists Working Party. *World Hospitals* 1992;28(1):7-13.
39. Anonymous. Influence of the Royal College of Radiologists' guidelines on hospital practice: a multicentre study. Royal College of Radiologists Working Party. *Bmj* 1992;304(6829):740-3.
40. Kamyar M, Johnson BJ, McAnaw JJ, Lemmens-Gruber R, Hudson SA. Adherence to clinical guidelines in the prevention of coronary heart disease in type II diabetes mellitus. *Pharm World Sci* 2008;30(1):120-7.
41. Curry SJ. Organizational interventions to encourage guideline implementation. *Chest* 2000;118(2 SUPPL.):40S-46S.
42. Strike C, Wenghofer E, Gnam W, Hillier W, Veldhuizen S, Millson M. Physician peer assessments for compliance with methadone maintenance treatment guidelines. *Journal of Continuing Education in the Health Professions* 2007;27(4):208-13.
43. Schectman JM, Schroth WS, Verme D, Voss JD. Randomized Controlled Trial of Education and Feedback for Implementation of Guidelines for Acute Low Back Pain. *Journal of General Internal Medicine* 2003;18(10):773-80.
44. Amalberti R, Vincent C, Auroy Y, De Saint Maurice G. Violations and migrations in health care: A framework for understanding and management. *Quality and Safety in Health Care* 2006;15(SUPPL. 1):i66-i71.
45. Graber MA, Hartz A, James P, Nugent A, Green MD. An alternative method of determining standard of care in alleged cases of malpractice. *Journal of the American Board of Family Practice* 2005;18(6):453-8.

46. 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, Supplement* 2001;35(1):20-28.
47. Kun LE, Haffty BG, Bosma J, Strife JL, Hattery RR. American Board of Radiology Maintenance of Certification-Part IV: Practice Quality Improvement for Radiation Oncology. *International Journal of Radiation Oncology Biology Physics* 2007;68(1):7-12.
48. Wheeland RG. The pitfalls of regulating office-based surgery by state legislatures and boards of medical examiners. *Seminars in Cutaneous Medicine and Surgery* 2005;24(3):124-27.
49. Katz PO. American Board of Internal Medicine gastroenterology recertification: How can ACG help? *American Journal of Gastroenterology* 2004;99(7):1205-06.
50. Anonymous. Is your process for ID'ing disruptive docs complete? What your peer review process should contain. *Healthcare Benchmarks & Quality Improvement* 2008;15(10):105-06.
51. Zbar RIS, Taylor LD, Canady JW. The disruptive physician: righteous Maverick or dangerous Pariah? *Plast Reconstr Surg* 2009;123(1):409-15.
52. McKay J, Murphy DJ, Bowie P, Schmuck ML, Lough M, Eva KW. Development and testing of an assessment instrument for the formative peer review of significant event analyses. *Qual Saf Health Care* 2007;16(2):150-53.
53. Dannefer EF, Henson LC, Bierer SB, Grady-Weliky TA, Meldrum S, Nofziger AC, et al. Peer assessment of professional competence. *Medical Education* 2005;39(7):713-22.
54. Hekelman FP, Flynn SP, Glover PB, Galazka SS, Phillips Jr JA. Peer coaching in clinical teaching: Formative assessment of a case. *Evaluation and the Health Professions* 1994;17(3):366-81.
55. Rodrigues J, Sengupta A, Mitchell A, Kane C, Kane C, Maxwell S, et al. The Southeast Scotland Foundation Doctor Teaching Programme--is "near-peer" teaching feasible, efficacious and sustainable on a regional scale? *Medical Teacher* 2009;31(2):e51-7.
56. Hulsman RL, Harmsen AB, Fabriek M. Reflective teaching of medical communication skills with DiViDU: assessing the level of student reflection on recorded consultations with simulated patients. *Patient Education & Counseling* 2009;74(2):142-9.
57. Alexander BJ, Lindow LE, Schock MD. Measuring the impact of cooperative learning exercises on student perceptions of peer-to-peer learning: a case study. *Journal of Physician Assistant Education* 2008;19(3):18-25.
58. Adriaansen MJM, Frederiks CMA. Design of a postgraduate course in palliative care. *Journal of Continuing Education in Nursing* 2002;33(6):279-82.
59. Redwood S, Lloyd H, Carr E, Hancock H, McSherry R, Campbell S, et al. Evaluating nurse consultants' work through key informant perceptions. *Nursing Standard* 2007;21(17):35-40.

60. Anderson G, Sheps SB, Cardiff K. Hospital-based utilization management: a cross-Canada survey. *CMAJ Canadian Medical Association Journal* 1990;143(10):1025-30.
61. Ankone A. One surgeon is not the other. Doctors must regulate themselves. *Ziekenhuis* 1988;18(4):135-38.
62. Anderson CE. The physician profile system. *California Medicine* 1969;110(4):342-5.
63. Gruber T, Rudnitsky B. Can we be certain of PRO accuracy and accountability? Inaccurate "report cards" generated by medicare claims data as a marker for laboratory testing in diabetic patients. *American Journal of Medical Quality* 2002;17(5):171-4.
64. Dans PE, Weiner JP, Otter SE. Peer review organizations. Promises and potential pitfalls. *New England Journal of Medicine* 1985;313(18):1131-37.
65. Hickey EC, Savage AM. Improving the quality of inter-hospital transfers. *Journal of Quality Assurance* 1991;13(4):16-20.
66. Granatir T. PROs pose challenge to physician practice. *Hospital Medical Staff* 1984;13(11):17-22.
67. Butman SM. Percertification for percutaneous transluminal coronary angioplasty in Medicare beneficiaries: A melting pot or a need for better national standards? *Catheterization and Cardiovascular Diagnosis* 1990;21(4):227-32.
68. Dippe SE, Bell MM, Wells MA, Lyons W, Clester S. A peer review of a peer review organization. *Western Journal of Medicine* 1989;151(1):93-96.
69. Anonymous. Medicare program; acquisition, protection and disclosure of utilization and quality control peer review organization (PRO) information--HCFA. Proposed rule. *Federal Register* 1984;49(74):14977-86.
70. Jost TS. Medicare peer review organizations. *Quality Assurance in Health Care* 1989;1(4):235-48.
71. Sprague L. Contracting for quality: Medicare's quality improvement organizations. *NHPF Issue Brief* 2002(774):1-15.
72. Weinmann C. Quality improvement in health care. A brief history of the medicare peer review organization (PRO) initiative. *Evaluation and the Health Professions* 1998;21(4):413-18.
73. Weingart SN, Mukamal K, Davis RB, Davies DT, Jr., Palmer RH, Cahalane M, et al. Physician-reviewers' perceptions and judgments about quality of care. *International Journal for Quality in Health Care* 2001;13(5):357-65.
74. Anonymous. Peer review organizations may need to reconsider the use of physician and nurse reviews of quality of care. *AHRQ Research Activities* 2002(261):18-19.
75. Adler NE, Milstein A. Evaluating the impact of physician peer review: Factors associated with successful PSROs. *American Journal of Public Health* 1983;73(10):1182-85.

76. Marshall CL, Bluestein M, Briere E, Chapin C, Darling B, Davis K, et al. Improving outpatient diabetes management through a collaboration of six competing, capitated Medicare managed care plans. *American Journal of Medical Quality* 2000;15(2):65-71.
77. Stevenson KB, McMahon JW, Harris J, Hillman JR, Helgerson SD. Increasing pneumococcal vaccination rates among residents of long-term-care facilities: provider-based improvement strategies implemented by peer-review organizations in four western states. *Infection Control & Hospital Epidemiology* 2000;21(11):705-10.
78. Anderson FA, Jr., Audet A. Physician practices in the prevention of deep vein thrombosis: the MassPRO DVT study. *Orthopedics* 1996;19:9-11.
79. Darling WD, Gleitz HG. PRO sanctions: a practical guide to the process. *Medical Staff Counselor* 1988;2(1):12-8.
80. Bradley EH, Carlson MDA, Gallo WT, Scinto J, Campbell MK, Krumholz HM. From adversary to partner: Have quality improvement organizations made the transition? *Health Services Research* 2005;40(2):459-76.
81. Brooks JH, Renz KK, Richardson SL, White S, Hosey J. Systems versus performance problems: a peer review organization's perspective. *Qrb* 1992;Quality Review Bulletin. 18(5):172-7.
82. Lau LM, Warner HR. Performance of a diagnostic system (Iliad) as a tool for quality assurance. *Computers & Biomedical Research* 1992;25(4):314-23.
83. Rothschild RD. Providers question PROs' effectiveness. Critics contend peer review organizations are too costly and fail to improve the quality of care. *Health Progress* 1992;73(6):28-32.
84. Anderson BJ. Confidentiality of peer review. *Journal of Medical Practice Management* 2001;16(6):317-20.
85. Goldstein RL. When doctors divulge: is there a "threat from within" to psychiatric confidentiality? *Journal of Forensic Sciences* 1989;34(2):433-8.
86. Isackson C, McCahill M. Arnett v. Dal Cielo: peer review confidentiality threatened by medical board investigational subpoenas. *Health Care Law Newsletter* 1995;10(11):3-5.
87. Abruzzo MD. Peer review may not be confidential when fairness of process is at issue. *Manag Care* 1999;8(11):68-9.
88. Ward NO. The essential and ethical expert: Serving as a medical expert witness in the 21<sup>st</sup> century. *Otolaryngology - Head and Neck Surgery* 2007;136(1):4-7.
89. Schofferman J. Opinions and testimony of expert witnesses and independent medical evaluators. *Pain Medicine* 2007;8(4):376-82.
90. Andrew LB. Expert Witness Testimony: The Ethics of Being a Medical Expert Witness. *Emergency Medicine Clinics of North America* 2006;24(3):715-31.
91. Royal College of General Practitioners. *What sort of doctor? Assessing quality of care in general practice*. Lonon: Royal College of General Practitioners, 1985.



92. Hyman NH, Osler T, Cataldo P, Burns EH, Shackford SR. Anastomotic leaks after bowel resection: what does peer review teach us about the relationship to postoperative mortality? *Journal of the American College of Surgeons* 2009;208(1):48-52.
93. American College of Physicians. American College of Physicians Ethics Manual (3rd Edition). *Annals of Internal Medicine* 1992;117(11):947-60.
94. Hoop JG, Smyth AC, Roberts LW. Ethical Issues in Psychiatric Research on Children and Adolescents. *Child and Adolescent Psychiatric Clinics of North America* 2008;17(1):127-48.
95. Toohey J, Shakespeare TP, Morgan G. RANZCR 2006 Peer review audit instrument. *Journal of Medical Imaging and Radiation Oncology* 2008;52(4):403-13.
96. Merkur S, Mossialos E, Long M, McKee M. Physician revalidation in Europe. *Clinical Medicine, Journal of the Royal College of Physicians of London* 2008;8(4):371-76.
97. Crossley J, McDonnell J, Cooper C, McAvoy P, Archer J, Davies H. Can a district hospital assess its doctors for re-licensure? *Medical Education* 2008;42(4):359-63.
98. Murphy DJ, Bruce D, Eva KW. Workplace-based assessment for general practitioners: using stakeholder perception to aid blueprinting of an assessment battery. *Medical Education* 2008;42(1):96-103.
99. Allen JW, DeSimone KJ. Valid peer review for surgeons working in small hospitals. *American Journal of Surgery* 2002;184(1):16-18.
100. Thompson AM, Ritchie W, Stonebridge PA. Could sequential individual peer reviewed mortality audit data be used in appraisal? *Surgeon* 2005;3(4):288-92.
101. Murphy DJ, Bruce DA, Mercer SW, Eva KW. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Advances in Health Sciences Education* 2009;14(2):219-32.
102. McKay J, Shepherd A, Bowie P, Lough M. Acceptability and educational impact of a peer feedback model for significant event analysis. *Medical Education* 2008;42(12):1210-17.
103. Paul R, Van Der Vleuten C, Rethans JJ, Grol R, Aretz K. Assessment of practicing family physicians: Comparison of observation in a multiple-station examination using standardized patients with observation of consultations in daily practice. *Academic Medicine* 1999;74(1):62-69.
104. McMillan R, Cameron N. Factors influencing the submission of videotaped consultations by general practitioners for peer review and educational feedback. *Quality in Primary Care* 2006;14(2):85-89.
105. Martin D, Regehr G, Hodges B, McNaughton N. Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Academic Medicine* 1998;73(11):1201-06.
106. Agius RM. Peer review audit in occupational medicine. *Journal of the Society of Occupational Medicine* 1990;40(3):87-88.

107. Davies H, Archer J, Southgate L, Norcini J. Initial evaluation of the first year of the Foundation Assessment Programme. *Medical Education* 2009;43(1):74-81.
108. Staton LJ, Kraemer SM, Patel S, Talente GM, Estrada CA. 'Correction:'Peer chart audits: A tool to meet Accreditation Council on Graduate Medical Education (ACGME) competency in practice-based learning and improvement. *Implement Sci* 2007;2:24.
109. Goulet F, Jacques A, Gagnon R, Racette P, Sieber W. Assessment of family physicians' performance using patient charts: interrater reliability and concordance with chart-stimulated recall interview. *Eval Health Prof* 2007;30(4):376-92.
110. Russo RM, Gururaj VJ, Laude TA, Rajkumar SV, Ventre SA. A chart audit peer review system in an ambulatory service. *Pediatrics* 1975;56(2):246-54.
111. Vautrey R, Neal R. Using peer review visits to carry out the clinical governance baseline assessment in one primary care group. *Journal of Clinical Governance* 2000;8(3):135-40.
112. Bruno-Bazureault P, Touboul P, Hoffliger P, Marine-Barjoan E, Pradier C, Patuano C, et al. Peer visits to community-based general practitioners and pediatricians as part of the "Antibiotics only when necessary" campaign in the Alpes-Maritimes district. Methodology, feasibility, and interest. *Presse Medicale* 2006;35(5 Pt 1):749-54.
113. Van Den Hombergh P, Grol R, Van Den Hoogen HJM, Van Den Bosch WJHM. Practice visits as a tool in quality improvement: Acceptance and feasibility. *Quality in Health Care* 1999;8(3):167-71.
114. Grol R, Van Den Hoogen HJM, Van Den Bosch WJHM. Practice visits as a tool in quality improvement: mutual visits and feedback by peers compared with visits and feedback by non-physician observers. *Quality in Health Care* 1999;8(3):161-66.
115. Rickenbach M, Dunleavey J, Little P, Mullee M. Senior house officer training. Impact of existing peer review visits needs to be increased. *Bmj* 1997;314(7097):1829-30.
116. Balla M, Knothe B, Lancaster J, Prager S, Beatson J. Group peer review in psychiatry: The relationship to quality improvement and quality care. *Australian and New Zealand Journal of Psychiatry* 1996;30(5):653-59.
117. Beatson J, Knothe B. Peer review groups: Problems and solutions. *Australian and New Zealand Journal of Psychiatry* 1997;31(5):655-63.
118. Kuyvenhoven M, Pieters R, Jacobs M, Touw-Otten F, Spreeuwenberg C. Consultation among peers in general practice; from no consultation to peer review. *Scandinavian Journal of Primary Health Care* 1988;6(1):5-12.
119. Beyer M, Gerlach FM, Flies U, Grol R, Krol Z, Munck A, et al. The development of quality circles/peer review groups as a method of quality improvement in Europe. Results of a survey in 26 European countries. *Family Practice* 2003;20(4):443-51.
120. Kuhlein T. From peer group to the quality assurance cycle - Evaluation of a project on lipid lowering therapy. *Zeitschrift fur Allgemeinmedizin* 2005;81(1):18-23.

121. Tracey JM. Peer review. Part 1: Attitudes to internal and external standards. *New Zealand Medical Journal* 1991;104(904):8-10.
122. Tracey JM. Peer review. Part 2: The influence of internal and external standards on outcome. *New Zealand Medical Journal* 1991;104(906):66-7.
123. Schuwirth LWT, Southgate L, Page GG, Paget NS, Lescop JM, Lew SR, et al. When enough is enough: a conceptual basis for fair and defensible practice performance assessment. *Medical Education* 2002;36(10):925-30.
124. Adshead L, White PT, Stephenson A. Introducing peer observation of teaching to GP teachers: A questionnaire study. *Medical Teacher* 2006;28(2):e68-e73.
125. Agee C. Professional review committee improves the peer review process. *Healthc Exec* 2007;33(1):52-3.
126. Norcini JJ. Peer assessment of competence. *Medical Education* 2003;37(6):539-43.
127. Keely E, Myers K, Dojeiji S, Campbell C. Peer assessment of outpatient consultation letters--feasibility and satisfaction. *BMC Medical Education* 2007;7:13.
128. Keely E, Myers K, Dojeiji S, Campbell C. Peer assessment of outpatient consultation letters - Feasibility and satisfaction. *BMC Medical Education* 2007;7(13).
129. Malling B, Bonderup T, Mortensen L, Ringsted C, Scherpbier A. Effects of multi-source feedback on developmental plans for leaders of postgraduate medical education. *Medical Education* 2009;43(2):159-67.
130. Calhoun AW, Rider EA, Meyer EC, Lamiani G, Truog RD. Assessment of communication skills and self-appraisal in the simulated environment: feasibility of multirater feedback with gap analysis. *Simulation in Healthcare: The Journal of The Society for Medical Simulation* 2009;4(1):22-9.
131. Violato C, Lockyer JM, Fidler H. Changes in performance: a 5-year longitudinal study of participants in a multi-source feedback programme. *Medical Education* 2008;42(10):1007-13.
132. Archer J, Norcini J, Southgate L, Heard S, Davies H. mini-PAT (Peer Assessment Tool): a valid component of a national assessment programme in the UK? *Advances in Health Sciences Education* 2008;13(2):181-92.
133. Campbell JL, Richards SH, 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 & Safety in Health Care* 2008;17(3):187-93.
134. 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.
135. Allen MH, Carpenter D, Sheets JL, Miccio S, Ross R. What do consumers say they want and need during a psychiatric emergency? *Journal of Psychiatric Practice* 2003;9(1):39-58.

136. Weingart SN, Davis RB, Palmer RH, Cahalane M, Hamel MB, Mukamal K, et al. Discrepancies between explicit and implicit review: physician and nurse assessments of complications and quality. *Health Services Research* 2002;37(2):483-98.
137. Hall W, Violate 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):52-57.
138. Hays RB, Davies H, Beards JD, Caldon LJ, Farmer EA, Finucane PM, et al. Selecting performance assessment methods for experienced physicians. *36* 2002;10(910-917).
139. Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *British Medical Journal* 2005;330(7502):1251-53.
140. Lockyer JM, Violato C, Fidler HM. Assessment of radiology physicians by a regulatory authority. *Radiology* 2008;247(3):771-78.
141. Baigent ML, Harding R. IRQN AWARD Article: Radiology Audit: A Program Developed for Remote or Solo-Practicing Radiologists: A New Zealand Quality Initiative. *JACR Journal of the American College of Radiology* 2007;4(11):841-43.
142. Borgstede JP, Wilcox PA. Quality care and safety know no borders. *Biomedical Imaging and Intervention Journal* 2007;3(3)(e34).
143. Wood L, Wall D, Bullock A, Hassell A, Whitehouse A, Campbell I. 'Team observation': a six-year study of the development and use of multi-source feedback (360-degree assessment) in obstetrics and gynaecology training in the UK. *Medical Teacher* 2006;28(7):e177-84.
144. Davies H, Archer J, Bateman A, Dewar S, Crossley J, Grant J, et al. Specialty-specific multi-source feedback: Assuring validity, informing training. *Medical Education* 2008;42(10):1014-20.
145. Healthcare Assessment and Training. *Healthcare Assessment and Training*. Sheffield: Sheffield Children's Hospital. <http://www.hcat.nhs.uk/rcpch/rcpch-home> (accessed 20 April 2009).
146. Physician Achievement Review. *Program description*. Edmonton: College of Physicians & Surgeons of Alberta. <http://www.par-program.org/PAR-History2.htm> (accessed
147. Sargeant J, Mann K, van der Vleuten C, Metsemakers J. "Directed" self-assessment: practice and feedback within a social context. *Journal of Continuing Education in the Health Professions* 2008;28(1):47-54.
148. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Advances in Health Sciences Education* 2008;13(3):275-88.
149. Hadjianastassiou VG, Karadaglis D, Gavalas M. A comparison between different formats of educational feedback to junior doctors: A prospective pilot intervention study. *Journal of the Royal College of Surgeons of Edinburgh* 2001;46(6):354-57.

150. Peiperl MA. Getting 360 degrees feedback right. *Harvard Business Review* 2001;79(1):142-7.
151. Healthcare Assessment and Training. "Interpreting CFA Feedback" Slideshow. Sheffield: Sheffield Children's Hospital, 2009. <http://www.hcat.nhs.uk/rcpch/rcpch-home> (accessed
152. Prins FJ, Sluijsmans DMA, Kirschner PA. Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education* 2006;11(3):289-303.
153. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education* 2007;41:583-91.
154. Boylan O, Bradley T, McKnight A. GP appraisal: A qualitative study highlights implications for future training. *Education for Primary Care* 2005;16(1):51-56.
155. Sargeant JM, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Hayes VM, et al. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: A pilot study. *Academic Medicine* 2003;78(10 SUPPL.):S42-S44.
156. Malling B, Bonderup T, Mortensen L, Ringsted C, Scherpbier A. Effects of multi-source feedback on developmental plans for leaders of postgraduate medical education. *Medical Education* 2009;43(2):159-67.
157. Hulse NC, Del Fiol G, Bradshaw RL, Roemer LK, Rocha RA. Towards an on-demand peer feedback system for a clinical knowledge base: A case study with order sets. *Journal of Biomedical Informatics* 2008;41(1):152-64.
158. Rice HE, Frush DP, Harker MJ, Farmer D, Waldhausen JH, Committee AE. Peer assessment of pediatric surgeons for potential risks of radiation exposure from computed tomography scans. *Journal of Pediatric Surgery* 2007;42(7):1157-64.
159. Rotthoff T, Baehring T, Dicken H-D, Fahron U, Richter B, Fischer MR, et al. Comparison between Long-Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial. *BMC Medical Education* 2006;6:50.
160. Hughes C, Toohey S, Velan G. eMed teamwork: A self-moderating system to gather peer feedback for developing and assessing teamwork skills. *Medical Teacher* 2008;30(1):5-9.
161. Lockyer J, Blackmore D, Fidler H, Crutcher R, Salte B, Shaw K, et al. A study of a multi-source feedback system for international medical graduates holding defined licences. *Medical Education* 2006;40(4):340-7.
162. Greenwood J, Williams R. Continuing professional development for Australian rural psychiatrists by videoconference. *Australasian Psychiatry* 2008;16(4):273-76.
163. Williams J. Appraisal for consultants. *Current Paediatrics* 2003;13(7):549-54.

164. Lewis M, Elwyn G, Wood F. Appraisal of family doctors: An evaluation study. *British Journal of General Practice* 2003;53(491):454-60.
165. Roth RR, Porter PJ, Bisbey GR, May CR. The attitudes of family physicians toward the peer review process. *Arch Fam Med* 1993;2(12):1271-5.
166. McKinstry B, Peacock H, Shaw J. GP experiences of partner and external peer appraisal: A qualitative study. *British Journal of General Practice* 2005;55(516):539-43.
167. Grol R, Wensing M. Implementation of quality assurance and medical audit: General practitioners' perceived obstacles and requirements. *British Journal of General Practice* 1995;45(399):548-52.
168. Wood L, Hassell A, Whitehouse A, Bullock A, Wall D. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Medical Teacher* 2006;28(7):e185-91.
169. Williams BW. The prevalence and special educational requirements of dyscompetent physicians. *Journal of Continuing Education in the Health Professions* 2006;26(3):173-91.
170. Page GG, Bates J, Dyer SM, Vincent DR, Bordage G, Jacques A, et al. Physician-assessment and physician-enhancement programs in Canada. *Canadian Medical Association Journal* 1995;153(12):1723-28.
171. Bret J, Atwater L. 360-degree feedback: accuracy, reactions and perceptions of usefulness. *Journal of Applied Psychology* 2001;86(5):254-84.
172. Paige JT, Aaron DL, Yang T, Howell DS, Hilton CW, Cohn I, Jr., et al. Implementation of a preoperative briefing protocol improves accuracy of teamwork assessment in the operating room. *American Surgeon* 2008;74(9):817-23.
173. Palmer JK, Loveland JM. The influence of group discussion on performance judgments: rating accuracy, contrast effects, and halo. *Journal of Psychology* 2008;142(2):117-30.
174. Ramsey PG, Wenrich M. Evaluation of humanistic qualities and communication skills. *Journal of General Internal Medicine* 1993;8(3):164.
175. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *Journal of the American Medical Association* 1993;269(13):1655-60.
176. Ramsey PG, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Academic Medicine* 1996;71(4):364-70.
177. Lockyer J. Multisource feedback in the assessment of physician competencies. *Journal of Continuing Education in the Health Professions* 2003;23(1):4-12.
178. Lockyer JM, Violato C, Fidler H. The assessment of emergency physicians by a regulatory authority. *Academic Emergency Medicine* 2006;13(12):1296-303.
179. Notzer N, Eldad A, Donchin Y. Assessment of physician competence in prehospital trauma care. *Injury* 1995;26(7):471-4.

180. Siegal EC, Angelakis EJ, Hartman A. Can peer review contribute to earlier detection of breast cancer? A quality initiative to learn from false-negative mammograms. *Breast Journal* 2008;14(4):330-34.
181. Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical Education* 2008;42(4):364-73.
182. Chaillet N, Dube E, Dugas M, Francoeur D, Dube J, Gagnon S, et al. Identifying barriers and facilitators towards implementing guidelines to reduce caesarean section rates in Quebec. *Bulletin of the World Health Organization* 2007;85(10):791-97.
183. Wenghofer EF, Way D, Shaw Moxam R, Wu H, Faulkner D, Klass DJ. Effectiveness of an enhanced peer assessment program: Introducing education into regulatory assessment. *Journal of Continuing Education in the Health Professions* 2006;26(3):199-208.
184. Chen DR, Chung KP, Wu LC, Tseng CY, Chang CC. The relationship between peer comparison feedback and TG test ordering among internists: An exploratory study. *Taiwan Journal of Public Health* 2006;25(1):26-36.
185. Callahan CW, Chan DS, Brads-Pitt T, Underwood GH, Jr., Imamura DT. Performance feedback improves pediatric residents' adherence to asthma clinical guidelines. *Journal of Clinical Outcomes Management* 2003;10(6):317-22.
186. Ramsey PG, Wenrich MD. Peer ratings: An assessment tool whose time has come. *Journal of General Internal Medicine* 1999;14(9):581-82.
187. Lurie SJ, Nofziger AC, Meldrum S, Mooney C, Epstein RM. Effects of rater selection on peer assessment among medical students. *Medical Education* 2006;40(11):1088-97.
188. Johnson D, Cujec B. Comparison of self, nurse and physician assessment of residents rotating through an intensive care unit. *Critical Care Medicine* 1998;26:1811-6.
189. Machado JLM, Machado VMP, Grec W, Bollela VR, Vieira JE. Self- and peer assessment may not be an accurate measure of PBL tutorial process. *BMC Medical Education* 2008;8(55).
190. Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Medical Teacher* 2007;29(5):e122-e32.
191. Evans AW, Leeson RMA, Petrie A. Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery. *Medical Education* 2007;41(9):866-72.
192. Lelliott P, Williams R, Mears A, Andiappan M, Owen H, Reading P, et al. Questionnaires for 360-degree assessment of consultant psychiatrists: development and psychometric properties. *British Journal of Psychiatry* 2008;193(2):156-60.
193. Violato C, Lockyer J. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: implications for self-directed learning. *Advances in Health Sciences Education* 2006;11(3):235-44.

194. Lurie SJ, Meldrum S, Nofziger AC, Sillin ILF, Mooney CJ, Epstein RM. Changes in self-perceived abilities among male and female medical students after the first year of clinical training. *Medical Teacher* 2007;29(9-10):921-26.
195. Musick DW, McDowell SM, Clark N, Salcido R. Pilot study of a 360-degree assessment instrument for physical medicine & rehabilitation residency programs. *American Journal of Physical Medicine & Rehabilitation* 2003;82(5):394-402.
196. Bitterman M, Bitterman RA. ED discussions after patient death not covered by peer review privilege. *ED Legal Letter* 2009;20(3):25-28.
197. Hedge R, Hughson AV. Consultant appraisal of consultant appraisal. *Psychiatric Bulletin* 2008;32(6):214-17.
198. Physician Achievement Review. *PAR Patient questionnaire for Medical Specialists*. Edmonton: College of Physicians & Surgeons of Alberta, 2005. [http://www.par-program.org/MS\\_Patient.pdf](http://www.par-program.org/MS_Patient.pdf) (accessed
199. Healthcare Assessment and Training. *ePaedCbD*. Sheffield: Sheffield Children's Hospital, 2009. <http://www.hcat.nhs.uk/rcpch/epaedcbd> (accessed
200. Healthcare Assessment and Training. *Guidance on presenting eSPRAT feedback to trainees*. Sheffield: Sheffield Children's Hospital, 2009. <http://www.hcat.nhs.uk/rcpch/rcpch-home> (accessed
201. Borgstede JP, Lewis RS, Bhargavan M, Sunshine JH. RADPEER quality assurance program: A multifacility study of interpretive disagreement rates. *JACR Journal of the American College of Radiology* 2004;1(1):59-65.
202. Massagli TL, Carline JD. Reliability of a 360-degree evaluation to assess resident competence. *American Journal of Physical Medicine & Rehabilitation* 2007;86(10):845-52.
203. 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.
204. Nair BR, Alexander HG, McGrath BP, Parvathy MS, Kilsby EC, Wenzel J, et al. The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Med J Aust* 2008;189(3):159-61.
205. Ramsey P, Carline J, Blank L, Wenrich M. Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Academic Medicine* 1996;71:364-70.
206. Roark RM, Schaefer SD, Yu GP, Branovan DI, Peterson SJ, Lee WN. Assessing and documenting general competencies in otolaryngology resident training programs. *Laryngoscope* 2006;116(5):682-95.
207. Stark R, Korenstein D, Karani R. Impact of a 360-degree professionalism assessment on faculty comfort and skills in feedback delivery.[see comment]. *Journal of General Internal Medicine* 2008;23(7):969-72.



208. Staton LJ, Kraemer SM, Patel S, Talente GM, Estrada CA. Peer chart audits: A tool to meet Accreditation Council on Graduate Medical Education (ACGME) competency in practice-based learning and improvement. *Implementation Science* 2007;2:24.

## **Appendix A: Summary evidence sheet**

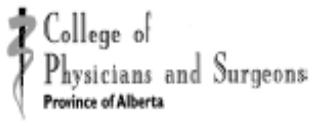
**Table 6: Summary evidence sheet for peer review**

Topic area	Peer review
<b>Definition:</b>	Peer review is the process by which individuals of the same profession, grade or setting, critically assess their colleague(s)' performance, in order to reinforce areas of strength and quality, and identify areas for development.
<b>Origin:</b>	Peer review in some form has been used in medicine since the inception of case review in the 1880s. It has been in use in Australia since the 1970s. In recent years changes to credentialing processes, along with concerns about the quality and safety of healthcare in general have resulted in a renewed interest in the use of peer review as an assessment and quality monitoring technique.
<b>Description:</b>	Peer review provides individual doctors with the opportunity to reflect on their current performance, and to compare their self assessment with that of their colleagues. It provides professional bodies and health services with a method for assessing competence in doctors, particularly in areas which are difficult to assess, such as communication, interprofessionalism, teamwork and relationship building with patients. Peer review enables these institutions to identify doctors who are at risk of poor performance and require assistance, and to formulate remedial strategies.
<b>Evidence base:</b>	There is a growing evidence base for peer review as an assessment and quality improvement process, both in health care and in related industries. Several validated review tools are currently in use. Peer review has been shown to be effective in improving use of guidelines, and in facilitating (positive) changes in practice amongst medical practitioners in several disciplines.
<b>Current use:</b>	A variety of peer review methods are currently in use. Reviews can include self, peer (medical practitioner to medical practitioner), colleague (clinician and non-clinician) and patient. Assessment. Peer review can be conducted using observation methods, document review, or surveys (either paper or online). There are an increasing number of validated peer review tools in use, both general and discipline specific. Peer review is currently in use around the world by individual clinicians, professional groups, professional bodies, accrediting bodies, and educational institutions.
<b>Applications for clinical practice improvement:</b>	Peer review is an effective method: for comparing self assessment of otherwise difficult to assess competencies, with colleagues and patients; to assist in the identification of medical practitioners experiencing difficulties; and to aid in changes in practice behaviour, such as uptake of guidelines.


## **Appendix B: Examples of peer review instruments**

Appendix B1: PAR patient questionnaire for medical specialists<sup>198</sup>

8173160972



## Patient Questionnaire



**Gender:**  Male  Female

**Age:**     Less than 1     19 - 45  
               1 - 5                     46 - 65  
               6 - 10                    66 and over  
               11 - 18

**This questionnaire is being completed by:**  
 Self(patient)     Caregiver/parent

**Physician's Name:** Dr. \_\_\_\_\_

**Marking Instructions**

Please indicate your answer by filling in the bubbles like this, ● not like ☒ or ☑. Thank you!

**Interpretation of the Rating Scale**  
 This form is used by a variety of patients, therefore, not all of the following items may be relevant to you. If any of these items are **NOT** relevant to you, mark these "Unable to Assess".

Indicate how much you agree with the statements on the left side of the page using the following scale.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Unable to Assess UA
	1	2	3	4	5	UA
<b>Based on my MOST RECENT VISIT, this doctor:</b>						
1. Explained my illness or concern to me clearly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Explained my treatment choices or options	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Explained my follow-up plan to me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Told me how and when to take my medicine, if medicine was prescribed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Told me of side effects of the medicine, if medicine was prescribed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Based on ALL OF YOUR VISITS to this doctor, how do you feel about this doctor's attitude and behavior towards you? This doctor:</b>						
6. Spends enough time with me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Shows interest in my problems	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Asks details about my personal life, when appropriate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Answers my questions well	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Examines me appropriately for my problems	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Treats me with respect	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Helps me with my fears and worries	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Rate each statement about this doctor's office. The office:</b>						
13. Is easy to get into (e.g. wheelchair accessible, parking)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. Has appropriate waiting areas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. Examining rooms are adequately sized and have adequate equipment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. Is clean and in good repair	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. Provides adequate privacy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please turn over

2611160973

	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5	Unable to Assess UA
<b>How do you feel this doctor runs his or her practice?</b>						
<b>Telephone:</b>						
18. I can reach the office by phone during the day	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. I receive an appropriate explanation if my appointment is delayed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20. My messages are returned	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>The Staff:</b>						
21. Are helpful and pleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. Are respectful of patients	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. Behave in a professional manner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. Work well with the doctor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25. Prevent patients from hearing confidential information about other patients	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Office Practices:</b>						
26. In an emergency situation this doctor's office provides me with clear instructions on what I am to do	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27. This doctor provides reports to my family doctor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
28. When asked, this doctor provides insurance and medico legal reports in a timely manner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
29. When asked, this doctor provides reports, files or copies of letters in a timely manner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. This doctor arranges appointments with other specialists when necessary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
31. This doctor's office follows-up on serious problems	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
32. I am told what to do if my problems do not get better	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>General:</b>						
33. I am asked about prescription and non-prescription medicine I may be taking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
34. This doctor talks to me about preventative care [e.g. quitting smoking, weight control, sleeping, alcohol, exercise, etc.]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
35. This doctor has good written health information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
36. This doctor refers me to appropriate educational resources [i.e., web sites, brochures, patient support groups, books]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
37. I would go back to this doctor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
38. I would send a friend to this doctor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
39. This doctor presents him/herself in a professional manner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
40. I was helped by this doctor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Appendix B2: ePaedCbD<sup>199</sup>**

*This text is taken from the HcAT website as an example of how it provides information to clinicians wishing to undertake peer reviews. The example provided is for ePaedCbD or online paediatric case based discussion.*

Case-based discussion (CbD) is used to enable the documenting of conversations about, and presentations of, cases by trainees. This activity happens throughout training, but is rarely conducted in a way that provides systematic assessment and structured feedback. The approach is called chart stimulated recall (CSR) in the US and Canada, and is widely used for the assessment of residents and of established doctors who are in difficulty. CbD is designed to assess clinical reasoning and decision-making and the application or use of medical knowledge in relation to patient care for which the trainee has been directly responsible. It also enables the discussion of the ethical and legal framework of practice, and in all instances, it allows trainees to discuss why they acted as they did. In the UK it has been extensively used by the GMC and NCAS for the assessment of doctors in difficulty. It is one of the four tools in the Foundation Assessment Programme and is also being used by a number of other Specialties.

The table below highlights the relevant RCPCH assessment standard in relation to each question:

Question area	RCPCH PMETB standards
Medical record keeping	Effective written communication skills (clear, accurate and facilitate safe handover to other professionals)
Clinical assessment	Responsibility for an analytic and focused 3 way consultation Focussed and analytic assessments of common and complex clinical problems in paediatrics The ability to come to a safe decision about the 'most likely' diagnosis in paediatrics and lead management for common and complex cases seeking additional advice and opinion as appropriate
Investigations and referrals	Leads management for common and complex conditions in paediatrics effectively seeking additional advice and opinion as appropriate Demonstrates effective collaboration with other specialists in using and interpreting complex investigations undertaken in children
Management of challenging and complex situations	Leads management for common and complex conditions in paediatrics seeking additional advice and opinion as appropriate Leadership skills ensuring effective management of behavioural, emotional and psychosocial aspects of illness in children and families Responsibility for an effective response to complex challenge and stress within paediatrics
Risk assessment	Independent thinking including being able to challenge guidelines and procedures where appropriate Knowledge and application of risk assessment strategies in all areas of paediatric practice
Treatment	Responsibility for safe prescribing in paediatrics in common and complex situations and for the supervision of others Effective skills to maintain and develop knowledge and clinical skills required of a specialist in paediatric

### **Who can complete it?**

SpRs and ST4+ trainees can be assessed by Consultants and SASGs only.  
ST1-ST3 trainees can be assessed by a ST4+ trainee, SpR, SASG or Consultant.

### **How is it completed?**

The focus of the discussion should be around an actual entry the trainee has made in the notes, exploring the thought processes that underpinned the entry. It should not be seen as an opportunity to conduct a viva but should be seen as an opportunity to explore their clinical reasoning and decision making.

A few possible questions are set out below to provide an example of the sort of style of questioning that is effective;

- What was going through your mind when you wrote that management plan, just talk me through your thought processes?
- I see that you have written down a number of different investigations – how did you think the results would help you work out what was going on and what you needed to do?
- You have referred to the ward guidelines in your notes – tell me a bit about how you used the guidelines to help plan management and whether there were any aspects that didn't fit for this case?
- I see that you have decided to treat this child with xx – talk me through how you decided to prescribe that regime and what the alternatives you considered were?
- You have written down that you were going to ask Dr X for their advice – what specifically did you want to discuss with them, why was it important in this case, how did their advice help and what did you learn from it?

Ideally half the cases should be selected by the trainee and half by the assessor to allow exploration of the effect of this on ePaedCbD ratings. The cases selected should cover a range of clinical problem areas and should be cases that reflect their stage of training.

Following discussion of the case the ePaedCbD must be completed online only by the assessor via their own HcAT account, not the trainee's. A paper form is available but this only provided to facilitate and record the discussion.

### **What feedback is given?**

Immediate feedback focusing on aspects of the discussion of the case where the assessor felt the trainee did especially well (anything especially good), suggestions for development and any agreed action should be provided and documented.

A collated feedback report for DOPS, PAEDmini-CEX and ePaedCbD assessments will be provided for each trainee once a year. This will include group comparator data and can be discussed with their educational supervisor to inform the ARCP process.



**Appendix C: Advice given by HcAT  
to assessors providing feedback**

## Guidance on presenting eSPRAT feedback to trainees<sup>200</sup>

*This text is taken from the HcAT website as an example of how it provides information to assessors providing feedback to clinicians. The example provided is for SPRAT.*

This guidance is to assist Educational Supervisors provide SPRAT feedback to trainees, especially those whose feedback suggests they are experiencing difficulties in certain areas of their practice.

### 1. PRESENTING FEEDBACK

Make sure that the discussion focuses on areas of strength, as well as ensuring there is adequate discussion of any concerns. Where possible, if there are any problems try to review any other assessments or evidence available.

#### 1.1 Identifying the problem

- Give the trainee the opportunity to present their views on the feedback first.
- Identify strengths and ensure these are highlighted during discussion.
- Clarify areas of concern, both in relation to absolute score (borderline/below standard) and to the rest of the peer group.
- Explore the trainee's insight into problem areas – do they agree with concerns raised? Try to use the evidence provided by SPRAT to help the trainee gain a better understanding of their performance.
- Discuss whether problems are sufficiently clearly defined to be able to develop a remediation plan or whether more evidence is needed to help clarify the problem.
- Consider whether there are factors outside work (distractors), e.g. problems in home life, stress, ill health, etc. that may be interfering with the trainee's ability to function well at work.

#### 1.2 Addressing the problems and supporting the trainee

- Ask the trainee to suggest ideas to address problem areas – how will they act to improve their performance?
- Set possible learning objectives and timescales for improvement. The objectives should be documented in a personal development plan and must be reviewed in the future. The learning objectives should be agreed and signed by both the trainee and the supervisor.
- Suggest targeted training and feedback in the areas identified as needing further development.
- Where the problem is not clear or where clarification of the nature of the problem is needed, agree on ways in which more evidence in relation to this area might be collected.
- Discuss ways in which you can help and support the trainee.
- Offer neutral support networks: local or Deanery counselling service, or national services such as the British Medical Association.
- Send a letter to the trainee following the meeting if you have discussed problem areas and agreed means of addressing them. This serves as a useful way of ensuring that both the supervisor and the trainee have a clear record of the discussion and its outcomes.

#### 1.3 What to do if the SPRAT feedback does not reflect the view of the Educational Supervisor

- Seek more detailed feedback from colleagues who work closely with this individual in different settings.
- Identify whether negative feedback in a specific domain has been observed frequently or is a reflection of a single incident.
- Check for discrimination or bias against the candidate ('halo or horns').

#### 1.4 What to do if the trainee refuses to accept that there is a problem

- Remind the trainee that a number of their assessors, not just one, have raised concerns.
- Reinforce that although the trainee perceives their behaviour to be appropriate, it might not come across to their colleagues as they would hope.
- If the trainee has consistently rated their own performance as much better than their assessor scores, it suggests the trainee may have a lack of insight into their problems.
- Be aware of any potential sources of discrimination that the trainee may feel contributes to a judgement of poor performance.
- Collect more evidence, either to confirm that there is a problem or to support the trainee's view that they are actually performing at a satisfactory level.
- Ensure that you have carefully documented all discussions, including telephone calls and emails.

## 2 TRAINEES IN SERIOUS DIFFICULTY

**A trainee can be said to be in serious difficulty if:**

- a number of colleagues have raised significant concerns over practice in a number of different areas  
**or**
- one or more serious incident(s) has been identified  
**or**
- concerns have been raised about a doctor's ability to function safely because of personal difficulties, including ill health.

Trainees in serious difficulty may find receiving the feedback distressing. It is therefore suggested that:

- the first meeting should be short and acknowledge there are some difficult discussions that need to take place. Ask if they wish to bring anyone with them to a future meeting for support
- arrange the second meeting after a very short interval. Explain to the trainee that written notes will be taken and nominate a scribe
- following the meeting, send the trainee a record of the discussion and agreed actions and ask them to sign it, acknowledging it as a correct record
- if there is serious cause for concern let the trainee know it will be the responsibility of the Educational Supervisor to inform the Regional Advisor/Programme Director and/or College Tutor.

**Remember the Educational Supervisor should not try to sort out this problem alone.**

The Postgraduate Deanery is able to provide useful advice and feedback. The Director of Postgraduate Medical Education or the Clinical Tutor often have experience dealing with trainees in difficulty, as do Programme Directors, Regional Advisors and College Tutors. These people are there to help you and the trainee.

## **Appendix D: Examples of recent studies of peer review**

**Table 7: Examples of recent studies of peer review**

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
Archer et al (2005) <sup>139</sup>	UK	To determine whether a multisource feedback questionnaire, SPRAT (Sheffield peer review assessment tool), is a feasible and reliable assessment method to inform the record of in-training assessment for paediatric senior house officers and specialist registrars.	Participants: 112 paediatric senior house officers and middle grades	<p>Trainees' clinical performance was evaluated using SPRAT sent to clinical colleagues of their choosing. Responses were analysed to determine variables that affected ratings and their measurement characteristics. Setting: Three tertiary hospitals and five secondary hospitals across a UK deanery.</p> <p>Main outcome measures: 95% confidence intervals for mean ratings; linear and hierarchical regression to explore potential biasing factors; time needed for the process per doctor</p>	20 middle grades and 92 senior house officers were assessed using SPRAT to inform their record of in-training assessment; 92/1120 (82%) of their proposed raters completed a SPRAT form. As a group, specialist registrars (mean 5.22, SD 0.34) scored significantly higher ( $t = -4.765$ ) than did senior house officers (mean 4.81, SD 0.35) ( $P < 0.001$ ). The grade of the doctor accounted for 7.6% of the variation in the mean ratings. The hierarchical regression showed that only 3.4% of the variation in the means could be additionally attributed to three main factors (occupation of rater, length of working relationship, and environment in which the relationship took place) when the doctor's grade was controlled for (significant $F$ change $< 0.001$ ). 93 (83%) of the doctors in this study would have needed only four raters to achieve a reliable score if the intent was to determine if they were satisfactory. The mean time taken to complete the questionnaire by a rater was six minutes. Just over an hour of administrative time is needed for each doctor	SPRAT seems to be a valid way of assessing large numbers of doctors to support quality assurance procedures for training programmes. The feedback from SPRAT can also be used to inform personal development planning and focus quality improvements.

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
Archer et al (2008) <sup>132</sup>		To design, implement and evaluate a multi-source feedback instrument	Six hundred and ninety-three mini-PAT assessments were undertaken for 553 trainees across 12 Deaneries in England, Wales and Northern Ireland. Two hundred and nineteen trainees were F1s or PRHOs and 334 were F2s. Trainees identified 5544 assessors of whom 67% responded	Tool mapped against Foundation Curriculum Trainees' clinical performance assessed using 16 questions rated against a six-point scale on two occasions in the pilot period. Responses were analysed to determine internal structure, potential sources of bias and measurement characteristics.	The mean score for F2 trainees was 4.61 (SD = 0.43) and for F1s was 4.44 (SD = 0.56). An independent t test showed that the mean scores of these 2 groups were significantly different ( $t = -4.59$ , $df = 390$ , $p < 0.001$ ). 43 F1s (19.6%) and 19 F2s (5.6%) were assessed as being below expectations for F2 completion. The factor analysis produced 2 main factors, one concerned clinical performance, the other humanistic qualities. Seventy-four percent of F2 trainees could have been assessed by as few as 8 assessors (95% CI +/-0.6) as they either scored an overall mean of 4.4 or above or 3.6 and below. Fifty-three percent of F1 trainees could have been assessed by as few as 8 assessors (95% CI +/-0.5) as they scored an overall mean of 4.5 or above or 3.5 and below. The hierarchical regression when controlling for the grade of trainee showed that bias related to the length of the working relationship, occupation of the assessor and the working environment explained 7% of the variation in mean scores when controlling for the year of the Foundation Programme ( $R$ squared change = 0.06, $F$	Content validity was assured by a mapping exercise against the Foundation Curriculum  As part of an assessment programme, mini-PAT appears to provide a valid way of collating colleague opinions to help reliably assess Foundation trainees.

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
					<i>change = 8.5, significant F change &lt;0.001).</i>	
Borgstede et al (2004) <sup>201</sup>	US	<i>To develop and test a radiology peer review system that adds minimally to workload, is confidential, uniform across practices, and provides useful information to meet the mandate for "evaluation of performance in practice" that is forthcoming from the American Board of Medical Specialties as one of the four elements of maintenance of certification.</i>	<i>31 radiologists</i>	<i>RADPEER has radiologists who review previous images as part of a new interpretation record their ratings of the previous interpretations on a 4-point scale. Reviewing radiologists' ratings of 3 and 4 (disagreements in nondifficult cases) are reviewed by a peer review committee in each practice to judge whether they are misinterpretations by the original radiologists. Final ratings are sent for central data entry and analysis. A pilot test of RADPEER was conducted in 2002.</i>	<i>Fourteen facilities participated in the pilot test, submitting a total of 20,286 cases. Disagreements in difficult cases (ratings of 2) averaged 2.9% of all cases. Committee-validated misinterpretations in nondifficult cases averaged 0.8% of all cases. There were considerable differences by modality. There were substantial differences across facilities; few of these differences were explicable by mix of modalities, facility size or type, or being early or late in the pilot test. Of 31 radiologists who interpreted over 200 cases, 2 had misinterpretation rates significantly (<math>P &lt; .05</math>) above what would be expected given their individual mix of modalities and the average misinterpretation rate for each modality in their practice.</i>	<i>A substantial number of facilities participated in the pilot test, and all maintained their participation throughout the year. Data generated are useful for the peer review of individual radiologists and for showing differences by modality. RADPEER is now operational and is a good solution to the need for a peer review system with the desirable characteristics listed above.</i>
Campbell et al (2008) <sup>133</sup>	UK	<i>To investigate the utility of the GMC patient and colleague questionnaires in assessing the professional performance of</i>	<i>541 doctors gave preliminary agreement to take part in the study. Responses were received from 13 754 patients</i>	<i>Cross-sectional questionnaire surveys. Settings were a range of UK clinical practice settings.</i>	<i>Questionnaire performance and standardised scores for each doctor derived from patient and colleague responses. Participant doctors were similar to non-participants in respect of age and gender. The patient and colleague questionnaires were acceptable to participants as</i>	<i>The GMC patient and colleague questionnaires offer a reliable basis for the assessment of professionalism among UK doctors. If used in the revalidation of doctors' registration, they would be</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>a large sample of UK doctors.</i>	<i>attending one of 380 participant doctors, and from 4269 colleagues of 309 participant doctors.</i>		<i>evidenced by low levels of missing data. One patient questionnaire item seemed to cause confusion for respondents and requires rewording. Both patient and colleague responses were highly skewed towards favourable impressions of doctor performance, with high internal consistency. To achieve acceptable levels of reliability, a minimum of 8 colleague questionnaires and 22 patient questionnaires are required. G coefficients for both questionnaires were comparable with internationally recognised survey instruments of broadly similar intent. Patient and colleague assessments provided complementary perspectives of doctors' performance. Older doctors had lower patient-derived and colleague-derived scores than younger doctors. Doctors from a mental health trust and doctors providing care in a variety of non-NHS settings had lower patient scores compared with doctors providing care in acute or primary care trust settings.</i>	<i>capable of discriminating a range of professional performance among doctors, and potentially identifying a minority whose practice should be subjected to further scrutiny.</i>
Davies et al (2009) <sup>107</sup>	UK	<i>This study represents an initial evaluation of the first year</i>	<i>The study sample included 3640 F1 trainees from 10</i>	<i>Descriptive analyses were undertaken for total number of encounters, assessors</i>	<i>A total of 2929 trainees submitted at least one of all four methods. A mean of 16.6 case-focused assessments were submitted per</i>	<i>The FAP is feasible and achieves acceptable reliability. There is some evidence to support its</i>



Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>(F1) of the Foundation Assessment Programme (FAP), in line with Postgraduate Medical Education and Training Board (PMETB) assessment principles.</i>	<i>English deaneries.</i>	<i>and trainees, mean number of assessments per trainee, mean number of assessments per assessor, time taken for the assessments, mean score and standard deviation for each method. Reliability was estimated using generalisability coefficients. Pearson correlations were used to explore relationships between instruments.</i>	<i>F1 trainee. Based on a return per trainee of six of each of the case-focused assessments, and eight assessors for multi-source feedback, 95% confidence intervals (CIs) ranged between 0.4 and 0.48. The estimated time required for this is 9 hours per trainee per year. Scores increased over time for all instruments and correlations between methods were in keeping with their intended focus of assessment, providing evidence of validity.</i>	<i>validity. Collated assessment data should form part of the evidence considered for selection and career progression decisions although work is needed to further develop the FAP. It is in any case of critical importance for the profession's accountability to the public.</i>
Davies et al (2008) <sup>144</sup>	UK	<i>The white paper 'Trust, Assurance and Safety: the Regulation of Health Professionals in the 21st Century' proposes a single, generic multi-source feedback (MSF) instrument in the UK. Multi-source feedback was proposed as part of the assessment programme for</i>	<i>92 trainees</i>	<i>An existing instrument was modified following blueprinting against the histopathology curriculum to establish content validity. Trainees were also assessed using an objective structured practical examination (OSPE). Factor analysis and correlation between trainees' OSPE performance and the MSF were used to explore validity.</i>	<i>All 92 trainees participated and the assessor response rate was 93%. Reliability was acceptable with eight assessors (95% confidence interval 0.38). Factor analysis revealed two factors: 'generic' and 'histopathology'. Pearson correlation of MSF scores with OSPE performances was 0.48 (<math>P = 0.001</math>) and the histopathology factor correlated more highly (histopathology <math>r = 0.54</math>, generic <math>r = 0.42</math>; <math>t = -2.76</math>, d.f. = 89, <math>P &lt; 0.01</math>). Trainees scored least highly in relation to ability to use histopathology to solve clinical problems (mean = 4.39) and provision of good reports (mean = 4.39). Three of six doctors whose means were &lt;</i>	<i>This paper highlights the importance of validating an MSF instrument within the specialty-specific context as, in addition to assuring content validity, the PATH-SPRAT (Histopathology-Sheffield Peer Review Assessment Tool) also demonstrates the potential to inform training as part of a quality improvement model.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>Year 1 specialty training in histopathology.</i>			<i>4.0 received free text comments about report writing. There were 83 forms with aggregate scores of &lt; 4. Of these, 19.2% included comments about report writing. Specialty-specific MSF is feasible and achieves satisfactory reliability. The higher correlation of the 'histopathology' factor with the OSPE supports validity.</i>	
Joshi et al (2004) <sup>23</sup>	US	<i>To test the reliability of the 360-degree evaluation instrument for assessing residents' competency in interpersonal and communication skills</i>	<i>Residents and evaluators</i>	<i>Ten-item questionnaires were distributed to residents and evaluators at Monmouth Medical Center in Long Branch, New Jersey, in March/April, 2002. The scoring scale was 1-5; the highest score was 50. Data were maintained strictly confidential; each resident was assigned a code. Completed data sheets were collated by category and entered into a spreadsheet. The total and mean scores by each category of evaluator were calculated for each resident and a rank order list created. Shrout-Fleiss (model 2)</i>	<i>Intraclass correlation coefficients showed a narrow range, from .85-.54. The highest ranked resident overall ranked high and the lowest was low with most evaluators. The rank order among fellow residents was markedly different from other evaluator categories. Pearson correlation coefficients showed significant correlation between faculty and ancillary staff, (<math>p = .002</math>). Patients as evaluators did show intraclass correlation, but did not correlate significantly with other categories. Scores from colleagues correlated negatively with all other categories of evaluators.</i>	<i>The 360-degree instrument appears to be reliable to evaluate residents' competency in interpersonal and communication skills. Information from the assessment may provide feedback to residents. Areas of improvement identified by the scores would suggest areas for improvement and further ongoing assessment.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
				<p><i>intraclass correlation coefficients measured reliability of ratings within each group of evaluators. Reliability/reproducibility among evaluators' scores were tested by the Pearson correlation coefficient (<math>p &lt; .05</math>).</i></p>		
Keely et al (2007) <sup>128</sup>	UK	<p><i>Written correspondence is one of the most important forms of communication between health care providers, yet there is little feedback provided to specialists. The objective of this study was to determine the feasibility and satisfaction of a peer assessment program on consultation letters and to determine inter-rater reliability between family physicians and</i></p>	<p><i>Nine Internal Medicine specialists/ Sub-specialists from two tertiary care centres submitted 10 letters with patient and physician identifiers removed. Two Internal Medicine specialists and 2 family physicians from the other centre rated each letter (to protect writer anonymity).</i></p>	<p><i>A rating scale of nine 5-point Likert scale items including specific content, style items, education value of the letter and an overall rating was developed from a previous validated tool. A satisfaction survey was sent to each writer and rater after collation of the results. A follow-up survey was sent 6-8 months later.</i></p>	<p><i>There was a high degree of satisfaction with the process and feedback. The rating scale information was felt to be useful and appropriate for evaluating the quality of consultation letters by 6/7 writers. 5/7 seven writers felt that the feedback they received resulted in immediate changes to their letters. Six months later, 6/9 writers indicated they had maintained changes in their letters. Raters rank ordered letters similarly (Cronbach's alpha 0.57-0.84) but mean scores were highly variant. At site 1 there were significant differences in scoring brevity (<math>p &lt; 0.01</math>) between family physician and specialist raters; whereas, at site 2 there were differences in scoring of history (<math>p &lt; 0.01</math>), physical examination (<math>p &lt; 0.01</math>) and educational value (<math>p &lt; 0.01</math>) of the letter.</i></p>	<p><i>Most participants found peer assessment of letters feasible and beneficial and longstanding changes occurred in some individuals. Family physicians and specialists appear to have different expectations on some items. Further studies on reliability and validity, with a larger sample, are required before high stakes professional assessments include consultation letters.</i></p>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>specialists.</i>				
Lelliott et al (2008) <sup>192</sup>	UK	<i>Expert clinical judgement combines technical proficiency with humanistic qualities. To test the psychometric properties of questionnaires to assess the humanistic qualities of working with colleagues and relating to patients using multisource feedback.</i>		<i>Analysis of self-ratings by 347 consultant psychiatrists and ratings by 4422 colleagues and 6657 patients.</i>	<i>Mean effectiveness as rated by self, colleagues and patients, was 4.6, 5.0 and 5.2 respectively (where 1=very low and 6=excellent). The instruments are internally consistent (Cronbach's alpha &gt;0.95). Principal components analysis of the colleague questionnaire yielded seven factors that explain 70.2% of the variance and accord with the domain structure. Colleague and patient ratings correlate with one another (r=0.39, P&lt;0.001) but not with the self-rating. Ratings from 13 colleagues and 25 patients are required to achieve a generalisability coefficient (Erho(2)) of 0.75.</i>	<i>Reliable 360-degree assessment of humane judgement is feasible for psychiatrists who work in large multiprofessional teams and who have large case-loads.</i>
Lockyer et al (2008) <sup>140</sup>	Canada	<i>To determine whether it is possible to develop a feasible, valid, and reliable multisource feedback program for radiologists.</i>	<i>190 radiologists, eight radiologic colleagues (peers), eight referring physicians, and eight co-workers (eg, technicians).</i>	<i>Surveys with 38, 29, and 20 items were developed to assess individual radiologists by eight radiologic colleagues (peers), eight referring physicians, and eight co-workers (eg, technicians), respectively, by using five-point scales along with an "unable to assess" category. Radiologists completed a self-assessment on</i>	<i>Data from 190 radiologists were available. The mean numbers of respondents per physician were 7.5 of eight (1259 of 1520, 83%), 7.15 of eight (1337 of 1520, 88%), and 7.5 of eight (1420 of 1520, 93%) for peers, referring physicians, and co-workers, respectively. The internal consistency reliability indicated all instruments had a Cronbach alpha of more than 0.95. The generalizability coefficient analysis indicated that the peer, referring physicians, and</i>	<i>The psychometric examination of the data suggests that the instruments developed to assess radiologists are a feasible way to assess radiology practice and provide evidence for validity and reliability.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
				<i>the basis of the peer questionnaire. Items addressed key competencies related to clinical competence, collegiality, professionalism, workplace behavior, and self-management. The study was approved by the University of Calgary Conjoint Health Ethics Research Board.</i>	<i>coworker instruments achieved a generalizability coefficient of 0.88, 0.79, and 0.87, respectively. The factor analysis indicated that four factors on the colleague questionnaire accounted for 70% of the total variance: clinical competence, collegiality, professional development, and workplace behavior. For the referring physician survey, three factors accounted for 64.1% of the variance: professional development, professional consultation, and professional responsibility. Two factors on the co-worker questionnaire accounted for 63.2% of the total variance: professional responsibility and patient interaction.</i>	
Malling et al (2009) <sup>129</sup>	US	<i>Multi-source feedback (MSF) is a widely used developmental tool for leaders in organisations including those dealing with health care. This study was performed to</i>	<i>Consultants responsible for PGME in clinical departments (CREs) Heads of departments and young doctors-in-training.</i>	<i>An MSF instrument was developed based on literature on the subject and previous investigations. The instrument was used by consultants responsible for PGME in clinical departments (CREs). Apart from CREs' self-ratings, MSF responses were collected from heads of</i>	<i>The MSF instrument was found to be feasible, valid and reliable (Cronbach's alpha = 0.98). The study included 52 CREs from various specialties and 498 MSF respondents. The response rate was 84%. Low ratings and negative gaps between others' ratings and self-ratings were identified for both management and leadership performance areas. The developmental plans</i>	<i>An MSF process might in itself lead to development in administrative areas. However, MSF carried through as a single stand-alone procedure was not sufficient to foster plans for the development of leadership performance.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>examine the effects of an MSF process on developmental plans made by leaders of postgraduate medical education (PGME) in clinical departments.</i>		<i>departments, consultants and young doctors-in-training. The MSF process included individual feedback as well as guidance on drafting developmental plans for both the department and the CREs. Themes emerging in the developmental plans were analysed and compared with the areas in need of improvement identified by the MSF process.</i>	<i>mainly focused on management initiatives, whereas plans for the development of leadership performance were few. Areas rated low by all respondents were scarcely represented in CREs' developmental plans.</i>	
Massagli et al (2007) <sup>202</sup>	US	<i>To determine the feasibility and psychometric qualities of a 360-degree evaluation of physical medicine and rehabilitation (PM&amp;R) residents' competence.</i>	<i>56 residents</i>	<i>Nurses, allied health staff, and medical students completed a 12-item questionnaire after each PM&amp;R resident rotation from January 2002 to December 2004. The items were derived from five of the six competencies defined by the Accreditation Council for Graduate Medical Education (ACGME).</i>	<i>Nine hundred thirty evaluations of 56 residents were completed. The alpha reliability coefficient for the instrument was 0.89. Ratings did not vary significantly by resident gender. Senior residents had higher ratings than junior residents. A reliability of &gt;0.8 could be achieved by ratings from just five nurses or allied health staff, compared with 23 ratings from medical students. Factor analysis revealed all items clustered on one factor, accounting for 84% of the variance. In a subgroup of</i>	<i>Resident assessment tools should be valid, reliable, and feasible. This Web-based 360-degree evaluation tool is a feasible way to obtain reliable ratings from rehabilitation staff about resident behaviors. The assignment of higher ratings for senior residents than junior residents is evidence for the general validity of this 360-degree evaluation tool in the assessment of resident performance. Different rater groups may need distinct instruments</i>



Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
					<i>residents with low scores, raters were able to differentiate among skills.</i>	<i>based on the exposure of rater groups to various resident activities and behaviors</i>
McKay et al (2006) <sup>203</sup>	UK	<i>Clinical audit has a central role in the NHS clinical governance agenda and the professional appraisal of medical practitioners in the UK. However, concerns have been raised about the poor design and impact of clinical audit studies and the ability of practitioners to apply audit methods. One method of making informed judgements on audit performance is by peer review. In the west of Scotland a voluntary peer review model</i>	<i>1002 submissions by general practitioners in the UK</i>	<i>Participants submitted a criterion audit or significant event analysis in standard formats for review by two informed general practitioners (GPs) using appropriate instruments. Peer review outcome data and the professional status of doctors participating were generated by computer search. Differences in proportions of those gaining a satisfactory peer review for each group were calculated.</i>	<i>Of 1002 criterion audit submissions, 552 (55%) were judged to be satisfactory. GP registrars were significantly more likely than GP trainers (<math>P &lt; 0.001</math>) and other established GP groups (<math>P &lt; 0.001</math>) to gain a satisfactory peer review. GPs in non-training practices were less likely to achieve a satisfactory review than registrars (<math>P &lt; 0.001</math>) and colleagues in training practices (<math>P &lt; 0.001</math>). Of 883 SEA submissions, 541 (65%) were judged as satisfactory, with all groups gaining a similar proportion of satisfactory assessments, although GP registrars may have outperformed non-training practice GPs (<math>P = 0.05</math>).</i>	<i>A significant proportion of GPs may be unable to adequately apply audit methods, potentially raising serious questions about the effectiveness of clinical audit as a health care improvement policy in general medical practice.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>has been open to general practitioners since 1999, while general practice trainees are compelled to participate as part of summative assessment. The study aimed to compare the outcomes of peer review for two methods of audit undertaken by different professional and academic groups of doctors.</i>				
Murphy et al (2009) <sup>101</sup>	UK	<i>To investigate the reliability and feasibility of six potential workplace-based assessment methods in general practice training: criterion audit, multi-source feedback from clinical and</i>	<i>171 GP registrars and their trainers, drawn from nine deaneries (representing all four countries in the UK), participated.</i>	<i>Performance of GP registrars (trainees) was evaluated with each tool to assess the reliabilities of the tools and feasibility, given raters and number of assessments needed. Participant experience of process determined by questionnaire. The ability of each tool</i>	<i>Multi-source feedback from colleagues and patient feedback on consultations emerged as the two methods most likely to offer a reliable and feasible opinion of workplace performance. Reliability co-efficients of 0.8 were attainable with 41 CARE Measure patient questionnaires and six clinical and/or five non-clinical colleagues per doctor when assessed on two</i>	<i>The combination of patient and colleague views of doctors' performance, coupled with reliable competence measures, may offer a suitable evidence-base on which to monitor progress and completion of doctors' training in general practice.</i>



Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>non-clinical colleagues, patient feedback (the CARE Measure), referral letters, significant event analysis, and video analysis of consultations.</i>		<i>to differentiate between doctors (reliability) was assessed using generalisability theory. Decision studies were then conducted to determine the number of observations required to achieve an acceptably high reliability for "high-stakes assessment" using each instrument. Finally, descriptive statistics were used to summarise participants' ratings of their experience using these tools.</i>	<i>occasions. For the other four methods tested, 10 or more assessors were required per doctor in order to achieve a reliable assessment, making the feasibility of their use in high-stakes assessment extremely low. Participant feedback did not raise any major concerns regarding the acceptability, feasibility, or educational impact of the tools.</i>	
Nair et al (2008) <sup>204</sup>	AU	<i>To evaluate the feasibility, reliability and acceptability of the mini clinical evaluation exercise (mini-CEX) for performance assessment among international medical graduates (IMGs)</i>	<i>28 IMGs and 35 examiners</i>	<i>Observational study of 209 patient encounters involving 28 IMGs and 35 examiners at three metropolitan teaching hospitals in New South Wales, Victoria and Queensland, September–December 2006.</i>	<i>The reliability of the mini-CEX was estimated using generalisability (G) analysis, and its acceptability was evaluated by a written survey of the examiners and IMGs.</i>  <i>The G coefficient for eight encounters was 0.88, suggesting that the reliability of the mini-CEX was 0.90 for 10 encounters. Almost half of the IMGs (7/16) and most examiners (14/18) were satisfied with the mini-CEX as a learning tool. Most of the IMGs and examiners enjoyed the immediate feedback, which is a</i>	<i>The mini-CEX is a reliable tool for performance assessment of IMGs, and is acceptable to and well received by both learners and supervisors.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
					<i>strong component of the tool.</i>	
Paige et al (2008) <sup>172</sup>	US	<i>This study examined the effect of implementing a new preoperative briefing protocol on self- and peer-assessments of individual operating room (OR) teamwork behaviors</i>	<i>10 rural community hospital staff (one surgeon, nine OR staff)</i>	<i>From July 2006 to February 2007, OR teamwork performance at a rural community hospital was evaluated before and after training and implementation of the protocol. After each case, every member on the team completed a 360-degree type teamwork behavior evaluation containing both self- and peer-assessments using a six-point Likert type scale (1 = definitely no to 6 = definitely yes). Individual behavior change was measured using the mean scale score of pre and postprotocol assessments.</i>	<i>Statistical analysis included t test for both pre/post and self/peer differences. Data were available for one general surgeon and nine OR staff (pre = 20 cases, post = 16 cases). The pre-protocol self-assessment mean score was significantly higher than peer-assessment (5.63 vs 5.29, <math>P &lt; 0.0267</math>). Pre and post-protocol peer assessment mean scores revealed a statistically significant gain in teamwork behaviors. No difference was observed in post-assessment mean scores for self- and peer-assessments. Individuals overestimated their teamwork behaviors before protocol implementation. Using a preoperative protocol seems to improve OR staff teamwork behaviors and self-assessment accuracy.</i>	<i>The use of a 360-degree assessment method targeting specific, observable behaviors may be useful in evaluating team-based interventions and enhancing teamwork effectiveness.</i>
Papinczak et al (2007) <sup>190</sup>	AU	<i>The purpose of this study was to explore self-, peer-, and tutor assessment of performance in tutorials among first year medical students in a</i>	<i>125 first year medical students</i>	<i>One hundred and twenty-five students enrolled in the first year of the Bachelor of Medicine and Bachelor of Surgery Program at the University of Queensland were recruited to participate in a study of metacognition</i>	<i>Scores obtained from tutor assessment correlated poorly with self-assessment ratings (<math>r = 0.31-0.41</math>), with students consistently under-marking their own performance to a substantial degree. Students with greater self-efficacy, scored their PBL performance more highly. Peer-assessment was a slightly more</i>	<i>According to the results of this study, first-year medical students in a problem-based learning curriculum were better able to accurately judge the performance of their peers compared to their own performance. This study has shown that self-assessment of process is</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>problem-based learning curriculum.</i>		<i>and peer- and self-assessment. Both quantitative and qualitative data were collected from the assessment of PBL performance within the tutorial setting, which included elements such as responsibility and respect, communication, and critical analysis through presentation of a case summary. Self-, peer-, and tutor assessment took place concurrently.</i>	<i>accurate measure, with peer-averaged scores correlating moderately with tutor ratings initially (<math>r = 0.40</math>) and improving over time (<math>r = 0.60</math>). Students consistently over-marked their peers, particularly those with sceptical attitudes to the peer-assessment process. Peer over-marking led to less divergence from the tutor scoring than under-marking of one's own work.</i>	<i>not an accurate measure, in line with the majority of research in this domain. Nevertheless, it has an important role to play in supporting the development of skills in reflection and self-awareness.</i>
Ramsey et al (1993) <sup>175</sup>	US	<i>To assess the feasibility and measurement characteristics of ratings completed by professional associates to evaluate the performance of practicing physicians.</i>	<i>Physician-subjects were selected from among practicing internists in New York, New Jersey, and Pennsylvania who received American Board of Internal Medicine certification 5 to 15 years previously.</i>	<i>The clinical performance of physicians was evaluated using written questionnaires mailed to professional associates (physicians and nurses). Physician-associates were randomly selected from lists provided by both the subjects and medical supervisors, and detailed information was collected concerning the professional and social relationships between the associate and the subject. Responses</i>	<i>Physician performance as assessed by peers. Peer ratings are not biased substantially by the method of selection of the peers or the relationship between the rater and the subject. Factor analyses suggest a two-dimensional conceptualization of clinical skills: one factor represents cognitive and clinical management skills and the other factor represents humanistic qualities and management of psychosocial aspects of illness. Strong correlation coefficient between the peer ratings medical knowledge and examination</i>	<i>These findings suggest that it is feasible to obtain assessments from professional associates of practicing physicians in areas such as clinical skills, humanistic qualities, and communication skills. Using a shorter version of the questionnaire used in this study, peer ratings provide a practical method to assess clinical performance in areas such as humanistic qualities and communication skills that are difficult to assess with other measures.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
				<i>were analyzed to determine factors that affect ratings and measurement characteristics of peer ratings.</i>	<i>medical board examination scores (between 0.5 – 0.6) Low correlation between communication and interprofessional relationship abilities and their examination scores (&lt;0.15) Ratings from 11 peer physicians are needed to provide a reliable assessment in these two areas.</i>	
Ramsey et al (1996) <sup>205</sup>	US	<i>To address the feasibility of obtaining reliable evaluations of individual physicians from peer ratings undertaken at diverse hospitals.</i>	<i>Eleven hospitals in diverse locations in the United States were recruited to participate. With the aid of the hospitals' medical directors, up to 40 board-certified internists with admitting privileges were recruited per hospital.</i>	<i>Participating physicians provided demographic data about themselves and nominated physician-associates to do peer ratings. Between April 1993 and January 1994, the physicians were rated by their peers, who received a single mailing with no follow-up. The raters used a nine-point Likert scale for 11 cognitive and noncognitive categories. Administrative procedures were coordinated from the American Board of Internal Medicine. Chi-square, Student's t-test, and factor analysis using varimax rotation were used to analyze the</i>	<i>Of the 4,139 questionnaires that were mailed to peer raters, 3,005 (73%) were returned. Of the 228 physicians who were rated, 187 received ten or more usable ratings, which were used for further analysis. The findings confirmed the results of previous research: The highest mean rating was for the category of integrity, and the lowest was for the category of psychosocial aspects of care. Ten to 11 responses per physician were necessary to achieve a generalizability coefficient of .7. Nearly 90% of the variance in the ratings was accounted for by two factors, one representing cognitive and clinical management skills and the other, humanistic qualities. For 16 physicians (9%), the ratings of overall clinical skills were less than 7 on a scale from 1 (low) to</i>	<i>The peer raters' response rate and the analysis of the ratings suggest that the rating process is acceptable to physicians and that it is feasible to obtain reliable, multidimensional peer evaluations of individual physicians practicing in diverse clinical settings.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
				<i>results.</i>	<i>9 (high); their ratings for all individual cognitive and noncognitive categories were below the ratings of the other physicians.</i>	
Roark et al (2006) <sup>206</sup>	US	<i>The objectives of this study were to: 1) implement web-based instruments for assessing and documenting the general competencies of otolaryngology resident education, as outlined by the Accreditation Council of Graduate Medical Education (ACGME); and 2) examine the benefit and validity of this online system for measuring educational outcomes and for identifying insufficiencies in the training program as they</i>	<i>26 surgical residents</i>	<i>We developed an online assessment system for a surgical postgraduate education program and examined its feasibility, usability, and validity. Evaluations of behaviors, skills, and attitudes of 26 residents were completed online by faculty, peers, and nonphysician professionals during a 3-year period. Analyses included calculation and evaluation of total average performance scores of each resident by different evaluators. Evaluations were also compared with American Board of Otolaryngology-administered in-service examination (ISE) scores for each resident. Convergent validity was examined statistically by comparing ratings among the different evaluator types.</i>	<i>Questionnaires and software were found to be simple to use and efficient in collecting essential information. From July 2002 to June 2005, 1,336 evaluation forms were available for analysis. The average score assigned by faculty was 4.31, significantly lower than that by nonphysician professionals (4.66) and residents evaluating peers (4.63) (<math>P &lt; .001</math>), whereas scores were similar between nonphysician professionals and resident peers. Average scores between faculty and nonphysician groups showed correlation in constructs of communication and relationship with patients, but not in those of professionalism and documentation. Correlation was observed in respect for patients but not in medical knowledge between faculty and resident peer groups. Resident ISE scores improved in the third year of the study and demonstrated high correlation with faculty perceptions of medical knowledge (<math>r = 0.65</math>, <math>P</math></i>	<i>Compliance for completion of forms was 97%. The system facilitated the educational management of our training program along multiple dimensions. The small perceptual differences among a highly selected group of residents have made the unambiguous validation of the system challenging. The instruments and approach warrant further study. Improvements are likely best achieved in broad consultation among other otolaryngology programs. copyright 2006 The American Laryngological, Rhinological and Otological Society, Inc.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
Stark et al (2008) <sup>207</sup>	US	<i>Professionalism is identified as a competency of resident education. Best approaches to teaching and evaluating professionalism are unknown, but feedback about professionalism is necessary to change practice and behavior. Faculty discomfort with professionalism may limit their delivery of feedback to residents.</i>	15 faculty members.	<i>A pilot program to implement a 360-degree evaluation of observable professional behaviors and determine how its use impacts faculty feedback to residents. Internal Medicine (IM) residents were evaluated during ambulatory rotations using a 360-degree assessment of professional behaviors developed by the National Board of Medical Examiners(R). Faculty used evaluation results to provide individual feedback to residents.</i>	<i>Faculty completed pre- and post-intervention surveys. Using a 7-point Likert scale, faculty reported increased skill in giving general feedback (4.85 vs 4.36, <math>p &lt; .05</math>) and feedback about professionalism (4.71 vs 3.57, <math>p &lt; .01</math>) after the implementation of the 360-degree evaluation. They reported increased comfort giving feedback about professionalism (5.07 vs 4.35, <math>p &lt; .05</math>) but not about giving feedback in general (5.43 vs 5.50).</i>	<i>A 360-degree professionalism evaluation instrument used to guide feedback to residents improves faculty comfort and self-assessed skill in giving feedback about professionalism.</i>
Staton et al (2007) <sup>208</sup>	US	<i>The Accreditation Council on Graduate Medical Education (ACGME) supports chart audit as a method to track competency in Practice-Based</i>	Residents	<i>A retrospective electronic chart review was performed on 347 patients with diabetes mellitus cared for by internal medicine residents in a university-based continuity clinic from May 2003 to September 2004. Residents abstracted</i>	<i>Significant improvement in the documentation of foot exams was observed over the course of the study. The percentage of patients receiving neurological, vascular, and skin exams increased by 20% (from 13% to 33%) (<math>p = 0.001</math>), 26% (from 45% to 71%) (<math>p &lt; 0.001</math>), and 18% (51%-72%) (<math>p = 0.005</math>), respectively. Similarly, the proportion of patients receiving</i>	<i>Peer chart audits performed by residents in the absence of formal feedback were associated with improved documentation of the foot exam in patients with diabetes mellitus. Although this study suggests that peer chart audits may be an effective tool to improve practice-based learning</i>



Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>Learning and Improvement. We examined whether peer chart audits performed by internal medicine residents were associated with improved documentation of foot care in patients with diabetes mellitus.</i>		<i>information pertaining to documentation of foot examinations (neurological, vascular, and skin) from the charts of patients followed by their physician peers. No formal feedback or education was provided.</i>	<i>a well-documented exam which includes all three components - neurological, vascular and skin foot exam - increased over time (6% to 24%, <math>p &lt; 0.001</math>).</i>	<i>and documentation of foot care in diabetic patients, evaluating the actual performance of clinical care was beyond the scope of this study and would be better addressed by a randomized controlled trial</i>
Violato et al (2008) <sup>131</sup>	Canada	<i>Multi-source feedback (MSF) enables performance data to be provided to doctors from patients, co-workers and medical colleagues. This study examined the evidence for the validity of MSF instruments for general practice, investigated changes in</i>	<i>250 doctors, 25 patients, eight co-workers and eight medical colleagues.</i>	<i>Data for 250 doctors included three datasets per doctor from, respectively, 25 patients, eight co-workers and eight medical colleagues, collected on two occasions.</i>	<i>There was high internal consistency (<math>\alpha &gt; 0.90</math>) and adequate generalisability (<math>E_p(2) &gt; 0.70</math>). D study results indicate adequate generalisability coefficients for groups of eight assessors (medical colleagues, co-workers) and 25 patient surveys. Confirmatory factor analyses provided evidence for the validity of factors that were theoretically expected, meaningful and cohesive. Comparative fit indices were 0.91 for medical colleague data, 0.87 for co-worker data and 0.81 for patient data. Paired t-test analysis showed significant change between the two assessments from medical</i>	<i>There is evidence for the construct validity of the instruments and for their stability over time. Upward changes in performance will occur, although their effect size is likely to be small to moderate.</i>

Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>performance for doctors who participated twice, 5 years apart, and determined the association between change in performance and initial assessment and socio-demographic characteristics.</i>			<i>colleagues and co-workers, but not between the two patient surveys. Multiple linear regressions explained 2.1% of the variance at time 2 for medical colleagues, 21.4% of the variance for co-workers and 16.35% of the variance for patient assessments, with professionalism a key variable in all regressions.</i>	
Whitehouse et al (2007) <sup>22</sup>	UK	<i>This study was to see if the team assessment of behaviours (TAB) 360 degree assessment tool was able to identify interpersonal behaviour problems in doctors in training, to see if feedback was useful, to gauge the value of the process by those involved, and to learn</i>	<i>171 Senior Health Officers</i>	<i>TAB was administered to assess interpersonal behaviours of senior house officers in four hospitals in the West Midlands, UK. In addition, questionnaires were sent to all participants, some were interviewed about the whole process, and records kept of the time involved.</i>	<i>One hundred and seventy-one SHO volunteers received 1378 assessments. The median number of ratings per SHO was 8 (mode 9). Sixty-four percent of SHOs received 'no concern' ratings in all four behaviours (domains) assessed. Twenty-one percent received one 'some concern' rating. Fifteen percent received more than one 'concern' rating</i>	<i>Assessors and trainees found the process practical, valuable and fair. Educational supervisors found it valuable, although only 23% learned something new about their trainees. Clinical tutors valued the system. Administrative staff found it time consuming. The TAB four-domain rating form with its single pass category identified specific concern about volunteer trainees' professional behaviour. Not all trainees received skilled feedback.</i>



Author	Country	Purpose	Participants	Design and method	Outcome measures and results	Conclusion
		<i>lessons about implementing the process for the future.</i>				
Wilkinson et al (2008) <sup>181</sup>	UK	<i>To evaluate the reliability and feasibility of assessing the performance of medical specialist registrars (SpRs) using three methods: the mini-clinical evaluation exercise (mini-CEX), directly observed procedural skills (DOPS) and multi-source feedback (MSF) to help inform annual decisions about the outcome of SpR training.</i>	<i>A total of 230 SpRs (from 17 specialties) in 58 UK hospitals took part from 2003 to 2004.</i>	<i>We conducted a feasibility study and generalisability analysis based on the application of these assessment methods and the resulting data. Main outcome measures included: time taken for each assessment, and variance component analysis of mean scores and derivation of 95% confidence intervals for individual doctors' scores based on the standard error of measurement. Responses to direct questions on questionnaires were analysed, as were the themes emerging from open-comment responses.</i>	<i>The methods can provide reliable scores with appropriate sampling. In our sample, all trainees who completed the number of assessments recommended by the Royal Colleges of Physicians had scores that were 95% certain to be better than unsatisfactory. The mean time taken to complete the mini-CEX (including feedback) was 25 minutes. The DOPS required the duration of the procedure being assessed plus an additional third of this time for feedback. The mean time required for each rater to complete his or her MSF form was 6 minutes.</i>	<i>This is the first attempt to evaluate the use of comprehensive workplace assessment across the medical specialties in the UK. The methods are feasible to conduct and can make reliable distinctions between doctors' performances. With adaptation, they may be appropriate for assessing the workplace performance of other grades and specialties of doctor. This may be helpful in informing foundation assessment.</i>

## **Appendix E: Bibliography and abstracts**

## Selected bibliography and abstracts

Abdulla, A. (2008). "A critical analysis of mini peer assessment tool (mini-PAT).[see comment]." *Journal of the Royal Society of Medicine* 101(1): 22-6.

The structured evaluation of doctors' performance through peer review is a relatively new phenomenon brought about by public demand for accountability to patients. Medical knowledge (as assessed by examination score) is no longer a good predictor of individual performance, humanistic qualities and communication skills. The process of peer review (or multi-source assessment) was developed over the last two decades in the USA and has started to pick up momentum in the UK through the introduction of Modernizing Medical Careers. However the concept is not new. Driven by market forces, it was initially developed by industrial organizations to improve leadership qualities with a view to increasing productivity through positive behaviour change and self-awareness. Multi-source feedback is not without its problems and may not always produce its desired outcomes. In this article we review the evidence for peer review and critically discuss the current process of mini peer assessment tool (mini-PAT) as the assessment tool for peer review employed in UK. [References: 28]

Adler, N. E. and A. Milstein (1983). "Evaluating the impact of physician peer review: Factors associated with successful PSROs." *American Journal of Public Health* 73(10): 1182-1185.

A five-component measurement method was developed and applied to the 1981 impact statements of 30 Professional Standards Review Organizations (PSROs) by four blind raters familiar with the PSRO program. High inter-rater reliability (.95) was achieved. Rater's scores for each PSRO were then averaged and regressed against five variables predicted to affect PSRO impact: geographical density of PSROs; PSRO affiliation with a medical society; surgical necessity review; use of data profiles; and pre-existing Medicare hospitalization rates. As a set, the variables accounted for 44% of the variance in PSRO performance ( $p < .05$ ). When entered in stepwise regression, geographical density and use of surgical necessity review accounted for the largest share of the variance. The findings are believed to reflect the recency of PSRO motivation to demonstrate significant impact, and the value of surgical necessity review as an indicator of PSRO courage to risk unpleasant backlash from their medical communities.

Agee, C. (2007). "Improving the peer review process: develop a professional review committee for better and quicker results." *Healthcare Executive* 22(3): 72-3.

Alderson, C. J. and P. Hogg (2003). "Advanced radiographic practice - The legal aspects." *Radiography* 9(4): 305-314.

Allied health and nursing professionals are continuing to expand their responsibilities into clinical areas outside their traditional spheres of interest; typically, many of these new responsibilities are found within the medical (doctor) domain. Such responsibilities are often at an advanced clinical level and consequently higher demands are placed upon the professionals, not least in terms of clinical updating, competence to practice and also legal liability. This article explores the legal implications of practising at an advanced clinical level with particular reference to legal claims. The first part of the article commences with an outline of pertinent law in England and Wales. The latter part of the article explores actual cases from which allied health professionals (eg radiographers) can gain valuable information. Throughout the article suggestions for good practice are indicated. Examples of good practice include: the need to base your practice on evidence and peer practice; the need to keep detailed records (protocols) of such practice; the need to know when you are at the limit of your ability; and as such when to ask for advice from a medical practitioner/radiologist. copyright 2003 The College of Radiographers. Published by Elsevier Ltd. All rights reserved.

Allerup, P., K. Aspegren, et al. (2007). "Use of 360-degree assessment of residents in internal medicine in a Danish setting: a feasibility study." *Medical Teacher* 29(2-3): 166-70.

**BACKGROUND:** The aim of the study was to explore the feasibility of 360 degree assessment in early specialist training in a Danish setting. Present Danish postgraduate training requires assessment of specific learning objectives. Residency in Internal Medicine was chosen for the study. It has 65 learning objectives to be assessed. We considered 22 of these suitable for assessment by 360-degrees assessment. **METHODS:** Medical departments of six hospitals contributed 42 interns to the study. Each resident was assessed by ten persons of whom one was a secretary, four were nurses and five senior doctors. The assessors spent 14.5 minutes (median) to fill in the forms. **RESULTS:** Of the 22 chosen objectives, 15 could reliably be assessed by doctors, 7 by nurses and none by

secretaries. CONCLUSIONS: The method was practical in busy clinical departments and was well accepted by the assessors. Reliability of the method was acceptable. It discriminated satisfactorily between the good and not so good performers.

Archer, J., J. Norcini, et al. (2008). "mini-PAT (Peer Assessment Tool): a valid component of a national assessment programme in the UK?" *Advances in Health Sciences Education* 13(2): 181-92.  
PURPOSE: To design, implement and evaluate a multisource feedback instrument to assess Foundation trainees across the UK. METHODS: mini-PAT (Peer Assessment Tool) was modified from SPRAT (Sheffield Peer Review Assessment Tool), an established multisource feedback (360 degrees ) instrument to assess more senior doctors, as part of a blueprinting exercise of instruments suitable for assessment in Foundation programmes (first 2 years postgraduation). mini-PAT's content validity was assured by a mapping exercise against the Foundation Curriculum. Trainees' clinical performance was then assessed using 16 questions rated against a six-point scale on two occasions in the pilot period. Responses were analysed to determine internal structure, potential sources of bias and measurement characteristics. RESULTS: Six hundred and ninety-three mini-PAT assessments were undertaken for 553 trainees across 12 Deaneries in England, Wales and Northern Ireland. Two hundred and nineteen trainees were F1s or PRHOs and 334 were F2s. Trainees identified 5544 assessors of whom 67% responded. The mean score for F2 trainees was 4.61 (SD = 0.43) and for F1s was 4.44 (SD = 0.56). An independent t test showed that the mean scores of these 2 groups were significantly different ( $t = -4.59$ ,  $df = 390$ ,  $p < 0.001$ ). 43 F1s (19.6%) and 19 F2s (5.6%) were assessed as being below expectations for F2 completion. The factor analysis produced 2 main factors, one concerned clinical performance, the other humanistic qualities. Seventy-four percent of F2 trainees could have been assessed by as few as 8 assessors (95% CI +/-0.6) as they either scored an overall mean of 4.4 or above or 3.6 and below. Fifty-three percent of F1 trainees could have been assessed by as few as 8 assessors (95% CI +/-0.5) as they scored an overall mean of 4.5 or above or 3.5 and below. The hierarchical regression when controlling for the grade of trainee showed that bias related to the length of the working relationship, occupation of the assessor and the working environment explained 7% of the variation in mean scores when controlling for the year of the Foundation Programme (R squared change = 0.06, F change = 8.5, significant F change <0.001). CONCLUSIONS: As part of an assessment programme, mini-PAT appears to provide a valid way of collating colleague opinions to help reliably assess Foundation trainees.

Armson, H., S. Kinzie, et al. (2007). "Translating learning into practice: Lessons from the practice-based small group learning program." *Canadian Family Physician* 53(9): 1477-1485.  
PROBLEM ADDRESSED: The need for effective and accessible educational approaches by which family physicians can maintain practice competence in the face of an overwhelming amount of medical information. OBJECTIVE OF PROGRAM: The practice-based small group (PBSG) learning program encourages practice changes through a process of small-group peer discussion - identifying practice gaps and reviewing clinical approaches in light of evidence. PROGRAM DESCRIPTION: The PBSG uses an interactive educational approach to continuing professional development. In small, self-formed groups within their local communities, family physicians discuss clinical topics using prepared modules that provide sample patient cases and accompanying information that distils the best evidence. Participants are guided by peer facilitators to reflect on the discussion and commit to appropriate practice changes. CONCLUSION: The PBSG has evolved over the past 15 years in response to feedback from members and reflections of the developers. The success of the program is evidenced in effect on clinical practice, a large and increasing number of members, and the growth of interest internationally.

Arnold, L. (2002). "Assessing professional behavior: Yesterday, today, and tomorrow." *Academic Medicine* 77(6): 502-515.

Purpose. The author interprets the state of the art of assessing professional behavior. She defines the concept of professionalism, reviews the psychometric properties of key approaches to assessing professionalism, conveys major findings that these approaches produced, and discusses recommendations to improve the assessment of professionalism. Method. The author reviewed professionalism literature from the last 30 years that had been identified through database searches; included in conference proceedings, bibliographies, and reference lists; and suggested by experts. The cited literature largely came from peer-reviewed journals, represented themes or novel approaches, reported qualitative or quantitative data about measurement instruments, or described pragmatic or theoretical approaches to assessing professionalism. Results. A circumscribed concept of professionalism is available to serve as a foundation for next steps in assessing professional

behavior. The current array of assessment tools is rich. However, their measurement properties should be strengthened. Accordingly, future research should explore rigorous qualitative techniques; refine quantitative assessments of competence, for example, through OSCEs; and evaluate separate elements of professionalism. It should test the hypothesis that assessment tools will be better if they define professionalism as behaviors expressive of value conflicts, investigate the resolution of these conflicts, and recognize the contextual nature of professional behaviors. Whether measurement tools should be tailored to the stage of a medical career and how the environment can support or sabotage the assessment of professional behavior are central issues. Final thought. Without solid assessment tools, questions about the efficacy of approaches to educating learners about professional behavior will not be effectively answered.

Arnold, L., C. K. Shue, et al. (2007). "Can there be a single system for peer assessment of professionalism among medical students? A multi-institutional study." *Academic Medicine* 82(6): 578-586.

**PURPOSE:** Peer assessment is a valuable source of information about medical students' professionalism. How best to facilitate peer assessment of students' professional behavior remains to be answered, however. This report extends previous research through a multi-institutional study of students' perspectives about system characteristics for peer assessment of professionalism. It examines whether students from different schools and year levels prefer different characteristics of peer assessment to assess each other candidly, or whether a single system can be designed. It then identifies the characteristics of the resulting preferred system(s). **METHOD:** At the beginning of academic year 2004-2005, students (1,661 of 2,115; 78%) in years one through four at four schools replied to a survey about which peer assessment characteristics-related to, for example, who receives the assessment, its anonymity, and timing-would prevent or encourage their participation. Multivariate analysis of variance was used to detect differences among institutions and students from each year level. **RESULTS:** Students across year levels and schools generally agreed about the characteristics of peer assessment. They prefer a system that is 100% anonymous, provides immediate feedback, focuses on both unprofessional and professional behaviors, and uses peer assessment formatively while rewarding exemplary behavior and addressing serious repetitive professional lapses. The system, they emphasize, must be embedded in a supportive environment. **CONCLUSIONS:** Students' agreement about peer-assessment characteristics suggests that one system can be created to meet the majority of students' preferences. Once implemented, the system should be monitored for student acceptability to maximize participation and to determine the formative and summative value of the process. Copyright 2007 Association of American Medical Colleges.

Aronson, M., R. Lawrence, et al. (1982). "Peer review in a primary care education program." *Journal of Medical Education* 57(6): 481-483.

The peer review conference serves many functions. The frequently noted objections to peer review among physicians, by and large, have not been a problem because of the conviction that feedback is important for both the monitoring of care and for faculty and resident education. Other primary care practices and educational programs may wish to consider using this model, as opposed to more coercive peer review measures or more didactic teaching conferences.

Atwater, L. E., J. F. Brett, et al. (2007). "Multisource feedback: lessons learned and implications for practice." *Human Resource Management* 46(2): 285-307.

Organizations around the world are using multisource, or 360-degree, feedback. Although many HR practitioners embrace it as an important mechanism for leadership development, organizations must attend to and address several issues in order to maximize the utility of multisource feedback (MSF). We discuss current research findings and highlight issues for managers to consider both before starting a multisource feedback process and after the feedback is given, plus we review potential outcomes of the process. We also describe lessons learned from an intensive three-year investigation of an MSF implementation in two organizations. © 2007 Wiley Periodicals, Inc.

Baigent, M. L. and R. Harding (2007). "IRQN AWARD Article: Radiology Audit: A Program Developed for Remote or Solo-Practicing Radiologists: A New Zealand Quality Initiative." *JACR Journal of the American College of Radiology* 4(11): 841-843.

New Zealand radiologists working in solo or locum tenens practice away from colleagues in hospital departments or at large practices may have problems satisfying peer review and audit requirements for the Medical Council of New Zealand and Royal Australian and New Zealand College of Radiologists. This paper describes a practical audit program developed by a small group of New

Zealand radiologists that commenced in 2005. Each year, a set of cases with known clinical and pathologic outcomes is sent from one radiologist to the next on a rota system. Reports are e-mailed back, then audited, with comments being returned again by e-mail. Each radiologist therefore acts as both reporter and auditor, ensuring a spectrum of audit around the group. The cases are reported blind, but usually, a limited amount of clinical information is provided, as would be the case on a request form in clinical practice. The process of audit is described, and the reaction to and acceptance of the program by the participants have been surveyed at the end of the first year of operation. The program has achieved overall acceptance by the group and is approved by the continuing medical education committee of the college, hence satisfying the Medical Council of New Zealand.

Balla, M., B. Knothe, et al. (1996). "Group peer review in psychiatry: The relationship to quality improvement and quality care." *Australian and New Zealand Journal of Psychiatry* 30(5): 653-659.  
Objective: This study sought to elucidate the contribution of peer review groups involving psychiatrists to quality improvement and quality care. Method: Audio-taped interviews of groups engaged in peer review were analysed using a qualitative methodology. Participants' views of the ways in which they experienced and conceptualised peer review were explored. Results: The views of participants in peer review groups were analysed, and categories evolved which identified differences in how they perceived the structure and function of group peer review. Conclusions: Participants in the groups studied perceived peer review as a professional growth forum within a quality improvement framework providing critical review of treatment, continuing education, and a sense of collegiality. Boundaries of acceptable practice were tested and defined. At its best, participation in peer review groups enhanced reflective practice which achieved new understandings of clinical work. In this regard, peer review is seen as a highly desirable method for the maintenance of professional standards.

Beatson, J., N. Rushford, et al. (1996). "Group peer review: A questionnaire-based survey." *Australian and New Zealand Journal of Psychiatry* 30(5): 643-652.

Objective: This paper presents the findings from a questionnaire-based survey of psychiatrists designed to elucidate the positive and negative aspects of group peer review and its perceived place in accountability procedures, and to provide information about how accountability through group peer review might be improved. Method: Three hundred and eighty-eight psychiatrists were surveyed via mail-out questionnaire. Demographic data, details of groups, and perceptions of beneficial and detrimental effects of group peer review were sought from group participants and non-participants. Attitudes of participants were compared with those of non-participants. Features of groups related to satisfaction in participants were examined. Results: The majority of the 170 respondents participating in groups regarded peer review as a means of maintaining and improving skills, sharing ideas and methods, receiving constructive criticism and feedback, of educational benefit and an important source of professional accountability. Non-participants, while less positive overall, responded equally that participation in peer review groups would be an effective response to accountability procedures. Potential detrimental effects and problems with the functioning of peer review groups were elucidated. Conclusions: Group peer review contributes significantly to professional accountability and education in well-functioning groups. Further strategies for the facilitation of group functioning and for the processing of problems arising in group peer review need to be developed to optimise its contribution to the maintenance and improvement of professional standards.

Beckett, P. A. (2008). "Lung cancer peer review survey." *Thorax* 63(3): 290-1.

Becton, J. B. and M. Schraeder (2004). "Participant Input into Rater Selection: Potential Effects on the Quality and Acceptance of Ratings in the Context of 360-Degree Feedback." *Public Personnel Management* 33(1): 23-32.

This article explores implications and potential benefits of giving individuals an opportunity to have input into the selection of raters involved in giving them 360 degree feedback. First, allowing participants some degree of input, but not total input into the selection of raters, may result in better perceptions of fairness, rater credibility and usefulness of the feedback. In addition, increased participation in the selection of raters could translate into the development of more challenging goals and higher levels of job performance. From a quantitative perspective, this could also produce more variance in the ratings.

Blesch, G. (2007). "Peer-review showdown. High court review could lead to opening records." *Modern Healthcare* 37(48): 10.

Borel-Rinkes, I. H. M., D. J. Gouma, et al. (2008). "Surgical training in the Netherlands." *World Journal of Surgery* 32(10): 2172-2177.

Surgical training in the Netherlands has traditionally been characterized by learning on the job under the classic master-trainee doctrine. Over the past decades, it has become regionally organized with intensive structural training courses, and a peer-based quality control system. Recently, the nationwide programme has been modernized further and now involves a systematic, competency-based education with structural training courses, formalized assessment and room for reflection by residents under the supervision of surgical teaching groups. To this end, a uniform web-based digital portfolio is being introduced to facilitate monitoring of the individual resident's progress. Though requiring inspirational leadership, commitment, and determination, this modernization has sparked enthusiasm among trainees and teachers.

Borgstede, J. P., R. S. Lewis, et al. (2004). "RADPEER quality assurance program: A multifacility study of interpretive disagreement rates." *JACR Journal of the American College of Radiology* 1(1): 59-65.

Purpose: To develop and test a radiology peer review system that adds minimally to workload, is confidential, uniform across practices, and provides useful information to meet the mandate for "evaluation of performance in practice" that is forthcoming from the American Board of Medical Specialties as one of the four elements of maintenance of certification. Method: RADPEER has radiologists who review previous images as part of a new interpretation record their ratings of the previous interpretations on a 4-point scale. Reviewing radiologists' ratings of 3 and 4 (disagreements in nondifficult cases) are reviewed by a peer review committee in each practice to judge whether they are misinterpretations by the original radiologists. Final ratings are sent for central data entry and analysis. A pilot test of RADPEER was conducted in 2002. Results: Fourteen facilities participated in the pilot test, submitting a total of 20,286 cases. Disagreements in difficult cases (ratings of 2) averaged 2.9% of all cases. Committee-validated misinterpretations in nondifficult cases averaged 0.8% of all cases. There were considerable differences by modality. There were substantial differences across facilities; few of these differences were explicable by mix of modalities, facility size or type, or being early or late in the pilot test. Of 31 radiologists who interpreted over 200 cases, 2 had misinterpretation rates significantly ( $P < .05$ ) above what would be expected given their individual mix of modalities and the average misinterpretation rate for each modality in their practice. Conclusions: A substantial number of facilities participated in the pilot test, and all maintained their participation throughout the year. Data generated are useful for the peer review of individual radiologists and for showing differences by modality. RADPEER is now operational and is a good solution to the need for a peer review system with the desirable characteristics listed above. Copyright 2004 American College of Radiology.

Borgstede, J. P. and P. A. Wilcox (2007). "Quality care and safety know no borders." *Biomedical Imaging and Intervention Journal* 3(3)(e34).

The public, governmental agencies and payers expect medical professional organisations to develop practice guidelines and technical standards. The American College of Radiology proactively addresses these topics as well as other quality and safety interests including appropriateness criteria and accreditation. The College is also actively involved in development of a national radiology data base to collect data regarding quality and safety metrics in multiple areas. In addition, the College has developed RADPEER[trademark], a simple, cost-effective process that allows peer review to be performed during the routine interpretation of current images. This paper discusses the efforts of the ACR in all of these areas. Copyright 2007 Biomedical Imaging and Intervention Journal. All rights reserved.

Boulet, L. P., E. Dorval, et al. (2008). "Towards excellence in asthma management: Final report of an eight-year program aimed at reducing care gaps in asthma management in Quebec." *Canadian Respiratory Journal* 15(6): 302-310.

Background and objectives: Asthma care in Canada and around the world persistently falls short of optimal treatment. To optimize care, a systematic approach to identifying such shortfalls or 'care gaps', in which all stakeholders of the health care system (including patients) are involved, was proposed. Methods: Several projects of a multipartner, multidisciplinary disease management program, developed to optimize asthma care in Quebec, was conducted in a period of eight years. First, two population maps were produced to identify regional variations in asthma-related morbidity and to prioritize interventions for improving treatment. Second, current care was evaluated in a

physician-patient cohort, confirming the many care gaps in asthma management. Third, two series of peer-reviewed outcome studies, targeting high-risk populations and specific asthma care gaps, were conducted. Finally, a process to integrate the best interventions into the health care system and an agenda for further research on optimal asthma management were proposed. Results: Key observations from these studies included the identification of specific patterns of noncompliance in using inhaled corticosteroids, the failure of increased access to spirometry in asthma education centres to increase the number of education referrals, the transient improvement in educational abilities of nurses involved with an asthma hotline telephone service, and the beneficial effects of practice tools aimed at facilitating the assessment of asthma control and treatment needs by general practitioners. Conclusions: Disease management programs such as Towards Excellence in Asthma Management can provide valuable information on optimal strategies for improving treatment of asthma and other chronic diseases by identifying care gaps, improving guidelines implementation and optimizing care. copyright 2008 Pulsus Group Inc. All rights reserved.

Boulet, L.-P., E. Dorval, et al. (2008). "Towards Excellence in Asthma Management: final report of an eight-year program aimed at reducing care gaps in asthma management in Quebec." *Canadian Respiratory Journal* 15(6): 302-10.

**BACKGROUND AND OBJECTIVES:** Asthma care in Canada and around the world persistently falls short of optimal treatment. To optimize care, a systematic approach to identifying such shortfalls or 'care gaps', in which all stakeholders of the health care system (including patients) are involved, was proposed. **METHODS:** Several projects of a multipartner, multidisciplinary disease management program, developed to optimize asthma care in Quebec, was conducted in a period of eight years. First, two population maps were produced to identify regional variations in asthma-related morbidity and to prioritize interventions for improving treatment. Second, current care was evaluated in a physician-patient cohort, confirming the many care gaps in asthma management. Third, two series of peer-reviewed outcome studies, targeting high-risk populations and specific asthma care gaps, were conducted. Finally, a process to integrate the best interventions into the health care system and an agenda for further research on optimal asthma management were proposed. **RESULTS:** Key observations from these studies included the identification of specific patterns of noncompliance in using inhaled corticosteroids, the failure of increased access to spirometry in asthma education centres to increase the number of education referrals, the transient improvement in educational abilities of nurses involved with an asthma hotline telephone service, and the beneficial effects of practice tools aimed at facilitating the assessment of asthma control and treatment needs by general practitioners. **CONCLUSIONS:** Disease management programs such as Towards Excellence in Asthma Management can provide valuable information on optimal strategies for improving treatment of asthma and other chronic diseases by identifying care gaps, improving guidelines implementation and optimizing care. [References: 36]

Bowie, P., J. McKay, et al. (2008). "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* 14(6): 1038-1043.

**Rationale, aims and objectives:** Clinical audit informs general practitioner (GP) appraisal and will provide evidence of performance for revalidation in the UK. However, objective evidence is now required. An established peer assessment system may offer an educational solution for making objective judgements on clinical audit quality. National Health Service (NHS) clinical audit specialists could potentially support this system if their audit assessments were comparable with established medical peer assessors. The study aimed to quantify differences between clinical audit specialists and medical peer assessors in their assessments of clinical audit projects. **Methods:** A comparison study of the assessment outcomes of clinical audit reports by two groups using appropriate assessment instruments was conducted. Mean scores were compared and 95% confidence intervals (CIs) and limits of agreement calculated. A two-point mean difference would be relevant. **Results:** Twelve significant event analysis (SEA) reports and 12 criterion audit projects were assessed by 11 experienced GP assessors and 10 NHS audit specialist novice assessors. For SEA, the mean score difference between groups was <1.0. The 95% CI for bias was -0.1 to 0.5 (P = 0.14). Limits of agreement ranged from -0.7 to 1.2. For criterion audit, a mean score difference of [less-than or equal to]1.0 was calculated for seven projects and scores between 1.1 and 1.9 for four. The 95% CI for bias was 0.8 to 1.5 (P < 0.001). Limits of agreement ranged from -2.5 to -0.0. **Conclusions:** The study findings suggest that a sample of NHS clinical audit specialists can give numerically accurate feedback scores to GPs on the quality of their clinical audit activity compared with established peer assessors as part of the model outlined. Copyright 2008 The Authors.



Bradshaw, R. W. (2000). "Using peer review for self-audits of medical record documentation." *Family Practice Management* 7(4): 28.

You might be surprised at how much doctors can help one another improve in coding and documentation.

Brett, J. F. and L. E. Atwater (2001). "360 degree feedback: accuracy, reactions, and perceptions of usefulness." *Journal of Applied Psychology* 86(5): 930-42.

This study examined how 360 degree feedback ratings and self-other rating discrepancies related to reactions to feedback, perceptions of feedback accuracy, perceived usefulness of the feedback, and recipients' receptivity to development. The results indicated that less favourable ratings were related to beliefs that feedback was less accurate and to negative reactions. Negative reactions and perceptions that feedback was less accurate were related to beliefs that the feedback was less useful. Those who found feedback less useful were perceived by a facilitator as less development-focused. Goal orientation did not moderate the relationship between ratings and perceptions of accuracy or reactions to feedback. Goal orientation was related to perceptions of usefulness of the process several weeks after receipt of feedback. The results question widely held assumptions about 360 degree feedback that negative and discrepant feedback motivates positive change.

Bucknall, V., E. M. Sobic, et al. (2008). "Peer assessment of resuscitation skills." *Resuscitation* 77(2): 211-215.

Introduction: Peer tuition has been identified as a useful tool for delivering undergraduate healthcare training in basic life support. The aim of this study was to test the expansion of the peer tuition model to include peer assessment of performance. The study also sought to establish the attitudes towards peer assessment among the course students and tutors. Methods: Students undergoing an end-of-course test in basic life support were simultaneously assessed by peer and faculty assessors, and the reliability of assessment results was measured. Students' and peer assessors' attitudes to peer assessment were also measured, by questionnaire. Results: In all 162 candidates were assessed by 9 sets of peers and faculty examiners. Inter-observer agreement was high (>95%) for all assessment domains apart from chest compressions (93%). Agreement on the final pass/fail decision was less consistent at 86%, because of the lower pass rate of 71% (115/162) afforded by peer assessors compared with 82% (132/162) by faculty assessors ( $p = 0.0008$ ). Peer assessor sensitivity and specificity were 85% and 90%, respectively, with positive predictive value of 97% and negative predictive value of 57%. Conclusion: Senior healthcare students can make reliable assessments of their peers' performance during an end-of-course test in basic life support. Students preferred peer assessment, and the peer assessment process was acceptable to the majority of students and peer assessors.

Bye, A. M. E., A. M. Connolly, et al. (2007). "A triangulated approach to the assessment of teaching in childhood epilepsy." *Medical Teacher* 29(2-3): 255-257.

Background: A comprehensive methodology is needed to assess student teaching. The present study employed a triangulated approach evaluating participant perceptions of learning, critical reflection by the lecturer and peer observation to measure confidence, interest and usefulness of the subject matter. Methods: Using an interactive lecturing style, seven teaching sessions were delivered to medical students and junior doctors. Rating scales, open-ended questions and focus group discussions evaluated participant perceptions. Critical reflections and observations were made by the lecturer and independent learning consultants. Results: Seventy per cent of participants rated the lecture on the highest scale for usefulness and interest. There was a significant post-lecture increase in clinical confidence in seizure identification ( $p < 0.0005$ ). Open-ended questions showed that videos were most useful (81/149) and interesting (109/149), and that the presentation of the syndromal classification provided a useful approach (114/149). Focus group discussion, lecturer and peer observation cross-validated these findings and highlighted the importance of expert commentary to the videos and the clinical relevance of material.

Calhoun, A. W., E. A. Rider, et al. (2009). "Assessment of communication skills and self-appraisal in the simulated environment: feasibility of multirater feedback with gap analysis." *Simulation in Healthcare: The Journal of The Society for Medical Simulation* 4(1): 22-9.

INTRODUCTION: Multirater assessment is a powerful means of measuring communication skills. The use of gap analysis to assess self-appraisal is a strength of this technique. On the basis of Kalamazoo Consensus Statement framework and 360-degree assessment models, we developed a

multirater instrument with gap analysis, with the goals of examining both communication skills and situational self-appraisal, and assessing the feasibility of the combined approach. **METHODS:** The multirater communication skills instrument was used to assess Pediatric and Neonatal Intensive Care fellows after participation in seven simulated family meetings. Instrument reliability was determined using Cronbach's Alpha and Factorial Analysis. Correlations between rater groups were examined with Spearman's Rank Coefficient. Gap analyses and rater perceptions of the instruments were analyzed using descriptive statistics. **RESULTS:** Seven pediatric intensive care unit and neonatal intensive care fellows were each assessed by 11 to 18 raters (108 total assessments). Correlations were identified between disciplinary groups within each encounter. Among the 7 fellows, 30 communication strengths or areas needing improvement and 24 significant gaps were identified, indicating self under-appraisals, 9 (38%) of which overlapped. The instrument was logistically feasible and well received. **CONCLUSIONS:** Our multirater communication skills instrument with gap analysis proved useful in identifying areas of strength and areas needing improvement, and in highlighting areas of self over- and under-appraisal that require focused feedback. The use of multirater assessment with gap analysis, in a simulated and "safe" environment, may assist in the delivery of feedback to trainees.

Cals, J. W., N. A. Scheppers, et al. (2007). "Evidence based management of acute bronchitis; sustained competence of enhanced communication skills acquisition in general practice." *Patient Education & Counseling* 68(3): 270-278.

**OBJECTIVE:** To determine if a communication skills training program for general practitioners involving context-rich learning experiences and 'peer review' of consultation transcripts results in communication skills acquisition and maintenance, while preserving time-efficiency in consultations. **METHODS:** A pre-test-post-test evaluation of training 20 general practitioners (GPs) in enhanced communication skills. Audio taped consultations with simulated patients in routine practice conducted before, within 2 weeks and again 6 months after communication skills training were analysed and consultation length measured. Transcripts were scored for specific skills to determine differences in short and longer-term competence of GPs for the communication skills. **RESULTS:** There was good evidence that GPs acquired key communication skills after training and that these were maintained over 6 months. Consultations remained within normal consultation length in primary care. **CONCLUSION:** Specific communication skills for acute bronchitis can be successfully acquired by GPs through context-rich communication training with peer review of transcripts with simulated patients, without making consultation length unfeasible. **PRACTICE IMPLICATIONS:** This approach to skill acquisition is useful for enhancing communication skills competence in general medical practice.

Cameron, N. and R. McMillan (2006). "Enhancing communication skills by peer review of consultation videos." *Education for Primary Care* 17(1): 40-48.

This paper describes an intervention designed to encourage general practitioners (GPs) to develop their communication skills and participate in analysis of video-recordings of their consultations. The need to develop communication skills teaching has long been recognised by providers of medical education. Video-recording of consultations offers one method of permitting reflection and analysis by individual doctors and provides material that can be submitted for external peer review. However, experience in this area is limited and the number of doctors involved in this type of educational activity has been small. Eleven doctors took part in a course that used facilitated small groups to review communication skills and to analyse video-recordings of participant's consultations. The participants were representative of the majority of the general practice population in that they were not part of the vocational training establishment and therefore had limited opportunities to access this type of educational activity. The results demonstrate that following this intervention all participants reported changes in their consulting behaviour and also submitted a video-recording for external peer review. This study indicates that small group video analysis and external peer review are effective methods of learning, developing and reviewing communication skills and describes a method that is both practical and feasible. This approach not only satisfies the requirements of the General Medical Council and appraisal, but participation in this type of communication skills training can also increase job satisfaction and enhance morale. copyright 2006 Radcliffe Publishing Limited.

Campbell, C. M. (2009). "Revalidation of Canadian physicians." *Canadian Medical Association Journal* 180(5): 539-540.

Campbell, J. L., S. H. Richards, et al. (2008). "Assessing the professional performance of UK doctors: an evaluation of the utility of the General Medical Council patient and colleague questionnaires." *Quality & Safety in Health Care* 17(3): 187-193.

**OBJECTIVE:** To investigate the utility of the GMC patient and colleague questionnaires in assessing the professional performance of a large sample of UK doctors. **DESIGN:** Cross-sectional questionnaire surveys. **SETTING:** Range of UK clinical practice settings. **PARTICIPANTS:** 541 doctors gave preliminary agreement to take part in the study. Responses were received from 13 754 patients attending one of 380 participant doctors, and from 4269 colleagues of 309 participant doctors. **MAIN OUTCOME MEASURES:** Questionnaire performance and standardised scores for each doctor derived from patient and colleague responses. **RESULTS:** Participant doctors were similar to non-participants in respect of age and gender. The patient and colleague questionnaires were acceptable to participants as evidenced by low levels of missing data. One patient questionnaire item seemed to cause confusion for respondents and requires rewording. Both patient and colleague responses were highly skewed towards favourable impressions of doctor performance, with high internal consistency. To achieve acceptable levels of reliability, a minimum of 8 colleague questionnaires and 22 patient questionnaires are required. G coefficients for both questionnaires were comparable with internationally recognised survey instruments of broadly similar intent. Patient and colleague assessments provided complementary perspectives of doctors' performance. Older doctors had lower patient-derived and colleague-derived scores than younger doctors. Doctors from a mental health trust and doctors providing care in a variety of non-NHS settings had lower patient scores compared with doctors providing care in acute or primary care trust settings. **CONCLUSIONS:** The GMC patient and colleague questionnaires offer a reliable basis for the assessment of professionalism among UK doctors. If used in the revalidation of doctors' registration, they would be capable of discriminating a range of professional performance among doctors, and potentially identifying a minority whose practice should be subjected to further scrutiny.

Cave, J., P. Washer, et al. (2007). "Explicitly linking teaching and assessment of communication skills." *Medical Teacher* 29(4): 317-322.

**Background:** Communication skills teaching is known to be effective, but students feel there are discrepancies between how communication skills are taught and how they are assessed. **Aims:** This study examined the effect of using standard assessment criteria during communication skills teaching on students' performance in an end-of-year summative OSCE. **Method:** Students attending their year 3 communication skills teaching were randomised to one of the following three conditions: the assessment criteria were available for reference on the medical school website; or students received the assessment criteria for use in the discussion and feedback; or each student's performance was graded by him- or her- self, his or her peers, the tutor and the actor using the standard assessment criteria. **Results:** There was no significant difference in the end-of-year OSCE performance of students who received the three different conditions. Actively using standard assessment criteria during teaching did not therefore improve OSCE performance. There were low but significant correlations between the tutors' assessment and the students' self-assessment and between the tutors' assessment and the peer group's assessment. **Conclusion:** The congruence between observers in the assessments of role-played consultations using the standard assessment criteria indicates that the criteria may be helpful for summarizing feedback to students.

Chaillet, N., E. Dube, et al. (2007). "Identifying barriers and facilitators towards implementing guidelines to reduce caesarean section rates in Quebec." *Bulletin of the World Health Organization* 85(10): 791-797.

**Objective:** To investigate obstetricians' perceptions of clinical practice guidelines targeting management of labour and vaginal birth after previous caesarean birth, and to identify the barriers to, facilitators of and obstetricians' solutions for implementing these guidelines in practice. **Methods:** This qualitative study was conducted in three hospitals in Montreal that represent around 10% of births in Quebec. Data was collected from 10 focus groups, followed by six semi-structured interviews. Two researchers jointly analysed the verbatim transcripts according to A manual for the use of focus groups. **Findings:** The identified barriers to and facilitators of the implementation of guidelines can be classified into four categories: 1) the hospital level, including management and hospital policies; 2) the departmental level, including local policies, leadership, organizational factors, economic incentive, and availability of equipment and staff; 3) the health professionals' motivations and attitudes, including medico-legal concerns, skill levels, acceptance of guidelines and strategies used to implement recommendations; and 4) patients' motivations. **Conclusion:** Identifying the barriers to and facilitators of the adoption of recommendations is an important way to guide the development of efficient

strategies. The findings of this study suggest that the adoption of guidelines may be improved if local health professionals' perceptions are considered to make recommendations more acceptable and useful. Our findings also support the assumption that obstetricians seek to implement best practices, but require evidence tools and support to assess their practices and enhance their performance. In addition, peer review activities championed by opinion leaders have been identified by obstetricians as the most suitable strategy to improve the use of the guidelines in their practices.

Cheadle, P. (2007). "Peer review." *Clinical Risk* 13(1): 30-33.

An explanation of the principles of peer review, charting the development of peer review to date and providing a guide to practitioners in preparing for the new ways of working with the Legal Services Commission focussing on clinical negligence.

Chenot, J. F., A. Simmenroth-Nayda, et al. (2007). "Can student tutors act as examiners in an objective structured clinical examination?" *Medical Education* 41(11): 1032-1038.

Context: The dissemination of objective structured clinical examinations (OSCEs) is hampered by requirements for high levels of staffing and a significantly higher workload compared with multiple-choice examinations. Senior medical students may be able to support faculty staff to assess their peers. The aim of this study is to assess the reliability of student tutors as OSCE examiners and their acceptance by their peers. Methods: Using a checklist and a global rating, teaching doctors (TDs) and student tutors (STs) simultaneously assessed students in basic clinical skills at 4 OSCE stations. The inter-rater agreement between TDs and STs was calculated by kappa values and paired t-tests. Students then completed a questionnaire to assess their acceptance of student peer examiners. Results: All 214 Year 3 students at the University of Gottingen Medical School were evaluated in spring 2005. Student tutors gave slightly better average grades than TDs (differences of 0.02-0.20 on a 5-point Likert scale). Inter-rater agreement at the stations ranged from 0.41 to 0.64 for checklist assessment and global ratings; overall inter-rater agreement on the final grade was 0.66. Most students felt that assessment by STs would result in the same grades as assessment by TDs (64%) and that it would be similarly objective (69%). Nearly all students (95%) felt confident that they could evaluate their peers themselves in an OSCE. Conclusions: On the basis of our results, STs can act as examiners in summative OSCEs to assess basic medical skills. The slightly better grades observed are of no practical concern. Students accepted assessment performed by STs. copyright 2007 Blackwell Publishing Ltd.

Collins, J. (2009). "Lifelong learning in the 21st century and beyond." *Radiographics* 29(2): 613-22.

Lifelong learning is now recognized by educators, governing bodies, accreditation organizations, certification boards, employers, third-party payers, and the general public as one of the most important competencies that people must possess. Promoting lifelong learning as continuous, collaborative, self-directed, active, broad in domain, everlasting, positive and fulfilling, and applicable to one's profession as well as all aspects of one's life has emerged as a major global educational challenge. Meeting this challenge will require changes in the way teachers teach and learners learn, as teachers take on a more facilitative role and learners take more responsibility for setting goals, identifying resources for learning, and reflecting on and evaluating their learning. For physicians, this will require less dependence on traditional educational venues, such as passive continuing medical education activities, and greater participation in self-assessment, peer assessment, evaluation of performance in practice, documentation of practice-based learning and improvement activities, and learning at the point of care. Radiologists in an academic setting are exposed to multiple opportunities for practicing lifelong learning, such as teaching others, participating in multidisciplinary conferences and journal clubs, and engaging in research. All radiologists can participate in self-audits and group audits of performance and become active participants in national radiology societies, where they can learn from each other. Participation in the American Board of Radiology's Maintenance of Certification program reflects a commitment to actively engage in lifelong learning and is one way of demonstrating to the general public a commitment to maintaining competence. (c) RSNA, 2009.

Conference of Medical Royal, C. and G. Their Faculties in the United Kingdom. Audit Working (1993). "Access to medical records for the purposes of medical audit." *BMJ* 306(6882): 896-7.

Cordeiro, L. and L. Newman (2001). "The 360 degrees feedback process: a tool for continuous board improvement." *Healthcare Management Forum* 14(1): 42-4.

To remain accountable to stakeholders, healthcare governing boards need meaningful and actionable feedback on their performance. In 1998, the Board of Governors of West Park Hospital (now West

Park Healthcare Centre) used the 360 degree feedback technique for the first time to gather feedback from a range of stakeholders. Board members found the technology accessible and the process valuable. They are using the individual results to improve individual performance, and have developed work plans to address areas for continued board development.

Craven, S. A. (2007). "Peer review." *British Journal of General Practice* 57(536): 245-246.

Crawford, S. M. (2009). "Opportunities for better cancer outcomes missed in UK primary care." *The Lancet Oncology* 10(1): 2-3.

Crossley, J., J. McDonnell, et al. (2008). "Can a district hospital assess its doctors for re-licensure?" *Medical Education* 42(4): 359-63.

CONTEXT: The Chief Medical Officer's recommendations on medical regulation in the UK suggest that National Health Service (NHS) trusts should assess their doctors and confirm whether they remain fit to practise medicine. OBJECTIVE: We set out to evaluate the utility of hospital trust-based assessment in a 'best-case scenario' within existing resources. METHODS: We carried out a generalisability analysis, and feasibility and validity evaluation, based on an assessment process for 137 career-grade doctors at Chesterfield Royal Hospital, Chesterfield, UK, using validated multi-source feedback (MSF) and patient rating (PR) instruments. RESULTS: Uptake and response rates were good for MSF (91% and 85%, respectively). However, only 6% of non-clinical doctors and anaesthetists, and 48% of clinical doctors, obtained sufficient PR ratings. Aggregate scores were acceptably reliable. Nine combined MSF ratings and 15 PR ratings produce standard errors of measurement of 0.19 on a 6-point scale and 0.15 on a 5-point scale, respectively. Overall aggregate scores did not identify any doctor as unsatisfactory, but 6 doctors were scored as unsatisfactory by 2 or more colleagues or patients. These performance concerns appear to merit further investigation. Patients rated female doctors better than male doctors (4.61 versus 4.46;  $P < 0.05$ ). Colleagues rated UK graduates better than non-UK graduates (5.31 versus 5.15;  $P < 0.05$ ). CONCLUSIONS: This study shows that the commissioning of professional services makes the implementation of an assessment process linked to appraisal feasible. However, trust-based assessment requires significant development: developmental appraisal needs protection; new instruments are needed for non-clinical specialties; PR requires specific administrative support, and guidance is required over concern thresholds and demographic effects. Disaggregated assessment data may help identify doctors with potential performance problems.

Crystal, B. (1994). "The 360 degree assessment. Healthcare executives receive letters from home about their leadership effectiveness." *Healthcare Executive* 9(6): 18-21.

Curfman, G. D., S. Morrissey, et al. (2008). "Peer review in the balance." *New England Journal of Medicine* 358(21): 2276-7.

Dalton, F. (2005). "Management. Using 360 degree feedback mechanisms." *Occupational Health & Safety* 74(7): 28.

You can harness the power of the 360 degree review process by following seven best practices.

Dalton, F. (2005). "Using 360 degree feedback mechanisms. You can harness their power and utility by following these seven best practices." *Occupational Health & Safety* 74(7): 28.

Davidson, M. L. (2007). "The 360degrees Evaluation." *Clinics in Podiatric Medicine and Surgery* 24(1): 65-94.

The 360degrees evaluation is an evaluation system representing the entire spectrum of resident assessment throughout residency training. Several questions still remain: Will the tremendous time and energy put into developing and instituting these assessment tools change the outcomes? Will the residents and ultimately the public benefit in the long term? No one yet has the answers. However, the process forces faculty to look much more closely at how residents are educated, identifying not only residents' weaknesses but also shortcomings within both specialties and programs. The process may be arduous but the concept is worthy, and the information gained along the way is well worth the effort. copyright 2006 Elsevier Inc. All rights reserved.

Davies, H., J. Archer, et al. (2008). "Specialty-specific multi-source feedback: Assuring validity, informing training." *Medical Education* 42(10): 1014-1020.

Context: The white paper 'Trust, Assurance and Safety: the Regulation of Health Professionals in the 21st Century' proposes a single, generic multi-source feedback (MSF) instrument in the UK. Multi-source feedback was proposed as part of the assessment programme for Year 1 specialty training in histopathology. Methods: An existing instrument was modified following blueprinting against the histopathology curriculum to establish content validity. Trainees were also assessed using an objective structured practical examination (OSPE). Factor analysis and correlation between trainees' OSPE performance and the MSF were used to explore validity. All 92 trainees participated and the assessor response rate was 93%. Reliability was acceptable with eight assessors (95% confidence interval 0.38). Factor analysis revealed two factors: 'generic' and 'histopathology'. Pearson correlation of MSF scores with OSPE performances was 0.48 ( $P = 0.001$ ) and the histopathology factor correlated more highly (histopathology  $r = 0.54$ , generic  $r = 0.42$ ;  $t = -2.76$ , d.f. = 89,  $P < 0.01$ ). Trainees scored least highly in relation to ability to use histopathology to solve clinical problems (mean = 4.39) and provision of good reports (mean = 4.39). Three of six doctors whose means were  $< 4.0$  received free text comments about report writing. There were 83 forms with aggregate scores of  $< 4$ . Of these, 19.2% included comments about report writing. Results: Specialty-specific MSF is feasible and achieves satisfactory reliability. The higher correlation of the 'histopathology' factor with the OSPE supports validity. This paper highlights the importance of validating an MSF instrument within the specialty-specific context as, in addition to assuring content validity, the PATH-SPRAT (Histopathology-Sheffield Peer Review Assessment Tool) also demonstrates the potential to inform training as part of a quality improvement model.

Davies, H., J. Archer, et al. (2009). "Initial evaluation of the first year of the Foundation Assessment Programme." *Medical Education* 43(1): 74-81.

OBJECTIVES: This study represents an initial evaluation of the first year (F1) of the Foundation Assessment Programme (FAP), in line with Postgraduate Medical Education and Training Board (PMETB) assessment principles. METHODS: Descriptive analyses were undertaken for total number of encounters, assessors and trainees, mean number of assessments per trainee, mean number of assessments per assessor, time taken for the assessments, mean score and standard deviation for each method. Reliability was estimated using generalisability coefficients. Pearson correlations were used to explore relationships between instruments. The study sample included 3640 F1 trainees from 10 English deaneries. RESULTS: A total of 2929 trainees submitted at least one of all four methods. A mean of 16.6 case-focused assessments were submitted per F1 trainee. Based on a return per trainee of six of each of the case-focused assessments, and eight assessors for multi-source feedback, 95% confidence intervals (CIs) ranged between 0.4 and 0.48. The estimated time required for this is 9 hours per trainee per year. Scores increased over time for all instruments and correlations between methods were in keeping with their intended focus of assessment, providing evidence of validity. CONCLUSIONS: The FAP is feasible and achieves acceptable reliability. There is some evidence to support its validity. Collated assessment data should form part of the evidence considered for selection and career progression decisions although work is needed to further develop the FAP. It is in any case of critical importance for the profession's accountability to the public.

Dawson, A. (1985). "The accountability of doctor to doctor." *Lancet* 2(8450): 323-4.

DeAngelis, C. D. and J. P. Thornton (2008). "Preserving confidentiality in the peer review process." *JAMA* 299(16): 1956.

Deming, W. (2008). "Clinical peer review: A plea for fairness." *Catheterization and Cardiovascular Interventions* 71(7): 1000-1001.

DeSimone, R. (1998). "Establishing the link: relating a 360 degree management assessment and development process to the bottom line." *Health Care Supervisor* 17(1): 31-6.

The challenges facing hospital management today demand that both medical and nonmedical personnel be as productive as possible, with the ultimate goal of establishing a link between employee growth and return on investment, while maintaining the highest standards of patient care. Using two case studies--River Parishes Hospital and Manatee Hospital--this article discusses how 360-degree assessment programs can contribute to individual and team development as well as the bottom line. The article will show how ACUMEN Skills, a 360 degree assessment tool, implemented

with appropriate facilitation, helped to increase productivity and efficiency, improve relationships among staff at all levels, reduces expenses, and strengthen the bottom line.

Douglas, R. M. and A. Blood (1978). "Evaluation and peer review of the role of specialist and general medical units in a teaching hospital." *Australian & New Zealand Journal of Medicine* 8(3): 337-43.

The trend toward increasing specialization in the medical wards of teaching hospitals has important implications for future health care delivery and student teaching. A study of the likely implications of a change to a specialty oriented medical admission system at the Royal Adelaide Hospital has shown that already the general physicians are seeing little renal and endocrine disease and, that work satisfaction is greater on the specialty units than on the general medical units. Patient satisfaction marginally favoured the specialty units. A peer review audit method was introduced to try and discern whether the quality of care was significantly different on the two types of units. Most physicians cooperated in at least one or two audits which involved them in judgements about adequacy of casenotes, management decisions, follow-up arrangements, length of stay and investigations in randomly assigned specialty pairs of patients (cared for by specialist or general units). No clear-cut trend was discerned favouring specialist or generalist care amongst 17 randomly selected pairs, but this part of the study was never completed, perhaps partly because of lack of enthusiasm for it by busy clinicians. The importance of maintaining a model of generalism for medical students on the teaching hospital campus is discussed in the light of the decline of the role of general physicians.

Edwards, J. (2004). "Practice and peer review for freelance GPs: A pilot study." *Education for Primary Care* 15(SUPPL.): 468-469.

In this pilot study, a feedback form was designed for freelance GPs to use to provide peer review of aspects of their performance for their personal portfolio. The form was found to be useful and acceptable to practices and GPs.

Ellis, P. M., G. W. Mellsop, et al. (1987). "Peer review as an aid to improving the completeness of psychiatric case notes." *Medical Education* 21(6): 493-7.

Clinical auditing and the setting of goals for continuing education are often based on case note reviews. Deficiencies in the comprehensiveness of the recorded case history place some restrictions on the usefulness of such reviews as guides to continuing education programmes. In this study, attempts were made to improve the quality of the data in psychiatric case notes by peer discussion, and by altering the case note recording guidelines. After each educational intervention a further detailed audit of the subsequent case histories was performed. Significant differences emerged in a few subsections of the case notes, but overall there was little change. Possible explanations for these findings are discussed.

Epstein, R. M., E. F. Dannefer, et al. (2004). "Comprehensive assessment of professional competence: the Rochester experiment." *Teaching & Learning in Medicine* 16(2): 186-96.

**BACKGROUND:** A required 2-week comprehensive assessment (CA) for 2nd-year medical students that integrates basic science, clinical skills, information management, and professionalism was implemented. **DESCRIPTION:** The CA links standardized patients (SPs) with computer-based exercises, a teamwork exercise, and peer assessments; and culminates in student-generated learning plans. **EVALUATION:** Scores assigned by SPs showed acceptable interrater reliability. Factor analyses defined meaningful subscales of the peer assessment and communication rating scales. Ratings of communication skills were correlated with information gathering, patient counseling, and peer assessments; these, in turn, were strongly correlated with the written exercises. Students found the CA fair, with some variability in opinion of the peer and written exercises. Useful learning plans and positive curricular changes were undertaken in response to the CA results. **Conclusion:** A CA that integrates multiple domains of professional competence is feasible, useful to students, and fosters reflection and change. Preliminary data suggest that this format is reliable and valid.

Essex, C., I. M. Rubin, et al. (2004). "Review of instruments for peer assessment of physicians." *British Medical Journal* 329(7458): 171.

Evans, A. W., R. M. A. Leeson, et al. (2007). "Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery." *Medical Education* 41(9): 866-872.

**Context:** It is sometimes claimed that self-assessment is inaccurate and that clinicians over-rate their performance. There is a need to find out why this should be. Is poor self-assessment caused by some clinicians' inability to accurately judge performance? Or does over-scoring result from a desire to

convey a more favourable impression? Peer assessment is widely advocated and is said to be of benefit to both assessor and assessee. Methods: In this study, we wanted to see if postgraduates were able to peer-assess and if this form of assessment was more reliable than self-assessment when compared with assessment by a trainer. We used checklist and global rating scales to evaluate surgical skills in removing a mandibular third molar tooth. Results: There was no statistically significant difference between peer-assessed and trainer-assessed scores. We found that, on average, peer assessment (especially global rating scales) reflected trainer scores more accurately than self-assessment of surgical skills. Self-assessment scores were significantly higher on average than those given in peer assessment. Discussion: Although peers and trainee surgeons came from the same group, the surgeons were more likely to over-score when measuring their own performances. The greatest variability (and over-scoring) between assessor and trainee surgeon appeared to occur in those with lower mean scores. Formative peer assessment may be a useful and less stressful mechanism for encouraging reflection.

Evans, R. G. B. (1992). "Accreditation process in the United Kingdom." *International Journal of Radiation Oncology Biology Physics* 24(5): 863-865.

Ewen, A. J. (1998). "360 degree feedback process drives culture change in health care." *Surgical Services Management* 4(4): 28.

Ferguson, K. J. and C. D. Kreiter (2007). "Assessing the relationship between peer and facilitator evaluations in case-based learning." *Medical Education* 41(9): 906-908.

Objectives: Attempts to validate peer evaluation and to incorporate it into the curriculum have met with mixed results. The purpose of this study was to assess the use of peer evaluations in a Year 1 case-based learning course. Methods: As part of the formal grading process for the course, all faculty facilitators (n = 69 over 3 years) completed a 12-item evaluation form for each student at the conclusion of each case. As part of a course assignment, students (n = 415 over 3 years) completed brief evaluations of their peers based on 2 criteria: the overall quality of written reports, and participation in group discussion. In addition, students provided anonymous feedback in the written end-of-course evaluation about the peer evaluation process, and faculty were asked to comment during the wrap-up luncheon for small-group facilitators. Results: Response rates for the 3 Year 1 medical student classes ranged from 95% to 99%. The average number of peer evaluations completed for each student was 4.6. The G coefficients for the rater-nested-within-person generalisability study were 0.52 for written reports and 0.60 for group participation; both were based on an average of 4-5 ratings. Correlation coefficients between peer and faculty evaluations in each of the 3 consecutive years of the course ranged from 0.46 to 0.63; all were statistically significant at  $P < 0.001$ . A correction for attenuation suggests that the true score correlation between faculty and peer measures is near 1.0. Discussion: This study provides strong evidence that facilitator and peer ratings measure similar constructs and shows that, even among Year 1 medical students, peer evaluation can be conducted in a valid manner. copyright 2007 Blackwell Publishing Ltd.

Flew, R. and S. Street (2008). "Assessing the appraisers: Quality assurance for the appraisal of general practitioners." *Education for Primary Care* 19(3): 265-273.

Organisations responsible for ensuring that appraisals are undertaken do not have a recognised method to assess the competence of newly trained or established appraisers. This pilot project has established a system for the quality assurance of the GP appraisal interview in which experienced appraisers are selected and trained to become assessors of newly trained appraisers. The appraiser assessors have developed the competencies and the expected standard of appraisal skills. With this system primary care organisations (PCOs) will have a dependable method to ensure the quality of their appraisal process.

Forster, A. J., K. O'Rourke, et al. (2007). "Combining ratings from multiple physician reviewers helped to overcome the uncertainty associated with adverse event classification." *Journal of Clinical Epidemiology* 60(9): 892-901.

OBJECTIVES: Adverse events (AEs) are poor patient outcomes, resulting from medical care. We performed this study to quantify the misclassification rate obtained using current AE detection methods and to evaluate the effect of combining physician AE ratings. STUDY DESIGN AND SETTING: Three physicians independently rated poor patient outcomes. We used latent class analysis to obtain estimates for AE prevalence and reviewer accuracy. These estimates were used as a base case for four simulations of 10,000 cases rated independently by five reviewers. We assessed



the effect of AE prevalence, reviewer accuracy, and the number of agreeing reviewers on the probability that cases were correctly classified as an AE. RESULTS: Reviewer sensitivity and specificity for AE classification were 0.86 and 0.94, respectively. When prevalence was 3%, the positive predictive value that an AE occurred when a single reviewer classified the case as such was 31%, whereas when 2/3 reviewers did so it was 51%. The positive predictive values of ratings for AE occurrence increased with AE prevalence, reviewer accuracy, and the number of reviewers. CONCLUSION: Current methods of AE detection overestimate the risk of AE. Uncertainty regarding the presence of an AE can be overcome by increasing the number of reviews.

Franco, J. M., P. De Chazournes, et al. (2007). "Impact of peer visits." *Revue du Praticien* 57(11): 1211-1217.

Objectives: To measure the impact of an intervention on the general practitioners (GP) of Reunion in order to improve the management of patients with type 2 diabetes, in conforming with the recommendations of the Anaes. Method: Randomised intervention study on a random sample of 120 practitioners out of a total of 630 GP in Reunion, 60 in the intervention group (IG) and 60 in the control group (CG). Each doctor of the IG received 2 visits by a << visiting GP >> which have had specific training. The period of observation included the 12 months before, and the 6 months after the date of the intervention. Data were collected retrospectively, at the end of 18 months of observation; in medical records of 25 diabetics seen consecutively in consultation, the GP collected the dates of performance of six procedures under surveillance: HbA1c, examination of the feet, funduscopy, ECG, estimation of the creatinine clearance, level of micro-albuminuria. Outcome measures were delays in performance of the procedures conforming to the recommendations. Results: 42 GP out of 60 in the IG, and 40 out of 60 in the CG participated to the study. Patients included, 792 in the IG and 789 in the CG, were comparable for age, sex-ratio and profession. The distribution of delays in performance before the intervention was comparable in the two groups. The comparison between the groups after the intervention showed a significantly important improvement in the IG for 4 to 6 of the procedures: examination of the feet, funduscopy, creatinine clearance and micro-albuminuria. Conclusion: In the short-term, a << outreach visit >> or << academic detailing >> improves the delay in performance of most of the surveillance procedures in type 2 diabetes.

Furnham, A. and P. Stringfield (1998). "Congruence in job-performance ratings: a study of 360 degree feedback examining self, manager, peers, and consultant ratings." *Human Relations* 51(4): 517-530. Seven task-oriented teams worked together for between 4 and 6 months on a project. At the end of the task, members were each subject to four task-performance ratings: from self, superior, team-peers, and a consultant who was part of the team. There were fewer than chance differences between the different teams on the congruence measures so the data was combined. While the congruence between self and manager, self and peer, and self and consultant ratings were very low, the manager-peer, manager-consultant, and peer-consultant congruence was overall high. Observable behaviors like forward planning and communication showed overall highest congruence while less observable cognitive variables showed much lower congruence. These results are similar to previous studies in the area. Implications of the use of these ratings in management development are considered.

Garbett, R., S. Hardy, et al. (2007). "Developing a qualitative approach to 360-degree feedback to aid understanding and development of clinical expertise." *Journal of Nursing Management* 15(3): 342-7. AIM: This paper presents one aspect of a 5-year multicentre action research study to develop an accreditation process for clinical nursing expertise. Part of the process consisted of the exploration, critique and refinement of qualitative 360-degree feedback as a tool for peer review. BACKGROUND: Three hundred and sixty-degree feedback is widely used as a personal and professional development strategy. This part of the overall study challenged assumptions about the necessity for anonymity and structured questionnaires to collect data. The study involved 32 experienced clinical nurses drawn from a range of clinical settings supported by 'critical companions' (colleagues from clinical practice, education, management and research, recruited to provide supervision and support). METHOD(S): Study participants, facilitated by the project team (the authors), engaged in critiquing and refining 360-degree feedback as a process to help them examine and develop their practice. CONCLUSION(S): On the basis of our findings this approach to gathering 360-degree feedback facilitates the collection of evidence that aids professional development. There are indications that it may also contribute to improved working relationships.

Garman, A. N., J. L. Tyler, et al. (2004). "Development and validation of a 360-degree-feedback instrument for healthcare administrators... including commentary by Lerner W." *Journal of Healthcare Management* 49(5): 307-322.

Goebel, L. J. (1997). "A peer review feedback method of promoting compliance with preventive care guidelines in a resident ambulatory care clinic." *Joint Commission Journal on Quality Improvement* 23(4): 196-202.

**BACKGROUND:** Simple distribution of clinical practice guidelines to physicians does not change practice behavior. A low-cost, continuous peer review feedback method was used to promote resident physicians' compliance with nine preventive care guidelines at the ambulatory care clinic at the Marshall University School of Medicine (Huntington, West Virginia). **METHODS:** Preventive care guidelines were distributed and a peer review feedback program was instituted in the resident physician primary care practice. The frequency of resident physician use of nine preventive care services was assessed and compared during three periods: preguideline (September 1, 1993, to March 1, 1994; 148 patients), guideline (September 1, 1994, to March 1, 1995; 148 patients), and one-year follow-up (September 1, 1995, to March 1, 1996; 150 patients). The patients in the three periods were similar in age, gender, and risk for influenza and pneumococcal infection. **RESULTS:** During the guideline period, resident physicians offered patients four preventive care services-tetanus toxoid immunization, clinical breast examination, Papanicolaou smear testing, and hemocult testing significantly more often than during the preguideline period. All services were offered significantly more often during the one-year follow-up period compared with the preguideline period and as often as in the guideline period. **CONCLUSION:** A low-cost, continuous peer review feedback program significantly and durably improves resident physician compliance with clinical practice guidelines on preventive care services. However, the effectiveness of the peer review feedback method may not generalize to private practice or other settings. Research on other methods to promote compliance with clinical practice guidelines and to influence physician behavior in general should continue.

Goulet, F., R. Gagnon, et al. (2007). "Influence of remedial professional development programs for poorly performing physicians." *Journal of Continuing Education in the Health Professions* 27(1): 42-48.

**Introduction:** The College des Medecins du Quebec (CMQ) offers an individualized remedial professional development program to help physicians overcome selected clinical shortcomings. To measure the influence of the remedial professional development program, physicians who completed the program between 1993 and 2004 and who were assessed by peer review during a 2-year period preceding or following the remedial activities were tracked. **Methods:** For each physician, 30 to 50 patient records were selected randomly for review. Ratings were assigned for the quality of record keeping and for 3 elements pertaining to the quality of care: the clinical investigation plan, diagnostic accuracy, and patient treatment and follow-up. The impact of the program was measured by comparing the proportion of physicians with satisfactory ratings assigned by peer review before and after the remedial professional development program. **Results:** Statistically significant improvements ( $p < .05$ ) were observed for a proportion of physicians ( $n = 51$ ) with satisfactory ratings with regard to record keeping (20% before and 54% after remediation), the clinical investigation plan (13% before and 59% after remediation), diagnostic accuracy (32% before and 61% after remediation), and patient treatment and follow-up (31% before and 67% after remediation). **Discussion:** Participation in a CMQ remedial professional development program can result in improved clinical performance, as assessed through peer review.

Goulet, F., A. Jacques, et al. (2007). "Assessment of family physicians' performance using patient charts: interrater reliability and concordance with chart-stimulated recall interview." *Evaluation & the Health Professions* 30(4): 376-92.

Peer-assessment processes with chart review have been used for many years to assess the clinical performance of physicians. The Quebec medical licensing authority has been required by provincial law to assess the practicing Quebec physicians on a nonvoluntary basis. During the period from January 2001 to November 2004, 25 family physicians in active practice were randomly selected from a pool of about 300. For each physician, 25 to 40 patients' medical charts were randomly selected to evaluate the interrater reliability of peer-review assessment of medical charts and to compare ratings based on chart review with a chart-stimulated recall interview to those based on chart review alone. The concordance between chart review alone and that of chart review with chart-stimulated recall interview was 75% for chart keeping, 69% for clinical investigation, 81% for diagnostic accuracy, and 74% for treatment plan. Ratings based on chart review alone achieve moderate levels of reliability

(Kappa = 0.44 to 0.56). It appears that some important information about quality of care is missed when only chart review is used.

Govindan, V. K. (2008). "Enhancing communication skills using an OSCE and peer review." *Medical Education* 42(5): 535-536.

Graber, M. A., A. Hartz, et al. (2005). "An alternative method of determining standard of care in alleged cases of malpractice." *Journal of the American Board of Family Practice* 18(6): 453-8.

**BACKGROUND:** Physician experts hired and prepared by litigants provide most information on standard of care for medical malpractice cases. Because this information may not be objective or accurate, we examined the feasibility and potential value of surveying peer physicians to assess standard of care. **METHODS:** The survey method was evaluated for a medical malpractice case involving a patient hospitalized with abdominal pain. An abstract of the medical record was created that included the patient characteristics and physician decisions most likely to influence patient outcome. The abstract and questionnaire were sent to 16 academic family physicians and to 20 randomly chosen primary care physicians in Iowa who practiced in communities of similar size to the defendant's community. **RESULTS:** All 16 academic and 18 (90%) community physicians completed the survey. All respondents judged the patient as presenting with an acute abdomen, and 89% of the community physicians and 100% of the academic physicians judged the care as below standard. More than half the physicians surveyed listed the autopsy diagnosis (perforated ulcer) in their differential. **CONCLUSION:** Surveys of randomly selected physicians are feasible to perform for medical malpractice cases. A pro-physician bias has little if any influence on the results.

Greenwood, J. and R. Williams (2008). "Continuing professional development for Australian rural psychiatrists by videoconference." *Australasian Psychiatry* 16(4): 273-276.

**Objective:** The Rural Special Interest Group (RSIG) of the Royal Australian and New Zealand College of Psychiatrists provided a program of educational meetings for psychiatrists and trainees in rural locations to decrease professional isolation and facilitate opportunities to obtain credit for continuing professional development (CPD) purposes in ongoing professional registration. A pilot peer review group by videoconference was also implemented in Western Australia, to provide peer review for psychiatrists working in isolation. **Method:** Six national interactive videoconferenced education seminars were provided. The membership was polled prior to the program for topic preferences and requested speakers, who were subsequently approached to present the seminars. A West Australian peer review group was formed, involving psychiatrists from five different rural locations and one psychiatrist from the metropolitan area. **Results:** National seminars were attended by 106 individuals (67% consultant psychiatrists, 17% trainees and 16% allied health professionals). Evaluation demonstrates favourable views of the content and quality of the presentations and impact on practice patterns. **Conclusions:** Attendance at seminars via videoconferencing provided rural psychiatrists with opportunities to obtain CPD credit without the need to travel, providing a cost-effective alternative for accessing education. Involvement enabled interaction among peers throughout Australia, alleviating professional isolation and generating networking opportunities. The West Australian peer review group demonstrated sustainability with a commitment to continuation and positive impact on practice.

Grol, R., J. van Eijk, et al. (1985). "Audit: a project on peer review in general practice." *Family Practice* 2(4): 219-24.

A project of group-based peer review by general practitioners is described. Attention is given to the motivation of general practitioners to participate in audits, to the threat posed by peer review and to the use of criteria. The recruitment and motivation were successful. A total of 253 general practitioners took part, divided into 23 groups. This represented approximately 30% of all general practitioners in the area and 72% of all general practitioners specifically approached by us. The threat posed by audit disappears rapidly for most physicians once they have started auditing. The use of predefined criteria developed by 'experts', representing an optimal way of work for the general practitioner, appeared to be quite feasible in this project. From the results several conclusions are drawn as regards the planning and implementation of peer review under general practitioners.

Gukas, I. D., S. Miles, et al. (2008). "Medical students' perceptions of peer feedback on an anatomy student-selected study module." *Medical Teacher* 30(8): 812-814.

**Objective:** The study aimed to determine student views of peer feedback on their student-selected study (SSS) module. **Methods:** A questionnaire was developed to study perceptions of three groups of medical students (N = 42) towards feedback received from peers about their anatomy SSS

presentation. Results: Most students felt comfortable receiving and giving feedback. They also felt that received feedback was fair, adequate and helpful, and that receiving feedback made them reflect. Slightly more students reported inadequate feedback from their peers about the presentations' content, compared to other aspects, due to their peers' relative lack of knowledge about their 'specialized' subject. Students would be reluctant to give feedback if anonymity was removed. Conclusion: The attitudes of medical students towards peer feedback were largely positive. We advocate further studies to evaluate quality of feedback, and the role of anonymity in peer feedback, and its effect on group dynamics and cohesion.

Gurevitz, H. (1976). "All marriages are not in heaven made: the joining of quality assurance and utilization review." *Journal - National Association of Private Psychiatric Hospitals* 8(3): 17-20.

Hagan, C. M., R. Konopaske, et al. (2006). "Predicting assessment center performance with 360-degree, top-down, and customer-based competency assessments." *Human Resource Management* 45(3): 357-390.

In the first criterion-related validity study of a complete 360-degree competency assessment process (i.e., where customer data are included), aggregated 360-degree assessment of 428 retail associate store managers on six competencies showed strong validity (.50) in the prediction of assessment center performance. In addition, 360-degree assessments on each of the six competencies were significantly correlated with the criteria. The aggregated 360-degree assessments also demonstrated incremental validity over managerial ratings alone in the prediction of assessment center criteria. Customer (mystery shopper) assessments were also significantly correlated with the assessment center criteria and exhibited incremental validity beyond supervisory assessments. © 2006 Wiley Periodicals, Inc.

Hagopian, A., P. West, et al. (1994). "Improving a rural hospital's peer review process." *Journal for Healthcare Quality: Promoting Excellence in Healthcare* 16(6): 17-19.

Haurani, M. J., I. Rubinfeld, et al. (2007). "Are the communication and professionalism competencies the new critical values in a resident's global evaluation process?" *Journal of Surgical Education* 64(6): 351-356.

Background: The ACGME requires the assessment of resident competency in 6 domains. Global evaluations covering all 6 competencies are routinely used. Evaluators may be overly influenced by resident affability and availability, thereby resulting in a halo effect. We hypothesized that the Interpersonal Skills and Communications (ICS) and Professionalism (PR) competencies would unduly influence other competency scores. Methods: General surgery resident evaluations are performed by staff and peers on a rotational basis using competency-based questions. Each question is scored using a 5-point Likert scale. Mean individual composite scores for each competency were calculated and then correlated with other mean composite competency scores. Data from patient evaluations were similarly analyzed. A final correlation of competency scores to ABSITE scores, as an objective, standardized measure of a specific competency, Medical knowledge (MK) was also performed. Results: Results were available for 37 residents (PGY 1-5). There was a significant association between ICS scores and higher scores in MK ( $r = 0.52$ ,  $p = 0.004$ ), PR ( $r = 0.826$ ,  $p < 0.0001$ ) and patient care (PC) ( $r = 0.619$ ,  $p < 0.0001$ ). No correlation, however, was found between patient evaluations of residents and their faculty/peer-based ICS scores. We found no association between ICS scores and improved patient evaluations. Lastly, we found no association between ICS or MK scores and ABSITE scores. Conclusions: It was difficult to ascertain whether residents with better ICS scores had higher PR, PC, and MK scores because of the halo effect, improper completion of evaluations, or whether those residents were truly performing better clinically. External measures of resident performance did not correlate with faculty/peer evaluations of ICS and PR. Residency programs should consider adopting a more standardized way to objectively evaluate residents.

Hedge, R. and A. V. M. Hughson (2008). "Consultant appraisal of consultant appraisal." *Psychiatric Bulletin* 32(6): 214-217.

Aims and Method: A postal survey was conducted to gauge opinion of consultant psychiatrists in the west of Scotland with regard to consultant appraisal. Results: We received 158 responses to our survey (a response rate of 77%). The results showed mixed feelings about appraisal - for example, its purpose was felt to be about the development of consultants (72%) rather than the protection of patients (53%). Practical problems were highlighted, like the provision of accurate data and the amount of time spent on the preparation for appraisal. Participation in 360-degree appraisal was

scant, and no consensus on the use of outcome measures was reached. Implications: More work needs to be done on the appraisal process for it to gain the full confidence of the profession.

Hershey, N. (1990). "Assessing peer review in the quest for improved medical services and the implications for education in quality assessment: Part IV." *Quality Assurance & Utilization Review* 5(4): 130-7.

Three court decisions reviewing medical peer review conducted in hospitals were discussed at length in part III of this article. In their opinions the courts gave at least tacit approval to the procedures followed in the hospitals, and they accepted that an evidentiary basis for adverse action against the physicians was present. But not all medical peer review in hospitals resulting in adverse actions is found satisfactory when challenged in court, and the most prominent litigation in the decade of the 1980s concerning medical peer review, *Patrick v. Burget* (1), is testimony to the potential for its perversion. Part II adverted to the potential for bias or lack of objectivity in assessing physician performance. Part I mentioned the problem of bias in the context of peer review of articles for publication and of research grant proposals. The objectives of Part IV are: (1) to examine the concern about bias in medical peer review and to indicate how it may be lessened, if not eliminated; (2) to address further the difficulty created by the relative lack of valid criteria to employ in medical peer review; (3) to review the extent of protection from liability afforded to participants in medical peer review; and (4) to describe the changes that should be anticipated in review of medical services in the future. Before addressing these subjects it is essential to remind the reader that medical peer review is not conducted primarily for disciplinary purposes; rather, its purpose is to evaluate the quality of care.(ABSTRACT TRUNCATED AT 250 WORDS)

Hershey, N. (1990). "Assessing peer review in the quest for improved medical services: Part III." *Quality Assurance & Utilization Review* 5(2): 63-8.

Part I of this article explored flaws in the operation of peer review in the selection of articles for journal publication and decision-making in the award of research grants, and Part II described medical peer review, its practitioners, and the distinctions between medical peer review and other activities and processes aimed at improving the quality of medical performance in hospitals. This part of the article directs attention to how the courts respond when a physician, aggrieved by an adverse determination with regard to appointment, reappointment, or clinical privileges (credentialing) by the hospital based on medical peer review, seeks redress in the courts.

Hershey, N. and L. C. Bontempo (1990). "Assessing peer review in the quest for improved medical services: Part II." *Quality Assurance & Utilization Review* 5(1): 7-11.

The first part of this article explored the flaws in the operation of peer review in two contexts: selection of articles for journal publication and decision-making in the award of research grants. There it was suggested that, to the extent these flaws had adverse effects on the information available to medical practitioners and those who conduct peer review of medical practice, the quest for improving the quality of health services is hampered. In this part of the article, medical peer review is defined, its practitioners noted, and the distinction between medical peer review and certain other activities and processes directed toward improving the quality of medical performance in hospitals illustrated.

Herzog, A. (2008). "The HAVEN Program colleagues helping colleagues." *Connecticut Medicine* 72(4): 239-40.

Higgins, R. S. D., J. Bridges, et al. (2004). "Implementing the ACGME general competencies in a cardiothoracic surgery residency program using 360-degree feedback." *Annals of Thoracic Surgery* 77(1): 12-7.

BACKGROUND: Methods to assess the six competency categories outlined by the Accreditation Council on Graduate Medical Education are essential to allow residency programs to develop reproducible evaluations of their educational curriculum. Current tools to evaluate competencies are insufficient to perform these tasks, particularly in subspecialty disciplines. The key objective of this initiative was to develop and implement an evaluative tool that would provide data to residents and program leadership regarding their performance and to provide the training program with a reliable way to assess this component of the program. METHODS: Utilizing a highly customized survey tool with a group of cardiothoracic residents, we implemented a 360-degree performance assessment process based on the six Accreditation Council on Graduate Medical Education competency areas. The full spectrum of associations in a resident's sphere of interaction were surveyed (ie, supervisors, peers, direct reports, nurses, and administrative personnel). Each resident received a comprehensive

report that included detailed documentation of the self-evaluation and the average rating of others by category. Each resident also received a transcript of the responses to the open-ended questions and summary of the data highlighting areas of excellence, areas for improvement, and suggested goals and recommendations. The program director received copies of all of these as well as a chart documenting the average scores on each item for the whole cohort. RESULTS: Each resident met with the 360-degree feedback specialist and the program director to develop and commit to an action plan based on the feedback. The feedback process was repeated approximately 8 months later. CONCLUSIONS: The 360-degree feedback results provided valuable information for the residents. It also provided our program with a reproducible, quantifiable tool to assess these competencies. Combined with other instruments, the 360-degree feedback was found to be a particularly valuable instrument.

Hoop, J. G., A. C. Smyth, et al. (2008). "Ethical Issues in Psychiatric Research on Children and Adolescents." *Child and Adolescent Psychiatric Clinics of North America* 17(1): 127-148.

Please approve the following synopsis as it will appear in the table of contents: Psychiatric research on children and adolescents is ethically justified by the need to reduce the burden that mental illnesses place on young people, their families, and society. Such research must be conducted with careful attention to the ethical principles of beneficence, justice, and respect for persons. Child and adolescent psychiatrists who collaborate on research trials or advise patients and families about research participation should consider nine domains when evaluating the ethical acceptability of particular protocols. These domains include scientific merit and design; expertise, commitment, and integrity; risks and benefits; confidentiality; participant selection and recruitment; informed consent and decisional capacity; incentives; institution and peer/professional review; and data presentation. Special ethical issues in child and adolescent psychiatry research concern the use of randomized, controlled treatment trials; the informed consent process for research involving adolescents; the therapeutic misconception; and conflicts of interest in physician referrals. copyright 2008 Elsevier Inc. All rights reserved.

Hrisos, S., J. C. Illing, et al. (2008). "Portfolio learning for foundation doctors: early feedback on its use in the clinical workplace." *Medical Education* 42(2): 214-23.

CONTEXT: A learning portfolio was developed to support the development of trainee doctors piloting Foundation Programme prototypes across the Northern Deanery in 2004 and 2005. Trainee doctors and their educational supervisors were surveyed about their experiences of using the portfolio in the clinical workplace. METHODS: The evaluation consisted of semi-structured interviews with trainee doctors and supervisors, followed by postal questionnaire surveys in 2004 and 2005. Quantitative and qualitative data were triangulated to identify core findings. RESULTS: Questionnaires were returned from 182/305 (60%) trainee doctors and 104 out of 179 (58%) educational supervisors. The portfolio was felt to be a 'good idea' by 55% supervisors and 48% trainees. Trainees' perceptions of the educational value of the portfolio remained consistently low over 2 surveys and they described a sense of 'burden', whereby they identified problems in workload and usability and in gaining feedback on performance. However, positive trainee attitudes towards the portfolio were significantly correlated with greater perceived educational benefits ( $r = 0.855$ ,  $P < 0.001$ ). DISCUSSION: Learning portfolios are now an integral part of Foundation Programme training but this evaluation suggests that many trainee doctors and educational supervisors are yet to be convinced of their educational value. Gaining multi-source feedback, a substantial component of trainee doctors' portfolios, impacts on the wider clinical team and presents a significant challenge to trainees. Educational supervisors continued to rely on feedback from clinical colleagues, rather than portfolio evidence, to monitor trainee doctors' development. Such factors may serve to disengage trainees with the portfolio process by overshadowing any perceived educational gains.

Hughes, C., S. Toohy, et al. (2008). "eMed teamwork: A self-moderating system to gather peer feedback for developing and assessing teamwork skills." *Medical Teacher* 30(1): 5-9.

Students in the six-year undergraduate medicine program at UNSW must submit a portfolio which demonstrates inter alia their development in teamwork skills. Much of the feedback they need to develop these skills, as well as the evidence they require to document their achievements, can only come from their peers. The eMed Teamwork system, developed for this purpose, is a computer-based system which gathers feedback from peers in project groups. The feedback submitted to the system is available to the recipient for formative purposes, and becomes part of both the author's and the recipient's portfolios for later summative assessment. This dual use ensures that the feedback is thoughtful and constructive and the system operates without significant moderation by teachers.

Hulse, N. C., G. Del Fioli, et al. (2008). "Towards an on-demand peer feedback system for a clinical knowledge base: A case study with order sets." *Journal of Biomedical Informatics* 41(1): 152-164.

Objective: We have developed an automated knowledge base peer feedback system as part of an effort to facilitate the creation and refinement of sound clinical knowledge content within an enterprise-wide knowledge base. The program collects clinical data stored in our Clinical Data Repository during usage of a physician order entry program. It analyzes usage patterns of order sets relative to their templates and creates a report detailing the usage patterns of the order set template. This report includes a set of suggested modifications to the template. Design: A quantitative analysis was performed to assess the validity of the program's suggested order set template modifications. Measurements: We collected and deidentified 2951 instances of POE-based orders. Our program then identified and generated feedback reports for thirty different order set templates from this data set. These reports contained 500 suggested modifications. Five order set authors were then asked to 'accept' or 'reject' each suggestion contained in his/her respective order set templates. They were also asked to categorize their rationale for doing so ('clinical relevance' or 'user convenience'). Results: In total, 62% (309/500) suggestions were accepted by clinical content authors. Of these, authors accepted 32% (36/114) of the suggested additions, 74% (123/167) of the suggested pre-selections, 76% (16/25) of the suggested de-selections, and 68% (131/194) of the suggested changes in combo box order. Conclusion: Overall, the feedback system generated suggestions that were deemed highly acceptable among order set authors. Future refinements and enhancements to the software will add to its utility. copyright 2007 Elsevier Inc. All rights reserved.

Huntoon, L. R. (2008). "Sham peer review: the Fifth Circuit Poliner decision." *Journal of American Physicians & Surgeons* 13(4): 98-99.

Hyman, N. H., T. Osler, et al. (2009). "Anastomotic leaks after bowel resection: what does peer review teach us about the relationship to postoperative mortality?" *Journal of the American College of Surgeons* 208(1): 48-52.

Background: Anastomotic leak is a dreaded complication of intestinal surgery and has been associated with a high mortality rate. But it is uncertain exactly which patient populations are at risk of death from the leak. We sought to assess the impact of surgeon volume on leak rate and to better understand the relationship of a leak to postoperative mortality. Study Design: All adult patients having a small or large bowel resection with anastomosis at a university hospital from July 2003 to June 2006 were entered into a prospectively maintained quality database; data were entered by a specially trained nurse practitioner who rounded daily with housestaff. Patients with a postoperative leak based on standardized criteria were identified. Patient characteristics, surgical procedure, and operating surgeon were noted. Overall complication and leak rates by surgeon were compared using Fisher's exact test. Individual case review by a group of peers was performed for all patients with a leak who died, to determine the relationship to mortality. Results: Five hundred fifty-six patients underwent resection with anastomosis during the study period. There were 27 patients with leaks (4.9%), 6 of whom died. Leak rate for the highest-volume surgeons ranged from 1.6% to 9.9% ( $p < 0.01$ ), and overall complication rate varied from 30.5% to 44% ( $p = 0.04$ ). In four of six deaths, leaks occurred in very ill patients undergoing emergency procedures and appeared to be pre-morbid events. In only one patient did the leak appear to be the primary cause of death. Conclusions: The variability in leak rate by surgeons doing similar operations suggests that many leaks may be preventable. But death after a leak is most often a surrogate for a critically ill patient and was infrequently the actual cause of death. copyright 2008 American College of Surgeons.

Jain, B. K. and D. Mohanty (2008). "Quality management in clinical practice: Role of stakeholders." *Journal of the Indian Medical Association* 106(3): 188.

Jelley, D. and T. Van Zwanenberg (2003). "Practice-based peer appraisal in general practice: An idea whose time has come?" *Education for Primary Care* 14(3): 329-337.

Practice-based peer appraisal is an emerging activity in UK general practice. It is a process in which doctors working in the same practice engage in a regular review of their clinical and non-clinical performance through both self-reflection and feedback from colleagues. Strengths and weaknesses are discussed, and training needs identified and formulated into a development plan. The plan is reviewed at subsequent appraisal meetings. In this study, which was carried out before the Department of Health guidelines on GP appraisal were published, we review progress in practice-based peer appraisal in the Northern Deanery. The possibility of practices adapting practice-based

appraisal systems in order to fulfil Department of Health requirements for annual appraisal is discussed.

Jimenez, P., C. Borrás, et al. (2006). "Accreditation of diagnostic imaging services in developing countries." *Pan American Journal of Public Health* 20(2-3): 104-12.

In recent decades, medical imaging has experienced a technological revolution. After conducting several surveys to assess the quality and safety of diagnostic imaging services in Latin America and the Caribbean, the Pan American Health Organization (PAHO) developed a basic accreditation program that can be implemented by the ministry of health of any developing country. Patterned after the American College of Radiology's accreditation program, the PAHO program relies on a national accreditation committee to establish and maintain accreditation standards. The process involves a peer review evaluation of: (1) imaging and processing equipment, (2) physician and technologist staff qualifications, (3) quality control and quality assurance programs, and (4) image quality and, where applicable, radiation dose. Public and private conventional radiography/fluoroscopy, mammography, and ultrasound services may request accreditation. The radiography/fluoroscopy accreditation program has three modules from which to choose: chest radiography, general radiography, and fluoroscopy. The national accreditation committee verifies compliance with the standards. On behalf of the ministry of health, the accreditation committee also issues a three-year accreditation certificate. As needed, the accreditation committee consults with foreign technical and clinical experts.

Johnson, C. (2009). "JMPT Peer Review for 2008." *Journal of Manipulative and Physiological Therapeutics* 32(2): 99.

Jolly, B. C. (2007). "The new curriculum framework and assessment practices: current challenges for postgraduate years 1 and 2." *Medical Journal of Australia* 186(7 Suppl): S33-6.

The new curriculum framework for doctors in postgraduate years 1 and 2 is a step towards seamless medical education. The framework will need additional components to make "the curriculum" deliverable. Assessment is an essential element of most curricula, and assessment systems should be carefully planned. Diligent observation and rating in the workplace may provide a suitable approach. In the future, Australia must also thoroughly engage with the debate on continuing validation of competence.

Jones, R. and A. Stephenson (2008). "Quality assurance of community based undergraduate medical curricula: a cross-sectional survey." *Education for Primary Care* 19(2): 135-142.

Because of the increasing contribution of university departments of general practice to the undergraduate curriculum in UK medical schools, involving large numbers of general practices and community based teachers, quality assurance of these 'distributed' curricula is essential. We have conducted a cross-sectional survey of all UK university departments of general practice to determine how they are approaching the task of quality assurance of their community based curricula. The response rate to our survey was over 85%. Most departments now contribute across all years of the undergraduate curriculum providing an average of 13% of total curriculum time. Over 85% of departments were confident that the teaching delivered in their community based programmes was consistent with the aims, objectives and content of their core curricula and also in the quality of their community based teachers. Whilst the majority of departments employed student feedback and feedback from external examiners as part of their quality assurance mechanisms, a minority (24%) used feedback from patients. Almost all departments undertook practice visits and direct observation of teaching and peer review. Most departments would dispense with poorly performing teachers identified through these mechanisms. Quality assurance is being taken seriously by the university departments of general practice. The results of this survey also indicate areas in which both established and newly created medical schools can make further progress towards ensuring high-quality community based undergraduate teaching.

Joshi, R., F. W. Ling, et al. (2004). "Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills." *Academic Medicine* 79(5): 458-63.

**OBJECTIVE:** To test the reliability of the 360-degree evaluation instrument for assessing residents' competency in interpersonal and communication skills. **METHOD:** Ten-item questionnaires were distributed to residents and evaluators at Monmouth Medical Center in Long Branch, New Jersey, in March/April, 2002. The scoring scale was 1-5; the highest score was 50. Data were maintained strictly confidential; each resident was assigned a code. Completed data sheets were collated by category and entered into a spreadsheet. The total and mean scores by each category of evaluator were



calculated for each resident and a rank order list created. Shrout-Fleiss (model 2) intraclass correlation coefficients measured reliability of ratings within each group of evaluators. Reliability/reproducibility among evaluators' scores were tested by the Pearson correlation coefficient ( $p < .05$ ). RESULTS: Intraclass correlation coefficients showed a narrow range, from .85-.54. The highest ranked resident overall ranked high and the lowest was low with most evaluators. The rank order among fellow residents was markedly different from other evaluator categories. Pearson correlation coefficients showed significant correlation between faculty and ancillary staff, ( $p = .002$ ). Patients as evaluators did show intraclass correlation, but did not correlate significantly with other categories. Scores from colleagues correlated negatively with all other categories of evaluators. CONCLUSIONS: The 360-degree instrument appears to be reliable to evaluate residents' competency in interpersonal and communication skills. Information from the assessment may provide feedback to residents. Areas of improvement identified by the scores would suggest areas for improvement and further ongoing assessment.

Kaegi, L. (1999). "AMA Clinical Quality Improvement Forum ties it all together: from guidelines to measurement to analysis and back to guidelines." *Joint Commission Journal on Quality Improvement* 25(2): 95-106.

BACKGROUND: In Chicago October 15, 1998, the American Medical Association (AMA) Department of Clinical Quality Improvement introduced a broadened scope for its Practice Parameters Forum, now retitled the Clinical Quality Improvement Forum. The Forum will now focus on integrating all the components of what the AMA has identified as the quality continuum-clinical practice guidelines, performance measurement, and process and outcomes analysis. KEYNOTE ADDRESS: The address "The Quality Continuum" heralded the ushering in of a third stage known as "clinical integration," which will become medicine's "industrial revolution." Unexplained clinical variation remains the major roadblock to lowering costs, improving quality, and establishing accountability. But several tools, from guidelines to case management, will help us work through our industrial revolution and integrate clinically. PANEL I: GUIDELINES AND EVIDENCE-BASED REPORTS: The international Cochrane Collaboration, through its Collaborative Review Groups and systematic reviews, and the Agency for Health Care Policy and Research, through its Evidence-Based Practice Centers and other programs, are providing the evidence for other groups to use in developing their own guidelines, performance measures, and other tools. But measuring the effectiveness of implementing clinical preventive services remains a difficult challenge. PANEL II: CLINICAL PROCESS MEASUREMENT: Presentations on the relationship between guidelines and performance measures, translating guidelines into review criteria/process measures, reasons for non-compliance with guideline recommendations, patient-reported process measurement, and specific process measurement experiences together reflected federal, national medical specialty society, state medical society/peer review organization, and university/health plan perspectives. PANEL III: CLINICAL OUTCOMES MEASUREMENT: If outcomes can be associated with specific processes of care, clinical practice guidelines can be refined. However, outcomes are dependent on factors outside physicians' behavior and control, such as patient compliance, comorbidities, and other risk factors, and consequently consideration of these variables is essential to appropriately interpreting the data.

Kaiser, J. (2008). "Peer review: Pfizer denied access to journals' files." *Science* 319(5870): 1601.

Kamrath, M., B. J. Johnson, et al. (2008). "Adherence to clinical guidelines in the prevention of coronary heart disease in type II diabetes mellitus." *Pharmacy World & Science* 30(1): 120-7.

OBJECTIVE: The study aimed to assess adherence of prescribed medication in primary care to nationally recognised guideline criteria using case note review applying a previously developed medication assessment tool for coronary heart disease (MAT-CHD). Setting Primary care medical centre serving 17,991 patients. A sample of 208 from 463 patients with type 2 diabetes aged 20-75 years with or without a history of ischaemic heart disease (IHD). METHOD: Patients' records were accessed via medical and pharmacy databases. The criteria of the 23-item audit tool were applied to medical records from case notes in order to quantify adherence to individual guideline criteria. MAIN OUTCOME MEASURE: Frequency of adherence to agreed definitions of medication use quality criteria. RESULTS: A total of 1,433 guideline criteria were applied and 1,107 (77.2%, CI: 75.0, 79.4) criteria standards were met with 326 (22.7%, CI: 20.6, 25.0) non-adherences. The overall adherence to guideline criteria was significantly lower for secondary prevention than for primary prevention (74.4 vs. 80.1%,  $P < 0.05$ ; Chi square). Justification recorded in the case notes was identified for 54 (17%, CI: 13, 21) of those non-adherences. CONCLUSION: The MAT-CHD highlighted areas for review and possible improvement. The tool can be used in primary care from case record examination and offers

a means of co-operation between community pharmacists and general practitioners in clinical guideline implementation.

Keely, E., K. Myers, et al. (2007). "Peer assessment of outpatient consultation letters - Feasibility and satisfaction." *BMC Medical Education* 7(13).

Background. Written correspondence is one of the most important forms of communication between health care providers, yet there is little feedback provided to specialists. The objective of this study was to determine the feasibility and satisfaction of a peer assessment program on consultation letters and to determine inter-rater reliability between family physicians and specialists. Methods. A rating scale of nine 5-point Likert scale items including specific content, style items, education value of the letter and an overall rating was developed from a previous validated tool. Nine Internal Medicine specialists/subspecialists from two tertiary care centres submitted 10 letters with patient and physician identifiers removed. Two Internal Medicine specialists, and 2 family physicians from the other centre rated each letter (to protect writer anonymity). A satisfaction survey was sent to each writer and rater after collation of the results. A follow-up survey was sent 6-8 months later. Results. There was a high degree of satisfaction with the process and feedback. The rating scale information was felt to be useful and appropriate for evaluating the quality of consultation letters by 6/7 writers. 5/7 seven writers felt that the feedback they received resulted in immediate changes to their letters. Six months later, 6/9 writers indicated they had maintained changes in their letters. Raters rank ordered letters similarly (Cronbach's alpha 0.57-0.84) but mean scores were highly variant. At site 1 there were significant differences in scoring brevity ( $p < 0.01$ ) between family physician and specialist raters; whereas, at site 2 there were differences in scoring of history ( $p < 0.01$ ), physical examination ( $p < 0.01$ ) and educational value ( $p < 0.01$ ) of the letter. Conclusion. Most participants found peer assessment of letters feasible and beneficial and longstanding changes occurred in some individuals. Family physicians and specialists appear to have different expectations on some items. Further studies on reliability and validity, with a larger sample, are required before high stakes professional assessments include consultation letters. copyright 2007 Keely et al; licensee BioMed Central Ltd.

Kennedy, D. (2008). "Confidential review--or not?" *Science* 319(5866): 1009.

King, L., G. Radford, et al. (2004). "Evaluation of peer review of infection control arrangements for healthcare-associated infection in four acute care trusts in the East of England." *Communicable Disease & Public Health* 7(4): 245-50.

In order to constructively use Department of Health surveillance data to address the issue of healthcare-associated infection, a peer review process investigating infection control arrangements for healthcare-associated infection in acute care trusts was initiated in the East of England. This study set out to evaluate that process and determine if it should be rolled out regionwide. A questionnaire survey of infection control doctors at trusts visited was conducted by telephone. Respondents commented that the peer review process raised the profile of infection control and healthcare-associated infection in participating trusts, and that a review of routine surveillance data was a useful way to initiate discussion about and focus attention on hospital infection control. Respondents felt the forum for discussion of infection control was most beneficial. A peer review process utilising routine surveillance data is of value in addressing issues related to infection control arrangements for healthcare-associated infection in acute care trusts.

Kraus, L. M. and L. W. Vernaglia (2007). "Rhode Island Supreme Court affirms limited nature of peer-review privilege." *Medicine & Health, Rhode Island* 90(6): 187-8.

Kun, L. E., B. G. Haffty, et al. (2007). "American Board of Radiology Maintenance of Certification-Part IV: Practice Quality Improvement for Radiation Oncology." *International Journal of Radiation Oncology Biology Physics* 68(1): 7-12.

Maintenance of Certification is a physician-based response to public concerns about the quality of medical care and physician competency in a rapidly evolving, technically demanding specialty. The American Board of Radiology (ABR) has previously described the first three components of the Maintenance of Certification. The ABR is currently developing a program in practice performance, completing Part IV of the competencies. The Practice Quality Improvement (PQI) program is meant to critically evaluate meaningful aspects of a physician's practice in a simple manner, using identifiable metrics and self-assessment to include an action plan for quality improvement. Each diplomate will be expected to complete three PQI projects during a full 10-year Maintenance of Certification cycle. Current diplomates with time-limited certificates will find prorated requirements determined by their

year of certification on the ABR Website. Diplomates will have the option of completing zero to two Type I PQI projects (assessing factors relevant to clinical practice by peer review and self-reporting) and one to three Type II projects (i.e., at least one Type II projects of the three required, assessing parameters of practice by comparison with evidence-based guidelines, consensus statements, or peer comparisons; Type II projects are initiated and managed by professional societies). Several examples of Type I projects that might be offered by societies or directly through the ABR are provided, as well as highlights of the two Type II projects that have sought ABR qualification: the American Society for Therapeutic Radiology and Oncology's Performance Assessment for the Advancement of Radiation Oncology Treatment program and American College of Radiology's RO-PEER program. An additional objective of the PQI is to develop national databases for future reference using aggregate data from the PQI projects. copyright 2007 The American Board of Radiology.

Lang, D. A. (1999). Medical staff peer review: motivation and performance in the era of managed care. Chicago, Illinois, Health Forum.

Lanza, M. L., R. A. Zeiss, et al. (2008). "Multiple perspectives on assault: the 360-degree interview." *Journal of the American Psychiatric Nurses Association* 14(6): 413-420.  
Workplace violence is common in health care settings. The authors review various models of this violence that have developed over time. From a linear model, understanding progressed to an interactional and then to a contextual model of assault that examines interactions of the aggressor, victim, and the environment. To date, there has not been a satisfactory research methodology to explore the complexities of the contextual model. This article proposes the 360-degree evaluation as an appropriate methodology for examination of multiple perspectives on assault. The 360-degree model allows comparison of perspectives of the assailant, victim, victim's peers, and victim's supervisor.

Larriviere, D. and H. R. Beresford (2008). "Invited Article: Professionalism in neurology: the role of law." *Neurology* 71(16): 1283-8.  
Professionalism may be defined as the obligation of the physician to uphold the primacy of patients' interests, to achieve and maintain medical competency, and to abide by high ethical standards. Recent commentary has suggested that medical professionalism is being threatened by commercialism and the legal system. Consideration of judicial rulings centered on primacy of patients' interests (informed consent, end-of-life care, and conflicts of interest), medical competence (standard of care in medical malpractice cases, medical futility cases, and confidentiality of peer review), and enforcement of ethical standards (peer review by professional organizations) demonstrates that the law generally defers to standards set by the medical profession, but competing views over what health care model is operative may generate non-deferential outcomes.

Law, S. and J. Bruce (2008). "Multi-source feedback: A salutary tale with a happy ending." *Education for Primary Care* 19(3): 313-316.

Lee, J. K. T. (2007). "Quality-A Radiology Imperative: Interpretation Accuracy and Pertinence." *JACR Journal of the American College of Radiology* 4(3): 162-165.  
Physicians as a group have neither consistently defined nor systematically measured the quality of medical practice. To referring clinicians and patients, a good radiologist is one who is accessible, recommends appropriate imaging studies, and provides timely consultation and reports with high interpretation accuracy. For determining the interpretation accuracy of cases with pathologic or surgical proof, the author proposes tracking data on positive predictive value, disease detection rates, and abnormal interpretation rates for individual radiologists. For imaging studies with no pathologic proof or adequate clinical follow-up, the author proposes measuring the concordance and discordance of the interpretations within a peer group. The monitoring of interpretation accuracy can be achieved through periodic imaging, pathologic correlation, regular peer review of randomly selected cases, or subscription to the ACR's RADPEER system. Challenges facing the implementation of an effective peer-review system include physician time, subjectivity in assessing discordant interpretations, lengthy and equivocal interpretations, and the potential misassignment of false-positive interpretations.

Legge, D. G. and D. E. Richmond (1986). "Continuing education and peer review in internal medicine (CEPRIME)." *Australian & New Zealand Journal of Medicine* 16(5): 679-85.

Research in adult education shows that educational initiatives are most effective when they are related to perceived needs. The CEPRIME system was devised as a means of using peer review to reveal the educational needs of physicians. Sixteen physicians volunteered to test the system, by allowing a panel of specialists to review copies of their reports to referring general practitioners. The comments of the reviewers, both commendatory and critical, were referred back to the participants, and their evaluation of the system was sought. Although the method was found to have limitations relating both to the adequacy of the report as a document for review, and to the objectivity and relevance of the comments made about it, the results were such as to encourage further development of CEPRIME as a means of appraising physicians of their specific educational needs.

Leiker, M. (2008). "Federal Appeals Court reverses lower Court ruling; reaffirms immunity for hospitals and physicians in peer review cases." *Wisconsin Medical Journal* 107(7): 344-346.

Peer review committees in Wisconsin should keep in mind both the Wisconsin peer review immunity statute and HCQIA when structuring their peer review processes. Documenting that actions were taken in good faith is critical. While this is not a consideration under HCQIA, it is under Wisconsin law. Temporary restrictions on a physician's privileges can be appropriate when patient safety requires it; in these circumstances, hospitals should make sure to keep the issue raised by HCQIA's emergency exception. By and large, courts have sided with hospitals in peer review disputes. The physician challenging the peer review action bears the burden of showing by a preponderance of the evidence that no reasonable belief supported the action. While the initial rulings in the Poliner case raised concerns, these rulings were not typical and were ultimately reversed. The recent ruling by the Appeals Court affirms that HCQIA stands on solid ground, provides reassurance to hospitals and physicians engaged in peer review, and advances the goals of Congress by allowing peer review committees to take actions that are in the best interest of patients. At the same time, physicians subject to review retain the right to seek remedies other than money damages and to hold a hospital and its medical staff to a reasonable standard conduct that they can reasonably implement. copyright 2008 Wisconsin Medical Society.

Lelliott, P., R. Williams, et al. (2008). "Questionnaires for 360-degree assessment of consultant psychiatrists: development and psychometric properties." *British Journal of Psychiatry* 193(2): 156-60. BACKGROUND: Expert clinical judgement combines technical proficiency with humanistic qualities. AIMS: To test the psychometric properties of questionnaires to assess the humanistic qualities of working with colleagues and relating to patients using multisource feedback. METHOD: Analysis of self-ratings by 347 consultant psychiatrists and ratings by 4422 colleagues and 6657 patients. RESULTS: Mean effectiveness as rated by self, colleagues and patients, was 4.6, 5.0 and 5.2 respectively (where 1=very low and 6=excellent). The instruments are internally consistent (Cronbach's alpha >0.95). Principal components analysis of the colleague questionnaire yielded seven factors that explain 70.2% of the variance and accord with the domain structure. Colleague and patient ratings correlate with one another ( $r=0.39$ ,  $P<0.001$ ) but not with the self-rating. Ratings from 13 colleagues and 25 patients are required to achieve a generalisability coefficient (Erho(2)) of 0.75. CONCLUSIONS: Reliable 360-degree assessment of humane judgement is feasible for psychiatrists who work in large multiprofessional teams and who have large case-loads.

Lembcke, B. T. (2009). "To be boarded or not to be boarded?... Reprinted from the president's column, *Medical Staff News*." *Baylor University Medical Center Proceedings* 22(1): 60-60.

Lemonick, D. M. (2008). ""Disruptive influences"." *Journal of American Physicians & Surgeons* 13(4): 97-97.

Lerner, S. (1976). "Accreditation of occupational health programs: development and validation of criteria." *American Industrial Hygiene Association Journal* 37(11): 636-9.

Defined standards and a mechanism for evaluation of performance can help promote new and improve existing occupational health programs. Criteria must be based on sound principles, be comprehensive, discriminating and attainable. They should apply to small and large establishments with varying hazards. A successful program requires a multi-disciplinary effort involving physicians, nurses, industrial hygienists, health physicists, safety professionals, managers, technicians and labor. Assessment by on-site peer review is necessary. The criteria and process for their use must be accepted by the professions, labor and management. The Occupational Health/Safety Programs Accreditation Commission plans to use the document entitled Standards, Interpretations and Audit

Criteria for Performance of Occupational Health Programs in its accreditation activities. The development and validation of this document and its use are described.

Lerner, W. (2004). "Development and validation of a 360-degree-feedback instrument for healthcare administrators: Practitioner application." *Journal of Healthcare Management* 49(5): 321-322.

Levine, R. E., P. A. Kelly, et al. (2007). "Peer evaluation in a clinical clerkship: Students' attitudes, experiences, and correlations with traditional assessments." *Academic Psychiatry* 31(1): 19-24.  
Objective: The authors performed this study to determine whether clerkship peer evaluations, initiated as part of our "team-based learning" curriculum in 2002, correlated with other student performance measures, and to determine what qualities students rate in their peer evaluations. Method: The authors correlated peer evaluation scores with other student performance measures and performed a qualitative examination of student comments to assess reasons students gave for giving high and low scores. Results: Peer evaluation scores correlated modestly with the National Board of Medical Examiners' (NBME) subject test, in-class quiz, and clinical scores. Qualitative comments demonstrated that students made assessments based on three thematic areas: personal attributes, team contributions, and cognitive abilities. Conclusions: Peers' evaluation scores modestly predict which students will perform well on other measures. However, there may be other qualities that are also important factors in peer evaluation. For example, most students value qualities of preparation and participation. Though students sometimes dislike peer evaluations, their assessments may enhance traditional course assessments and complement a comprehensive evaluation strategy. Copyright 2007 Academic Psychiatry.

Levinson, W. (2008). "Revalidation of physicians in Canada: are we passing the test?" *CMAJ Canadian Medical Association Journal* 179(10): 979-80.

Lockyer, J., D. Blackmore, et al. (2006). "A study of a multi-source feedback system for international medical graduates holding defined licences." *Medical Education* 40(4): 340-7.

OBJECTIVE: To develop and assess the feasibility and psychometric properties of multi-source feedback questionnaires to monitor international medical graduates practising in Canada under 'defined' licences. METHOD: Four questionnaires (patient, co-worker, colleague and self) were developed and administered in 2 phases through paper-based and telephone or Internet formats. Reliability was assessed with Cronbach's alpha and generalisability coefficient analyses. Validity was established through mean ratings, 'unable to respond' rates and factor analyses. RESULTS: A total of 37 doctors participated in the 2 phases. Overall response rates were 70% for patients, 86% for co-workers, 72% for medical colleagues and 92% for self, with response rates higher for the paper-based format than the Internet and phone formats. The instruments had high internal consistency reliability, with Cronbach's alphas of 0.83 for self-assessment and > 0.90 for the other instruments. The generalisability coefficients were  $Ep(2) = 0.71$  for 25 patients on a 13-item survey,  $Ep(2) = 0.59$  for 8 co-workers on a 13-item survey, and  $Ep(2) = 0.67$  for 8 colleagues on a 21-item questionnaire. The range of mean scores was narrow (between 4 and 5) for all items and all surveys. The factor analyses identified that 2 factors accounted for 70% or more of the variance for the patient and colleague surveys and 60% of the variance for the co-worker survey. CONCLUSION: These data suggest that the instruments have reasonable psychometric properties. Traditional survey methods (i.e. paper-based) yielded better results than Internet or phone methods for this group of doctors.

Lockyer, J. M., C. Violato, et al. (2008). "Assessment of radiology physicians by a regulatory authority." *Radiology* 247(3): 771-778.

Purpose: To determine whether it is possible to develop a feasible, valid, and reliable multisource feedback program for radiologists. Materials and Methods: Surveys with 38, 29, and 20 items were developed to assess individual radiologists by eight radiologic colleagues (peers), eight referring physicians, and eight co-workers (eg, technicians), respectively, by using five-point scales along with an "unable to assess" category. Radiologists completed a self-assessment on the basis of the peer questionnaire. Items addressed key competencies related to clinical competence, collegiality, professionalism, workplace behavior, and self-management. The study was approved by the University of Calgary Conjoint Health Ethics Research Board. Results: Data from 190 radiologists were available. The mean numbers of respondents per physician were 7.5 of eight (1259 of 1520, 83%), 7.15 of eight (1337 of 1520, 88%), and 7.5 of eight (1420 of 1520, 93%) for peers, referring physicians, and co-workers, respectively. The internal consistency reliability indicated all instruments had a Cronbach alpha of more than 0.95. The generalizability coefficient analysis indicated that the

peer, referring physicians, and coworker instruments achieved a generalizability coefficient of 0.88, 0.79, and 0.87, respectively. The factor analysis indicated that four factors on the colleague questionnaire accounted for 70% of the total variance: clinical competence, collegiality, professional development, and workplace behavior. For the referring physician survey, three factors accounted for 64.1% of the variance: professional development, professional consultation, and professional responsibility. Two factors on the co-worker questionnaire accounted for 63.2% of the total variance: professional responsibility and patient interaction. Conclusion: The psychometric examination of the data suggests that the instruments developed to assess radiologists are a feasible way to assess radiology practice and provide evidence for validity and reliability.

Lurie, S. J., D. R. Lambert, et al. (2007). "Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors." *Journal of General Internal Medicine* 22(1): 13-16.

BACKGROUND: It is not known to what extent the dean's letter (medical student performance evaluation [MSPE]) reflects peer-assessed work habits (WH) skills and/or interpersonal attributes (IA) of students. OBJECTIVE: To compare peer ratings of WH and IA of second- and third-year medical students with later MSPE rankings and ratings by internship program directors. DESIGN AND PARTICIPANTS: Participants were 281 medical students from the classes of 2004, 2005, and 2006 at a private medical school in the northeastern United States, who had participated in peer assessment exercises in the second and third years of medical school. For students from the class of 2004, we also compared peer assessment data against later evaluations obtained from internship program directors. RESULTS: Peer-assessed WH were predictive of later MSPE groups in both the second ( $F = 44.90, P < .001$ ) and third years ( $F = 29.54, P < .001$ ) of medical school. Interpersonal attributes were not related to MSPE rankings in either year. MSPE rankings for a majority of students were predictable from peer-assessed WH scores. Internship directors' ratings were significantly related to second- and third-year peer-assessed WH scores ( $r = .32 [P = .15]$  and  $r = .43 [P = .004]$ ), respectively, but not to peer-assessed IA. CONCLUSIONS: Peer assessment of WH, as early as the second year of medical school, can predict later MSPE rankings and internship performance. Although peer-assessed IA can be measured reliably, they are unrelated to either outcome. copyright 2007 Society of General Internal Medicine.

Lurie, S. J., S. Meldrum, et al. (2007). "Changes in self-perceived abilities among male and female medical students after the first year of clinical training." *Medical Teacher* 29(9-10): 921-926.

Background: It is known that male and female medical students have different experiences in their clinical training. Aims: To assess whether male and female medical students change in their self-rated work habits and interpersonal habits during the first year of clinical training. Method: Longitudinal study of self- and peer-assessment among 224 medical students in 3 consecutive classes at a private US medical school. Students rated themselves on global work habits (WH) and interpersonal attributes (IA). Students also rated and were rated by 6-12 peers on the same scale. Results: In the second year of medical school, there were no differences between men and women in quartiles of self-assessed WH or IA. At the end of the third year, however, women were more likely to be in the lower quartiles of self-assessed WH ( $X_{2} = 6.77; p = 0.03$ ), as well as the highest quartiles of self-assessed IA ( $X_{2} = 11.36; p = 0.003$ ). In both years, women rated their own WH skills significantly lower than they rated their peers, while men rated themselves similarly to peers. There were no sex differences in self-assessed IA. Conclusions: Although second-year male and female medical students appear similar to one another in terms of self-assessed WH and IA, by the end of the third year women rate themselves relatively lower in WH, while men rate themselves relatively lower in IA.

Lyman, J. A., J. Schorling, et al. (2008). "Development of a web-based resident profiling tool to support training in practice-based learning and improvement." *Journal of General Internal Medicine* 23(4): 485-488.

Multiple factors are driving residency programs to explicitly address practice-based learning and improvement (PBLI), yet few information systems exist to facilitate such training. We developed, implemented, and evaluated a Web-based tool that provides Internal Medicine residents at the University of Virginia Health System with population-based reports about their ambulatory clinical experiences. Residents use Systems and Practice Analysis for Resident Competencies (SPARC) to identify potential areas for practice improvement. Thirty-three (65%) of 51 residents completed a survey assessing SPARC's usefulness, with 94% agreeing that it was a useful educational tool. Twenty-six residents (51%) completed a before-after study indicating increased agreement (5-point

Likert scale, with 5=strongly agree) with statements regarding confidence in ability to access population-based data about chronic disease management (mean [SD] 2.5 [1.2] vs. 4.5 [0.5],  $p < .001$ , sign test) and information comparing their practice style to that of their peers (2.2 [1.2] vs. 4.6 [0.5],  $p < .001$ ). copyright 2007 Society of General Internal Medicine.

Macaulay, J. O. and P. Nagley (2008). "Student project cases: A learner-centred team activity broadly integrated across the undergraduate medical curriculum." *Medical Teacher* 30(1): e23-e33.

Background: Student Project Cases (SPCs) in the Monash University medical curriculum comprise a teamwork activity that emphasises interdisciplinary learning. SPCs provide a learning program contextualized within the broader medical curriculum, in which students research and present a medical disease. SPCs were introduced into the medical curriculum to integrate more active learning by creating a learning environment which: actively engages students in the learning process and encourages deep approaches to learning, promotes self-directed learning, develops the ability to work in teams, builds information gathering and processing skills, and develops presentation skills. Methods: The design of the SPCs requires students to: make decisions regarding the overall organisation of their group and presentations, communicate and work as members of small groups, research, identify, critically analyse, synthesize and integrate information, and prepare and communicate via written and oral presentations. To ensure effective and consistent assessment practices, an extensive series of simple and effective proformas were developed, including components of peer, self and staff assessment. Assessors utilise objective observations of student performance or achievement; numerical marks are independently based on specific descriptors aligned with learning objectives. The effectiveness of SPCs was evaluated through questionnaires, focus groups and informal feedback, involving both students and staff. Conclusion: The assessment proformas were successful learning aids for the students and allowed staff to critique student performance in a non-subjective manner. The SPCs are a successful active learning program which engages students and encourages deep learning and could readily be adapted by a diverse range of courses. The SPCs also develop graduate attributes of life-long learning skills including: capacity for inquiry and research, critical thought and analysis, problem solving, teamwork and written and oral communication. Additional benefits encompass enhanced student appreciation of learning and assessment as well as development of confidence in addressing new or complex tasks.

Machado, J. L. M., V. M. P. Machado, et al. (2008). "Self- and peer assessment may not be an accurate measure of PBL tutorial process." *BMC Medical Education* 8(55).

Background. Universidade Cidade de Sao Paulo adopted a problem-based learning (PBL) strategy as the predominant method for teaching and learning medicine. Self-, peer- and tutor marks of the educational process are taken into account as part of the final grade, which also includes assessment of content. This study compared the different perspectives (and grades) of evaluators during tutorials with first year medical students, from 2004 to 2007 ( $n = 349$ ), from seven semesters. Methods. The tutorial evaluation method was comprised of the students' self assessment (SA) (10%), tutor assessment (TA) (80%) and peer assessment (PA) (10%) to calculate a final educational process grade for each tutorial. We compared these three grades from each tutorial for seven semesters using ANOVA and a post hoc test. Results. A total of 349 students participated with 199 (57%) women and 150 (42%) men. The SA and PA scores were consistently greater than the TA scores. Moreover, the SA and PA groups did not show statistical difference in any semester evaluated, while both differed from tutor assessment in all semesters (Kruskal-Wallis, Dunn's test). The Spearman rank order showed significant ( $p < 0.0001$ ) and positive correlation for the SA and PA groups ( $r = 0.806$ ); this was not observed when we compared TA with PA ( $r = 0.456$ ) or TA with SA ( $r = 0.376$ ). Conclusion. Peer- and self-assessment marks might be reliable but not valid for PBL tutorial process, especially if these assessments are used for summative assessment, composing the final grade. This article suggests reconsideration of the use of summative assessment for self-evaluation in PBL tutorials.

Mackenzie, G. M., S. F. Hayes, et al. (2005). "Where next with revalidation?" *British Medical Journal* 331(7512): 352-353.

Mackway-Jones, K. and S. Carley (2003). "Accumulating knowledge: Five years of BETs." *Emergency Medicine Journal* 20(4): 304-305.

MacLean, P. A. (2007). "Court clears disclosure of doctor peer reviews: third federal court won't extend privilege; gap with states widens." *Medical Malpractice Law & Strategy* 24(11): 1.

Maher, E. J. (2008). "An integrated support and information centre in a large U.K. cancer centre established in 1993 and replicated in more than 60 units across the United Kingdom and Australia." *Current Oncology* 15 Suppl 2: s108.es64-7.

Established in 1993 after a 2-year consultation between professionals and cancer patients, the Lynda Jackson Macmillan Centre (ljmc) has been a catalyst for change in the United Kingdom. The Centre began with a small core staff in a purpose-built building next to a cancer centre, networking with outreach workers in 12 surrounding hospitals, with a mission to improve information, communication, and support for cancer patients. Since 1996, the ljmc model has been adopted and developed by the charity Macmillan Cancer Support and has been spread to more than 60 units across the United Kingdom and Australia. Introducing complementary therapies (cams) to a cancer centre was a particular early challenge. Establishing a shared understanding of the role of complementary therapies and developing nationally accredited written information about them, credible recruitment and governance procedures for therapy practitioners, agreed outcome measures, and peer-reviewed evaluation and research have all been important in engaging cancer physicians and managers; however, charitable funding is still required to support free access to most complementary therapies. An integrated supportive care service for cancer patients begins with a shift in the culture of cancer treatment organizations, moving from a professional-centred to a patient-centred agenda. Real reach and impact requires "new" ideas and services to be integrated into the routine practice of the cancer care delivery organizations. A key lesson learned over the last 15 years is that an integrated support centre must continually adapt to be viable. Sustaining meaningful user guidance is a particular challenge. Support for self-management and the testing and development of cam services are growing parts of the portfolio.

Maker, V. K. and M. B. Donnelly (2008). "Surgical Resident Peer Evaluations-What Have We Learned." *Journal of Surgical Education* 65(1): 8-16.

Purpose: Residents spend most of their time in the hospital with their respective junior and senior colleagues. Therefore, residents have a unique and valuable insight into their peers' professional qualities. Peer evaluation of values and virtues of each group may bring a different or an even better aspect of evaluating some ACGME core competencies. The purpose of this study was to identify the characteristics of junior and senior residents who are perceived to be role models in contrast to those who are not and to determine among these characteristics which is the most influential factor(s). Methods: Seventeen senior residents (4th and 5th) and 26 junior residents (1st, 2nd, and 3rd) evaluated each other in an anonymous manner 6 months apart on 3 separate occasions. Questionnaires for each group were designed by their evaluating group. Each questionnaire listed the qualities of a role-model resident anchored on a scale of 1-3. Results: A 1-way analysis of variance indicated that most of the time, the senior residents are perceived to be role models (the mean role-model rating = 2.87). A partial eta-squared analysis was performed to determine how well the specific performance factors differentiate the cases in which senior residents were perceived to be role models and when they were not. Clearly, professionalism along with attitude/expectations was the most differentiating factor. Similar statistical analyses were performed on seniors evaluating juniors. The role-model mean ratings were different significantly from one another ( $p = 0.014$ ), which indicated that the senior residents varied significantly in the degree to which juniors were perceived to be role models. Conclusion: In each group, both junior and senior residents value professionalism as the most important differentiating attribute of a role model. Seniors felt that knowledge was a less important attribute of a junior role model. Juniors felt teaching inside and outside of the operating room was a less important attribute of a senior role model. copyright 2008 Association of Program Directors in Surgery.

Malling, B., T. Bonderup, et al. (2009). "Effects of multi-source feedback on developmental plans for leaders of postgraduate medical education." *Medical Education* 43(2): 159-67.

OBJECTIVES: Multi-source feedback (MSF) is a widely used developmental tool for leaders in organisations including those dealing with health care. This study was performed to examine the effects of an MSF process on developmental plans made by leaders of postgraduate medical education (PGME) in clinical departments. METHODS: An MSF instrument was developed based on literature on the subject and previous investigations. The instrument was used by consultants responsible for PGME in clinical departments (CREs). Apart from CREs' self-ratings, MSF responses were collected from heads of departments, consultants and young doctors-in-training. The MSF process included individual feedback as well as guidance on drafting developmental plans for both the department and the CREs. Themes emerging in the developmental plans were analysed and compared with the areas in need of improvement identified by the MSF process. RESULTS: The MSF



instrument was found to be feasible, valid and reliable (Cronbach's alpha = 0.98). The study included 52 CREs from various specialties and 498 MSF respondents. The response rate was 84%. Low ratings and negative gaps between others' ratings and self-ratings were identified for both management and leadership performance areas. The developmental plans mainly focused on management initiatives, whereas plans for the development of leadership performance were few. Areas rated low by all respondents were scarcely represented in CREs' developmental plans. CONCLUSIONS: An MSF process might in itself lead to development in administrative areas. However, MSF carried through as a single stand-alone procedure was not sufficient to foster plans for the development of leadership performance.

Margalit, A. P., S. M. Glick, et al. (2007). "A practical assessment of physician biopsychosocial performance." *Medical Teacher* 29(8): e219-e226.

Background: A biopsychosocial approach to care seems to improve patient satisfaction and health outcomes. Nevertheless, this approach is not widely practiced, possibly because its precepts have not been translated into observable skills. Aim: To identify the skill components of a biopsychosocial consultation and develop an tool for their evaluation. Methods: We approached three e-mail discussion groups of family physicians and pooled their responses to the question "what types of observed physician behavior would characterize a biopsychosocial consultation?" We received 35 responses describing 37 types of behavior, all of which seemed to cluster around one of three aspects: patient-centered interview; system-centered and family-centered approach to care; or problem-solving orientation. Using these categories, we developed a nine-item evaluation tool. We used the evaluation tool to score videotaped encounters of patients with two types of doctors: family physicians who were identified by peer ratings to have a highly biopsychosocial orientation (n = 9) or a highly biomedical approach (n = 4); and 44 general practitioners, before and after they had participated in a program that taught a biopsychosocial approach to care. Results: The evaluation tool was found to demonstrate high reliability (alpha = 0.90) and acceptable interobserver variability. The average scores of the physicians with a highly biopsychosocial orientation were significantly higher than those of physicians with a highly biomedical approach. There were significant differences between the scores of the teaching-program participants before and after the program. Conclusions: A biopsychosocial approach to patient care can be characterized using a valid and easy-to-apply evaluation tool.

Marks, J. E. and J. Armbruster (1992). "Accreditation of radiation oncology in the United States." *International Journal of Radiation Oncology Biology Physics* 24(5): 857-860.

Mason, R. and H. Hayes (2007). "Telephone peer supervision and surviving as an isolated consultant." *Psychiatric Bulletin* 31(6): 215-217.

Mason, R., L. Zouita, et al. (2001). "Assessing clinical competence and revalidation of clinicians. Results from pilot study using portfolio and 360 degrees questionnaire." *BMJ* 322(7302): 1601.

Massagli, T. L. and J. D. Carline (2007). "Reliability of a 360-degree evaluation to assess resident competence." *American Journal of Physical Medicine & Rehabilitation* 86(10): 845-852.

OBJECTIVE:: To determine the feasibility and psychometric qualities of a 360-degree evaluation of physical medicine and rehabilitation (PM&R) residents' competence. DESIGN:: Nurses, allied health staff, and medical students completed a 12-item questionnaire after each PM&R resident rotation from January 2002 to December 2004. The items were derived from five of the six competencies defined by the Accreditation Council for Graduate Medical Education (ACGME). RESULTS:: Nine hundred thirty evaluations of 56 residents were completed. The alpha reliability coefficient for the instrument was 0.89. Ratings did not vary significantly by resident gender. Senior residents had higher ratings than junior residents. A reliability of >0.8 could be achieved by ratings from just five nurses or allied health staff, compared with 23 ratings from medical students. Factor analysis revealed all items clustered on one factor, accounting for 84% of the variance. In a subgroup of residents with low scores, raters were able to differentiate among skills. CONCLUSION:: Resident assessment tools should be valid, reliable, and feasible. This Web-based 360-degree evaluation tool is a feasible way to obtain reliable ratings from rehabilitation staff about resident behaviors. The assignment of higher ratings for senior residents than junior residents is evidence for the general validity of this 360-degree evaluation tool in the assessment of resident performance. Different rater groups may need distinct instruments based on the exposure of rater groups to various resident activities and behaviors.

McCormack, W. T., C. Lazarus, et al. (2007). "Peer nomination: A tool for identifying medical student exemplars in clinical competence and caring, evaluated at three medical schools." *Academic Medicine* 82(11): 1033-1039.

**PURPOSE:** Peer evaluation is underused in medical education. The goals of this study were to validate in a multiinstitutional study a peer nomination form that identifies outstanding students in clinical competency and interpersonal skills, to test the hypothesis that with additional survey items humanism could be identified as a separate factor, and to find the simplest method of analysis. **METHOD:** In 2003, a 12-item peer nomination form was administered to junior or senior medical students at three institutions. Factor analysis was used to identify major latent variables and the items related to those characteristics. On the basis of those results, in 2004 a simpler, six-item form was developed and administered. Student rankings based on factor analysis and nomination counts were compared. **RESULTS:** Factor analysis of peer nomination data from both surveys identified three factors: clinical competence, caring, and community service. New survey items designed to address humanism are all weighted with interpersonal skills items; thus, the second major factor is characterized as caring. Rankings based on peer nomination results analyzed by either factor analysis or simply counting nominations distinguish at least the top 15% of students for each characteristic. **CONCLUSIONS:** Counting peer nominations using a simple, six-item form identifies medical student exemplars for three characteristics: clinical competence, caring, and community service. Factor analysis of peer nomination data did not identify humanism as a separate factor. Peer nomination rankings provide medical schools with a reliable tool to identify exemplars for recognition in medical student performance evaluations and selection for honors (e.g., Gold Humanism Honor Society). Copyright 2007 Association of American Medical Colleges.

McCormick, D. (2008). "Peer review under review within NIH." *BioTechniques* 44(4): 451.

McGregor, J. R., P. Bowie, et al. (2005). "A pilot study of the feasibility of trainee hospital doctors undertaking significant event analysis." *British Journal of Hospital Medicine* 66(11): 638-41.

Participating in significant event analysis is proposed as an important method of reflective learning that can enhance patient care and safety. Peer review is one way of making informed judgments on the ability of participants to apply the technique.

McKay, J., A. Shepherd, et al. (2008). "Acceptability and educational impact of a peer feedback model for significant event analysis." *Medical Education* 42(12): 1210-1217.

**Context:** A model of independent, external review of significant event analysis by trained peers was introduced by NHS Scotland in 1998 to support the learning needs of general practitioners (GPs). Engagement with this feedback model has increased over time, but participants' views and experiences are largely unknown and there is limited evidence of its educational impact. This is important if external feedback is to play a potential role in appraisal and future revalidation. **Objective:** The study aimed to explore aspects of the acceptability and educational impact of this external feedback model with participating GPs. **Methods:** Semi-structured interviews were carried out with nine GPs. Participants were sampled to reflect their level of learning need (low, moderate or high) to gain a range of views and experiences. Transcribed interviews were analysed for content. **Results:** This system of external peer feedback is generally acceptable to participants. It complemented and enhanced the appraisal process. External feedback had positive educational outcomes, particularly in imparting technical knowledge on how to analyse significant events. Training issues for peer reviewers were suggested that would further enhance the educational gain from participation. There was disagreement over whether this type of feedback could or should be used as supporting evidence of the quality of doctors' work to educational and regulatory authorities. **Conclusions:** The findings add to the evidence for the acceptability and educational impact of external review by trained peers. Aligning such a model with the current national appraisal system may provide GPs with a more robust demonstration of participation in reflective learning.

McKinstry, B., H. Peacock, et al. (2005). "GP experiences of partner and external peer appraisal: A qualitative study." *British Journal of General Practice* 55(516): 539-543.

**Background:** Appraisal is being adopted both in the UK and internationally as a means of aiding personal development for family doctors. However, it is not clear by whom they should be appraised. **Aim:** To explore attitudes of GPs towards being appraised by externally appointed GP colleagues and by their own partners. **Design of study:** Semi-structured interviews of GPs who had experienced both forms of appraisal. **Setting:** Lothian, Scotland. **Method:** Sixty-six GPs agreed to take part in a study of partner (n = 46) and external (n = 20) peer-based appraisal. Six months later this group was followed

up by questionnaire to determine views of the process, in order to obtain a purposeful sample of 13 GPs who were interviewed in depth. Results: We uncovered concern and a need for clarity about the linkage of appraisal to revalidation. Interviewees felt that the potentially charged nature of appraisal could lead to collusion between appraiser and appraisee, which may lead to a superficial engagement. Similarly, lack of local knowledge of an appraisee potentially enabled a strategy of avoidance. GPs opting for partner appraisal were less likely to undergo appraisal due to lack of protected time. Conclusion: There are reported advantages and disadvantages to having an external peer or partner appraisal. The relationship between revalidation and appraisal needs to be clarified as this leads to collusion and avoidance strategies by both appraisers and appraisees. Good training is required to both recognise and address these strategies. Protected time is essential for effective appraisal. Copyright British Journal of General Practice.

McLellan, H., H. Bateman, et al. (2005). "The place of 360 degree appraisal within a team approach to professional development." *Journal of Interprofessional Care* 19(2): 137-48.

The role of appraisal within professional development has assumed a higher profile in response to national policies including those relating to clinical governance and workforce development. Under the guidance of a service development manager trained in human resource management, a new interprofessional Primary Care Practice Team near Cambridge defined and implemented a new appraisal system drawing on the principles of 360 degree appraisal. The appraisal system was designed to support the service function and the team culture, and took into account good management practice and the requirements of partner organizations and professional affiliations. This paper describes how the system was introduced, tailored to the particular needs of the interprofessional practice team, and evaluated. Early experience suggests that the approach enhances understanding of roles and responsibilities and is supportive to interprofessional team development. It is well received by team members involved and by some of the external partner organizations. However, it is demanding on resources and concerns were raised about the completeness of this approach to appraisal.

McMillan, R. and N. Cameron (2006). "Factors influencing the submission of videotaped consultations by general practitioners for peer review and educational feedback." *Quality in Primary Care* 14(2): 85-89.

Effective communication can both enhance patient satisfaction and lead to improved clinical outcomes. Peer review of consultations in general practice using video has been identified as an effective means of communication skills training. A postal survey of general practitioners in the west of Scotland was undertaken to identify motivating factors to submit a videotape of consultations for peer review and to identify barriers to engaging in this process. Analysis of the results demonstrated a number of motivating factors and perceived obstacles and also identified issues regarding the relevance of the feedback received by participants. These findings are examined, highlighting the benefits of this activity, and the paper discusses how these can be enhanced and how any difficulties may be minimised. The paper considers how the quality of feedback provided could be improved and delivered more effectively, ultimately aiming to improve the quality of patient care. Although the results may be considered predictable, the authors are unaware of any previous analysis of these factors. copyright 2006 Radcliffe Publishing.

Meadows, J. (2007). "Embracing clinical supervision." *British Journal of General Practice* 57(538): 412.

Merkur, S., E. Mossialos, et al. (2008). "Physician revalidation in Europe." *Clinical Medicine, Journal of the Royal College of Physicians of London* 8(4): 371-376.

Despite the increasing attention on patient mobility, there remains a lack of European-level interest in assuring the sustained competence of health professionals. Specifically, the existing European legal framework fails to recognise the introduction of periodic revalidation and requirements to participate in continuing professional development in some countries. This study shows that the definitions and mechanisms of revalidation vary significantly across member states. While some countries, eg Austria, Germany and Spain, look to continuing medical education as a means to promote recertification and quality of care, other countries, eg Belgium, France and the Netherlands, also incorporate peer review. In the UK the proposed revalidation scheme would include elements of relicensure through appraisal and feedback as well as physician recertification. Divergence between countries also exists in monitoring and enforcement. The European Commission should explore the

implications for professional mobility of the diversity in the regulation of the medical profession. copyright Royal College of Physicians, 2008. All rights reserved.

Miles, A. and J. Holden (2007). "Fellowship by Assessment 1989-2006." *British Journal of General Practice* 57(534): 74-75.

Moller, L. B. K. and M. W. Ejlskov (2008). "Evaluation of the 360-degree assessment method in a hospital." *Ugeskrift for Laeger* 170(38): 2973-6.

The present study examines the acceptability of the 360-degree assessment method as a means for evaluating the management and leadership competencies of the clinical staff of a university hospital. Twenty-eight consultants and registered nurses underwent evaluation. One group had debriefing with management consultants, the other with the head of the clinical department. Two months later, the applicability of the method was assessed. The strengths and weaknesses of the leaders were exposed, and areas for improvement were made visible, and acceptance of the method was widespread. Anonymity was required. The group coached by management consultants tended to benefit the most from the evaluation. Using a web-based solution to collect the data was unproblematic.

Morreim, E. H. (1993). "Am I my brother's warden? Responding to the unethical or incompetent colleague." *Hastings Center Report* 23(3): 19-27.

Moss, A. C., T. Dugal, et al. (2005). "Attitudes to peer review as a competence assurance structure - Results of a survey of Irish physicians." *Irish Journal of Medical Science* 174(3): 43-46.

Background: The Medical Council in Ireland is currently implementing Competence Assurance Structures (CAS). Peer review has been proposed as a tool to measure physician competence. Aims: To assess the attitudes of physicians working in the Irish healthcare system to a peer review programme of competence assurance prior to its implementation. Methods: A postal survey was sent to all physicians in the Irish Medical Directory in November 2003. Nine questions were asked to gauge attitudes to peer review as a CA tool. The returned questionnaires were collated and data extracted based on responses. Results: The response rate was 67%. The majority of respondents (92%) felt peer review would inform competence assurance in Ireland. Most physicians who were surveyed felt an on-site assessment (88%) every 5 years (87%) was the preferred method. Over 30% responded that there should be a financial incentive for completing a review, and 70% would pay to be assessed. The UK model of competence assurance was the model most physicians preferred for the Irish setting (42%). Conclusion: The majority of physicians practising in Ireland would favour a peer review system of competence assurance. The financial implications, and structure, of such a system would need to be explored prior to implementation.

Munshi, A. and N. Khattry (2008). "To grade a clinician." *Indian Journal of Medical Ethics* 5(4): 176.

Murphy, D. J., D. Bruce, et al. (2008). "Workplace-based assessment for general practitioners: using stakeholder perception to aid blueprinting of an assessment battery." *Medical Education* 42(1): 96-103.

CONTEXT: The implementation of an assessment system may be facilitated by stakeholder agreement that appropriate qualities are being tested. This study investigated the extent to which stakeholders perceived 8 assessment formats (multiple-choice questions, objective structured clinical examination, video, significant event analysis, criterion audit, multi-source feedback, case analysis and patient satisfaction questionnaire) as able to assess varying qualities of doctors training in UK general practice. METHODS: Educationalists, general practice trainers and registrars completed a blueprinting style of exercise to rate the extent to which each evaluation format was perceived to assess each of 8 competencies derived primarily from the General Medical Council document 'Good Medical Practice'. RESULTS: There were high levels of agreement among stakeholders regarding the perceived qualities tested by the proposed formats ( $G = 0.82-0.93$ ). Differences were found in participants' perceptions of how well qualities were able to be assessed and in the ability of the respective formats to test each quality. Multi-source feedback (MSF) was expected to assess a wide range of qualities, whereas Probity, Health and Ability to work with colleagues were limited in terms of how well they could be tested by the proposed formats. DISCUSSION: Awareness of the perceptions of stakeholders should facilitate the development and implementation of workplace-based assessment (WPBA) systems. These data shed light on the acceptability of various formats in a way

that will inform further investigation of WPBA formats' validity and feasibility, while also providing evidence on which to base educational efforts regarding the value of each format.

Murphy, D. J., D. A. Bruce, et al. (2009). "The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom." *Advances in Health Sciences Education* 14(2): 219-32.

To investigate the reliability and feasibility of six potential workplace-based assessment methods in general practice training: criterion audit, multi-source feedback from clinical and non-clinical colleagues, patient feedback (the CARE Measure), referral letters, significant event analysis, and video analysis of consultations. Performance of GP registrars (trainees) was evaluated with each tool to assess the reliabilities of the tools and feasibility, given raters and number of assessments needed. Participant experience of process determined by questionnaire. 171 GP registrars and their trainers, drawn from nine deaneries (representing all four countries in the UK), participated. The ability of each tool to differentiate between doctors (reliability) was assessed using generalisability theory. Decision studies were then conducted to determine the number of observations required to achieve an acceptably high reliability for "high-stakes assessment" using each instrument. Finally, descriptive statistics were used to summarise participants' ratings of their experience using these tools. Multi-source feedback from colleagues and patient feedback on consultations emerged as the two methods most likely to offer a reliable and feasible opinion of workplace performance. Reliability co-efficients of 0.8 were attainable with 41 CARE Measure patient questionnaires and six clinical and/or five non-clinical colleagues per doctor when assessed on two occasions. For the other four methods tested, 10 or more assessors were required per doctor in order to achieve a reliable assessment, making the feasibility of their use in high-stakes assessment extremely low. Participant feedback did not raise any major concerns regarding the acceptability, feasibility, or educational impact of the tools. The combination of patient and colleague views of doctors' performance, coupled with reliable competence measures, may offer a suitable evidence-base on which to monitor progress and completion of doctors' training in general practice.

Murphy, J. F. A. (2008). "Peer review confidentiality survives legal challenge." *Irish Medical Journal* 101(5): 132.

Naeem, A., A. Kent, et al. (2007). "Foundation programme assessment tools in psychiatry." *Psychiatric Bulletin* 31(11): 427-430.

Nair, B. R., H. G. Alexander, et al. (2008). "The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates." *Medical Journal of Australia* 189(3): 159-161.

Nash, D. B., L. E. Markson, et al. (1993). "Evaluating the competence of physicians in practice: From peer review to performance assessment." *Academic Medicine* 68(2 SUPPL.): S19-S22.

Evaluation of physician competence has traditionally been defined by the medical profession largely through standardized tests and board certifying examinations. However, a level of physician evaluation that takes place outside academic medical centers and certifying boards is rapidly developing. This article describes three programs for such physician evaluation: (1) the program of US Healthcare, a national managed health care company; (2) the DEMPAQ-Developing and Evaluating Methods to Promote Ambulatory Care Quality project, a joint research effort between Harvard University (Cambridge, Massachusetts) and the State of Maryland's (and District of Columbia's) Peer Review Organization (PRO); and (3) a project initiated by the American College of Physicians (ACP) in Philadelphia, Pennsylvania, to assess both medical competence and technical performance in the hospital setting. The authors argue for the need to inform physicians-in-training about the types of evaluation to which they will be subjected over the courses of their careers in order to further this goal, the authors advocate increased collaboration between leaders in the academic setting and those spearheading these new programs for assessment of physician performance.

Nayak, B. K., R. Maniar, et al. (2005). "The agony and the ecstasy of the peer-review process." *Indian Journal of Ophthalmology* 53(3): 153-155.

Nicol, J. and C. Hornsby (2008). "Peer review of a method for the quarterly assessment of specialty registrars." *Education for Primary Care* 19(3): 317-320.

Nielsen, H. G. and K. Bering (2008). "Supervision of practising physicians." *Ugeskrift for Laeger* 170(21): 1847.

Nijm, L. M. (2003). "2003 Schwartz Award. Pitfalls of peer review. The limited protections of state and federal peer review law for physicians." *Journal of Legal Medicine* 24(4): 541-56.

Nikendei, C., B. Kraus, et al. (2007). "An innovative model for teaching complex clinical procedures: Integration of standardised patients into ward round training for final year students." *Medical Teacher* 29(2-3): 246-252.

**Objectives:** Ward rounds are an essential activity for doctors in hospital settings and represent complex tasks requiring not only medical knowledge but also communication skills, clinical technical skills, patient management skills and team-work skills. However, although the need for ward round training is emphasized in the published literature, there are currently no reports of ward round training in a simulated setting with standardized patients. **Methods:** 45 final year students participated in a ward round training session lasting two hours with three standardized patient scenarios and role-plays. Final year students assumed the role of either doctor, nurse or final year student with role-specific instructions and provided each other with peer-feedback during the training session. Training was assessed using final year student focus groups and semi-structured interviews of standardized patients. Written protocols of the focus group as well as the interviews of standardized patients were content analysed. **Results:** In the course of five focus groups, 204 individual statements were gathered from participating final year students. Ward round training proved to be a feasible tool, well accepted by final year students. It was seen to offer a valuable opportunity for reflection on the processes of ward rounds, important relevant feedback from standardized patients, peer group and tutors. Semi-structured standardized patient interviews yielded 17 central comments indicating that ward rounds are a novel and exciting experience for standardized patients. **Conclusion:** Ward round training with standardized patients is greatly appreciated by final year students and is viewed as an important part of their education, easing the transition from observing ward rounds to conducting them on their own.

Norman, G. R., D. A. Davis, et al. (1993). "Competency assessment of primary care physicians as part of a peer review program." *Journal of the American Medical Association* 270(9): 1046-1051.

**Objective.** - To design and test a program that assesses clinical competence as a second stage in a peer review process and to determine the program's reliability. **Design and Setting.** - A three-cohort study of Ontario primary care physicians. **Participants.** - Reference physicians (n=26) randomly drawn from the Hamilton, Ontario, area; volunteer, self-referred physicians (n=20); and physicians referred by the licensing body (n=37) as a result of a disciplinary hearing or peer review. **Main Outcome Measures.** - Standardized patients, structured oral examinations, chart-stimulated recall, objective structured clinical examination, and multiple-choice examination. **Results.** - Test reliability was high, ranging from 0.73 to 0.91, and all tests discriminated among subgroups. Demographic variables relating to the final category were age, Canadian or foreign graduates, and whether or not participants were certified in family medicine. **Conclusions.** - The study demonstrated the feasibility, reliability, and validity of a multicomponent examination in the peer review process.

O'Brodovich, H., J. Beyene, et al. (2007). "Performance of a career development and compensation program at an academic health science center." *Pediatrics* 119(4): e791-e797.

**OBJECTIVE.** The academic physicians of our department developed a novel Career Development and Compensation Program to outline job expectations, enhance career development, and provide a peer-review process to assess performance. The Career Development and Compensation Program was founded on the principle that sustained achievement in education, clinical care, or research should be valued, supported, and rewarded in an equivalent manner and that reward for clinical work should not be limited by the focus of the university on research and education. The objective of this study was to determine whether the principles of the Career Development and Compensation Program were sustained during the initial 7 years of its implementation. **METHODS.** The outcome of the 7 triennial reviews that occurred from 1999 to 2005 was evaluated. For the purposes of some analyses, physicians were classified as predominately clinical (clinician-specialists and clinician-teachers), predominately education (clinician-educators), or predominately research (clinician-investigators and clinician-scientists). **RESULTS.** Each of the job profiles had a similar probability to increase a level within the Career Development and Compensation Program at the time of triennial review. Similarly, all 5 job profiles had a similar rate of increase in their level in relation to the total number of years of experience at an academic health science center. Neither the university academic

rank nor gender of the physician affected the probability of increasing a level at the time of the triennial review. CONCLUSION. The peer-reviewed Career Development and Compensation Program recognizes sustained achievement in each area of education, clinical care, and research in an equivalent manner with no detectable effect of academic rank or gender.

Oehlert, W. H. and S. D. Ketring (1997). "Accreditation: who's counting?" *Journal - Oklahoma State Medical Association* 90(2): 48-53.

Active physician involvement and leadership in their accreditation process can produce a cubic win for patients, payers, and providers. For health care quality to improve and everyone win, physicians need to understand the accountability system, the what and why of data collection, and be involved in short- and long-term performance assessments.

Ouellet, R. (2009). "Revalidation of Canadian physicians." *Canadian Medical Association Journal* 180(5): 540.

Overeem, K., M. J. Faber, et al. (2007). "Doctor performance assessment in daily practice: does it help doctors or not? A systematic review." *Medical Education* 41(11): 1039-1049.

CONTEXT: Continuous assessment of individual performance of doctors is crucial for life-long learning and quality of care. Policy-makers and health educators should have good insights into the strengths and weaknesses of the methods available. The aim of this study was to systematically evaluate the feasibility of methods, the psychometric properties of instruments that are especially important for summative assessments, and the effectiveness of methods serving formative assessments used in routine practise to assess the performance of individual doctors. METHODS: We searched the MEDLINE (1966-January 2006), PsychINFO (1972-January 2006), CINAHL (1982-January 2006), EMBASE (1980-January 2006) and Cochrane (1966-2006) databases for English language articles, and supplemented this with a hand-search of reference lists of relevant studies and bibliographies of review articles. Studies that aimed to assess the performance of individual doctors in routine practise were included. Two reviewers independently abstracted data regarding study design, setting and findings related to reliability, validity, feasibility and effectiveness using a standard data abstraction form. RESULTS: A total of 64 articles met our inclusion criteria. We observed 6 different methods of evaluating performance: simulated patients; video observation; direct observation; peer assessment; audit of medical records, and portfolio or appraisal. Peer assessment is the most feasible method in terms of costs and time. Little psychometric assessment of the instruments has been undertaken so far. Effectiveness of formative assessments is poorly studied. All systems but 2 rely on a single method to assess performance. DISCUSSION: There is substantial potential to assess performance of doctors in routine practise. The long term impact and effectiveness of formative performance assessments on education and quality of care remains hardly known. Future research designs need to pay special attention to unmasking effectiveness in terms of performance improvement.

Owen, C. and W. Ramsey (2007). "Medical student peer review: Making it work." *Clinical Teacher* 4(1): 31-35.

Owens, W. D. and F. P. Hughes (1994). "The American Board of Anesthesiology." *Annals of the Academy of Medicine, Singapore* 23(4): 619-22.

The American Board of Anesthesiology (ABA) has been in existence since 1937 for the purpose of advancing the practice of anaesthesiology in the United States. It is a credentialing organization which requires anaesthesiology residency training, medical licensure, and successful completion of written and oral examinations in order for an individual physician to be granted a diploma. Residency training requires four years and must culminate with a satisfactory certificate of clinical competence. The written and oral examinations follow residency training. The ABA also issues subspecialty certificates in critical care medicine and pain management. There is a voluntary mechanism for diplomates to demonstrate continuing qualifications. This is accomplished by peer review and a written examination.

Paige, J. T., D. L. Aaron, et al. (2008). "Implementation of a preoperative briefing protocol improves accuracy of teamwork assessment in the operating room." *American Surgeon* 74(9): 817-23.

This study examined the effect of implementing a new preoperative briefing protocol on self- and peer-assessments of individual operating room (OR) teamwork behaviors. From July 2006 to February 2007, OR teamwork performance at a rural community hospital was evaluated before and

after training and implementation of the protocol. After each case, every member on the team completed a 360-degree type teamwork behavior evaluation containing both self- and peer-assessments using a six-point Likert type scale (1 = definitely no to 6 = definitely yes). Individual behavior change was measured using the mean scale score of pre and postprotocol assessments. Statistical analysis included t test for both pre/post and self/peer differences. Data were available for one general surgeon and nine OR staff (pre = 20 cases, post = 16 cases). The preprotocol self-assessment mean score was significantly higher than peer-assessment (5.63 vs 5.29,  $P < 0.0267$ ). Pre and postprotocol peer assessment mean scores revealed a statistically significant gain in teamwork behaviors. No difference was observed in postassessment mean scores for self- and peer-assessments. Individuals overestimated their teamwork behaviors before protocol implementation. Using a preoperative protocol seems to improve OR staff teamwork behaviors and self-assessment accuracy. The use of a 360-degree assessment method targeting specific, observable behaviors may be useful in evaluating team-based interventions and enhancing teamwork effectiveness.

Palmer, J. K. and J. M. Loveland (2008). "The influence of group discussion on performance judgments: rating accuracy, contrast effects, and halo." *Journal of Psychology* 142(2): 117-30. The authors investigated the effect of group discussion, such as may occur formally in panel interview scenarios, assessment centers, or 360-degree feedback situations, on judgments of performance. Research on group polarization suggests that the effect of group discussion combined with raters' preexisting impressions of ratees or interviewees should result in an extremization of impressions. Thus, the authors hypothesized that group discussion would (a) make ratings less accurate, (b) polarize impressions that were already good or poor as reflected by greater contrast effects, and (c) increase positive halo. Results indicated that group discussion resulted in less accurate ratings and greater contrast effects. Additional analyses suggested that group discussion increased positive halo. The authors discuss implications for research on group or panel judgments.

Palmer, R., H. Rayner, et al. (2007). "Multisource feedback: 360-degree assessment of professional skills of clinical directors." *Health Services Management Research* 20(3): 183-188. For measuring behaviour of National Health Service (NHS) staff, 360-degree assessment is a valuable tool. The important role of a clinical director as a medical leader is increasingly recognized, and attributes of a good clinical director can be defined. Set against these attributes, a 360-degree assessment tool has been designed. The job description for clinical directors has been used to develop a questionnaire sent to senior hospital staff. The views of staff within the hospital are similar irrespective of gender, post held or length of time in post. Analysis has shown that three independent factors can be distilled, namely operational management, interpersonal skills and creative/strategic thinking. A simple validated questionnaire has been developed and successfully introduced for the 360-degree assessment of clinical directors.

Papinczak, T., L. Young, et al. (2007). "Peer assessment in problem-based learning: a qualitative study." *Advances in Health Sciences Education* 12(2): 169-86. Peer assessment provides a powerful avenue for students to receive feedback on their learning. Although student perceptions of peer assessment have been studied extensively in higher education, little qualitative research has been undertaken with medical students in problem-based learning (PBL) curricula. A qualitative study of students' attitudes to, and perceptions of, peer assessment was undertaken within the framework of a larger study of metacognition with first-year medical students at the University of Queensland. A highly structured format for provision of feedback was utilised in the study design. Many recommendations from the higher education literature on optimal implementation of peer-assessment procedures were put into practice. Results indicated the existence of six main themes: (1) increased responsibility for others, (2) improved learning, (3) lack of relevancy, (4) challenges, (5) discomfort, and (6) effects on the PBL process. Five of these themes have previously been described in the literature. However, the final theme represents a unique, although not unexpected, finding. Students expressed serious concerns about the negative impact of peer assessment on the cooperative, non-judgmental atmosphere of PBL tutorial groups. The practical implications of these findings are considered.

Papinczak, T., L. Young, et al. (2007). "An analysis of peer, self, and tutor assessment in problem-based learning tutorials." *Medical Teacher* 29(5): e122-e132. Objective: The purpose of this study was to explore self-, peer-, and tutor assessment of performance in tutorials among first year medical students in a problem-based learning curriculum. Methods: One hundred and twenty-five students enrolled in the first year of the Bachelor of Medicine and Bachelor of



Surgery Program at the University of Queensland were recruited to participate in a study of metacognition and peer- and self-assessment. Both quantitative and qualitative data were collected from the assessment of PBL performance within the tutorial setting, which included elements such as responsibility and respect, communication, and critical analysis through presentation of a case summary. Self-, peer-, and tutor assessment took place concurrently. Results: Scores obtained from tutor assessment correlated poorly with self-assessment ratings ( $r = 0.31-0.41$ ), with students consistently under-marking their own performance to a substantial degree. Students with greater self-efficacy, scored their PBL performance more highly. Peer-assessment was a slightly more accurate measure, with peer-averaged scores correlating moderately with tutor ratings initially ( $r = 0.40$ ) and improving over time ( $r = 0.60$ ). Students consistently over-marked their peers, particularly those with sceptical attitudes to the peer-assessment process. Peer over-marking led to less divergence from the tutor scoring than under-marking of one's own work. Conclusion: According to the results of this study, first-year medical students in a problem-based learning curriculum were better able to accurately judge the performance of their peers compared to their own performance. This study has shown that self-assessment of process is not an accurate measure, in line with the majority of research in this domain. Nevertheless, it has an important role to play in supporting the development of skills in reflection and self-awareness.

Peiperl, M. A. (2001). "Getting 360 degrees feedback right." *Harvard Business Review* 79(1): 142-7. Over the past decade, 360-degree feedback has revolutionized performance management. But one of its components--peer appraisal--consistently stymies executives and can exacerbate bureaucracy, heighten political tensions, and consume lots of time. For ten years, Maury Peiperl has studied 360-degree feedback and has asked: under what circumstances does peer appraisal improve performance? Why does peer appraisal sometimes work well and sometimes fail? And how can executives make these programs less anxiety provoking for participants and more productive for organizations? Peiperl discusses four paradoxes inherent to peer appraisal: In the Paradox of Roles, colleagues juggle being both peer and judge. The Paradox of Group Performance navigates between assessing individual feedback and the reality that much of today's work is done by groups. The Measurement Paradox arises because simple, straightforward rating systems would seem to generate the most useful appraisals--but they don't. Customized, qualitative feedback, though more difficult and time consuming to generate, is more helpful in improving performance. During evaluations, most people focus almost exclusively on reward outcomes and ignore the constructive feedback generated by peer appraisal. Ironically, it is precisely this overlooked feedback that helps improve performance--thus, the Paradox of Rewards. These paradoxes do not have neat solutions, but managers who understand them can better use peer appraisal to improve their organizations.

Potter, T. B. and R. G. Palmer (2003). "360-degree assessment in a multidisciplinary team setting." *Rheumatology* 42(11): 1404-7.

OBJECTIVES: To use the 360-degree assessment in the multidisciplinary setting of a rheumatology department and to evaluate its impact, recognizing that this process will become part of the revalidation process of NHS professionals in the future. METHODS: Seventeen team members completed an anonymous questionnaire to give confidential opinions about the clinical, humanistic and other skills of their colleagues. Results and comments were collated and given as feedback to each individual. Before feedback, participants were asked to predict their perceived strengths and weaknesses. After feedback they evaluated the process. RESULTS: A profile of abilities was established for each team member and discussed privately with the clinical director. Often team members had good insight into their perceived strengths and weaknesses. Some participants were hurt by negative comments made about them even if this was balanced by positive comments. There were mixed views on the relevance and usefulness of the process, and whether or not it should be repeated. Some team members found the process threatening. CONCLUSION: The 360-degree assessment can be used in a multidisciplinary setting, the questions being the same for all individuals. It is a very powerful tool that must be handled carefully so that it does not cause more harm than good.

Qureshi, A. I. (2007). "A new scheme for grading the quality of scientific reports that evaluate imaging modalities for cerebrovascular diseases." *Medical Science Monitor* 13(10): RA181-7. Imaging of head and neck vasculature continues to improve with the application of new technology. To judge the value of new technologies reported in the literature, it is imperative to develop objective standards optimized against bias and favoring statistical power and clinical relevance. A review of the existing literature identified the following items as lending scientific value to a report on imaging

technology: prospective design, comparison with an accepted modality, unbiased patient selection, standardized image acquisition, blinded interpretation, and measurement of reliability. These were incorporated into a new grading scheme. Two physicians tested the new scheme and an established scheme to grade reports published in the medical literature. Inter-observer reliability for both methods was calculated using the kappa coefficient. A total of 22 reports evaluating imaging modalities for cervical internal carotid artery stenosis were identified from a literature search and graded by both schemes. Agreement between the two physicians in grading the level of scientific evidence using the new scheme was excellent (kappa coefficient: 0.93,  $p < 0.0001$ ). Agreement using the established scheme was less rigorous (kappa coefficient: 0.39,  $p < 0.0001$ ). The weighted kappa coefficients were 0.95 and 0.38 for the new and established schemes, respectively. Overall agreement was higher for the newer scheme (95% versus 64%). The new grading scheme can be used reliably to categorize the strength of scientific knowledge provided by individual studies of vascular imaging. The new method could assist clinicians and researchers in determining appropriate clinical applications of newly reported technical advances.

Ramsey, P., J. Carline, et al. (1996). "Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians." *Academic Medicine* 71: 364-70.

Rashid, A., P. Musunuri, et al. (2008). "Section 5(2) MHA 1983: results of a three-year rapid cycle clinical audit." *Medicine, Science & the Law* 48(1): 69-71.

The aim of this paper was to establish whether section 5(2) of the Mental Health Act 1983 was being used appropriately in an inpatient psychiatric unit in the UK. A clinical audit was conducted over three consecutive years. Peer review of decisions to use section 5(2) on the hospital's adult and old age wards was conducted by junior medical staff. Ninety-eight per cent of the uses of section 5(2) were felt to be appropriate and most were reviewed within 72 hours. Very few patients (5%) were placed on this section again during their admission. The number placed on section 5(4) prior to section 5(2) increased from 2% to 12%. The proportion reviewed within the first 24 hours of detention nearly doubled over the audit period and fewer patients subsequently went on to sections 2 or 3 by the third year. There was no evidence of inappropriate use of section 5(2). Changes in the shift system for junior medical staff may have inadvertently led to a more frequent use of section 5(4). The use of further formal detention after section 5(2) declined, although this decline was not statistically significant.

Redwood, S., H. Lloyd, et al. (2007). "Evaluating nurse consultants' work through key informant perceptions." *Nursing Standard* 21(17): 35-40.

AIM: To evaluate the work of nurse consultants in the NHS by exploring the views of key informants and nurse consultants. METHOD: A multi-site evaluative study commissioned by and undertaken in four trusts. The evaluation was based on the 360 degree feedback process and used case study methodology, inviting key informants to provide information on their work with nurse consultants. FINDINGS: The findings are discussed in relation to the following themes: role aspirations and lived reality; challenging boundaries; impact and outcomes and leadership. The findings concur with previous studies demonstrating a series of common themes associated with leadership, clinical expertise, research and educational activity. These findings express the ways in which nurse consultants are working to develop unique services to meet patient needs. CONCLUSION: The nurse consultant has an important role in the modernisation of the NHS. The role's impact, in terms of the informants, is in leadership, clinical expertise, research and educational activity. The findings reveal an urgent need to support consultant nurses in developing their leadership potential and skills in researching practice.

Reed, D. A., C. P. West, et al. (2008). "Behaviors of highly professional resident physicians." *JAMA - Journal of the American Medical Association* 300(11): 1326-1333.

Context: Unprofessional behaviors in medical school predict high stakes consequences for practicing physicians, yet little is known about specific behaviors associated with professionalism during residency. Objective: To identify behaviors that distinguish highly professional residents from their peers. Design, Setting, and Participants: Comparative study of 148 first-year internal medicine residents at Mayo Clinic from July 1, 2004, through June 30, 2007. Main Outcome Measures: Professionalism as determined by multiple observation-based assessments by peers, senior residents, faculty, medical students, and nonphysician professionals over 1 year. Highly professional residents were defined as those who received a total professionalism score at the 80th percentile or higher of observation-based assessments on a 5-point scale (1, needs improvement; 5, exceptional).

They were compared with residents who received professionalism scores below the 80th percentile according to In-Training Examination (ITE) scores, Mini-Clinical Evaluation Exercise (mini-CEX) scores, conscientious behaviors (percentage of completed evaluations and conference attendance), and receipt of a warning or probation from the residency program. Results: The median total professionalism score among highly professional residents was 4.39 (interquartile range [IQR], 4.32-4.44) vs 4.07 (IQR, 3.91-4.17) among comparison residents. Highly professional residents achieved higher median scores on the ITE (65.5; IQR, 60.5-73.0 vs 63.0; IQR, 59.0-67.0;  $P=.03$ ) and on the mini-CEX (3.95; IQR, 3.63-4.20 vs 3.69; IQR, 3.36-3.90;  $P=.002$ ), and they completed a greater percentage of required evaluations (95.6%; IQR, 88.1%-99.0% vs 86.1%; IQR, 70.6%-95.0%;  $P<.001$ ) compared with residents with lower professionalism scores. In multivariate analysis, a professionalism score in the top 20% of residents was independently associated with ITE scores (odds ratio [OR] per 1-point increase, 1.07; 95% confidence interval [CI], 1.01-1.14;  $P=.046$ ), mini-CEX scores (OR, 4.64; 95% CI, 1.23-17.48;  $P=.02$ ), and completion of evaluations (OR, 1.07; 95% CI, 1.01-1.13;  $P=.02$ ). Six of the 8 residents who received a warning or probation had total professionalism scores in the bottom 20% of residents. Conclusion: Observation-based assessments of professionalism were associated with residents' knowledge, clinical skills, and conscientious behaviors.

Reid, W. A., E. Duvall, et al. (2007). "Relationship between assessment results and approaches to learning and studying in Year Two medical students." *Medical Education* 41(8): 754-762.

Context: Students can take different approaches to learning and studying: deep (understanding material); surface (memorising details), and strategic (motivated by assessments). It is important to know how assessments affect student choices of approach. Methods: Students' learning approaches in Year 2 of the medical programme were measured using the Approaches to Study Skills Inventory for Students. The course was specifically designed to promote a deep approach and deter a surface approach, with explicit learning objectives and assessment constructively aligned according to Biggs' Structure of the Observed Learning Outcome (SOLO) taxonomy. The marks of individual students in different components of the assessment were compared with their scores for the 3 learning approaches. Results: Marks correlated positively with deep and strategic approaches and negatively with surface approach across a range of assessment methods (relatively well with modified essays and multiple-choice questions, but poorly with in-course assignments). Peer assessment correlated particularly strongly with strategic approach. Discussion: The correlations met our expectations in terms of direction, but were weaker and less consistent than anticipated. Possible reasons include the drive to test basic (core) material, the use of questions that may limit students' scope of expression and markers' ability to detect a deep approach. It is, however, important to refine medical programmes, particularly assessments, so that they concur with and do not adversely affect students' learning approaches. copyright 2007 Blackwell Publishing Ltd.

Reyes-Morales, H., S. Flores-Hernandez, et al. (2009). "A multifaceted education intervention for improving family physicians' case management." *Family Medicine* 41(4): 277-84.

BACKGROUND AND OBJECTIVES: The Mexican Institute of Social Security (IMSS) is the largest public health care system in Mexico. IMSS family physicians' management of clinical problems is frequently not consistent with published evidence. This study aimed to evaluate the effects of a multifaceted educational intervention to improve management of acute respiratory infections (ARI) by IMSS family physicians. METHODS: A non-randomized pre-post intervention with comparison group design was conducted in eight IMSS family medicine clinics in which 106 family physicians practiced. An evidence-based clinical guideline for ARI management was developed, and clinical tutors were trained. The three-stage intervention comprised interactive workshops, individual tutorials, and round-table peer-review sessions. The main outcome was appropriate ARI case management. The intervention effect was calculated by using the differences-in-differences model, adjusting for cluster of physicians. RESULTS: In the intervention group, the difference in mean proportion of improvement compared with baseline evaluation was 22.6% (95% confidence interval [CI]=10.3 to 34.9) for appropriate prescription of antibiotics, 29.8% (95% CI=17.2 to 42.4) for indication of worsening signs, and 19.6% (95% CI=11.2 to 28.0) for overall appropriate case management. The comparison group showed no significant changes. CONCLUSIONS: The educational intervention improved ARI management. Further studies are needed to analyze organizational implications, cost, sustainability, and effects on health outcomes.

Rice, H. E., D. P. Frush, et al. (2007). "Peer assessment of pediatric surgeons for potential risks of radiation exposure from computed tomography scans." *Journal of Pediatric Surgery* 42(7): 1157-64.

**PURPOSE:** Radiology literature reports potential cancer risk from radiation exposure from computed tomography (CT). We hypothesized that pediatric surgeons' knowledge of potential risks of radiation exposure from CT scan is limited. **METHODS:** We used an anonymous, Internet-based peer assessment survey for members of the American Pediatric Surgical Association (APSA). The survey assessed surgeon's knowledge based on potential risks of radiation exposure from CT as well as current practice patterns for use of CT. The chi2 test of significance was used to detect any differences in responses based on years in training. **RESULTS:** Twenty percent (147/753) of the American Pediatric Surgical Association members completed the survey. About one half (54%) of surgeons believe that the lifetime risk of cancer was increased because of radiation from one abdominal/pelvic CT scan, although more than 75% of respondents underestimated the radiation dose from a CT scan compared to a chest radiograph. Most surgeons generally did not discuss the potential risks of CT scan with their patients. Surgeons demonstrated a range of responses for use of CT for select clinical scenarios. **CONCLUSIONS:** Pediatric surgeon's knowledge of potential risks of radiation exposure from CT scan is limited. As the radiology literature indicates an increasing awareness for potential cancer risks from radiation exposure from CT, there is also a need for education of subspecialties outside of radiology.

Roach, W. H., Jr. (1985). "Access to peer review committee records and incident reports, and photocopying fees." *Topics in Health Record Management* 6(1): 80-4.

Rodgers, K. G. and C. Manifold (2002). "360-Degree feedback: Possibilities for assessment of the ACGME core competencies for emergency medicine residents." *Academic Emergency Medicine* 9(11): 1300-1304.

The Accreditation Council for Graduate Medical Education (ACGME) has challenged residency programs to provide documentation via outcomes assessment that all residents have successfully mastered the six core competencies. A variety of assessment "tools" has been identified by the ACGME for outcomes assessment determination. Although rarely cited in the medical literature, 360-degree feedback is currently in widespread use in the business sector. This tool provides timely, consolidated feedback from sources in the resident's sphere of influence (emergency medicine faculty, emergency medicine residents, off-service residents and faculty, nurses, ancillary personnel, patients, out-of-hospital care providers, and a self-assessment). This is a significant deviation from both the peer review process and the resident review process that almost exclusively use physicians as raters. Because of its relative lack of development, utilization, and validation as a method of resident assessment in graduate medical education, a great opportunity exists to develop the 360-degree feedback tool for resident assessment.

Rosen, L. S., K. Schroeder, et al. (1996). "Adapting a statewide patient database for comparative analysis and quality improvement." *Joint Commission Journal on Quality Improvement* 22(7): 468-81. **BACKGROUND:** In 1991 the Michigan Health and Hospital Association established the Michigan Patient Outcome Measures Program to support quality improvement activities among member hospitals. Data submitted by each hospital are based on the Uniform Hospital Discharge Data Set (UHDDS). **REPORTING THE MEASURES:** Each participating hospital is provided with tabulations for each of the 18 indicators, along with comparative data for each of five hospital peer groups. **CASE STUDY 1: FACILITATING DEVELOPMENT OF A CLINICAL PATHWAY:** Review of data on stroke patients indicated that patients discharged to home represented the single largest group of patients and that their length of stay was lower than for other groups of stroke patients. Hospital A used a clinical pathway to achieve reduction in length of stay of almost a day for this group. **CASE STUDY 2: IMPROVING OBSTETRICS AND GYNECOLOGIC SERVICES THROUGH EDUCATION:** The Department of Obstetrics and Gynecological Surgery at Hospital B achieved improvement in birth trauma, hysterectomy, cesarean section (C-section) births, and vaginal births after C-section. Program participants are considering adopting a birth trauma indicator. **CASE STUDY 3: ESTABLISHING A HOSPITALWIDE QI PROGRAM:** The Newborn Team at Hospital C initiated several QI initiatives aimed at reducing the newborn mortality rate, including guidelines for earlier identification of high-risk obstetric cases and for earlier intervention by the perinatologist. **CONCLUSIONS:** Data that are not risk or severity adjusted have value in assessing hospital procedures and systems and can be used to educate and effect change in practice patterns.

Rosenfeld, R. M. (2007). "Peer review." *Otolaryngology - Head and Neck Surgery* 137(2): 179-181.

Roth, A. D. and S. Pilling (2007). "Clinical practice and the CBT competence framework: An update for clinical and counselling psychologists." *Clinical Psychology Forum*(179): 53-55.

Roth, R. R., P. J. Porter, et al. (1993). "The attitudes of family physicians toward the peer review process." *Archives of Family Medicine* 2(12): 1271-5.

We conducted a study to determine family physicians' attitudes and perceptions toward current peer review practices, and to discover if family physicians, general surgeons, and hospital-based physicians view the process differently. A survey instrument measured perceptions of physicians on the following four areas of the peer review process: (1) how peer reviews are administered, (2) the educational value of peer reviews, (3) the performance of peer review committees, and (4) the effect of the peer review process on physician morale. The survey was mailed to all 3528 practicing physicians who were members of a state medical society. A subgroup of 1695 family physicians, general surgeons, and hospital-based physicians was used for this study, of whom 774 (46%) responded to the questionnaire. Over one half of the family physicians responded negatively toward the peer review process on all items of the survey, with over 70% dissatisfied on five of the 17 items. Family physicians, general surgeons, and hospital-based physicians viewed the peer review process differently in the four areas measured. We found statistically significant differences of opinions regarding present peer review practices among the specialties cited. However, the overall dissatisfaction of the specialty groups studied may suggest that the concern resides more with the profession at large than with any one medical specialty group.

Rowland, J. M. (1988). "1987 Letourneau award. Enforcing hospital responsibility through self-evaluation and review committee confidentiality." *Journal of Legal Medicine* 9(3): 377-419.

Russo, M. R. (2005). "Scientific publishing: 360 degrees of satisfaction." *JALA - Journal of the Association for Laboratory Automation* 10(1): 6A.

Russo, R. M., V. J. Gururaj, et al. (1975). "A chart audit peer review system in an ambulatory service." *Pediatrics* 56(2): 246-54.

A system of providing rapid review of ambulatory patient records using a set of criteria was developed and tested. Of 337 records selected for review, 25% were not reviewable. Of the remaining 75%, virtually one half proved to be lacking in one or more of the established criteria. Deficiencies tended to occur most often in criteria related to the physical examination of the child. Physician compliance with the criteria could not be conclusively shown to improve during the five-month study period, although physician acceptability was judged as good. A large percentage (47.3%) of deficient records could be corrected by the responsible health professional. The chief advantages of the system are that it is rapid and easy to perform. It focuses on the major service provided by the facility, i.e., acute medical care, and on the performance of the individual providing that service.

Saberton, S. and C. Violato (2007). "Implementation of a multi-source feedback system for quality assurance for the College of Medical Radiation Technologists of Ontario." *Canadian Journal of Medical Radiation Technology* 38(1): 20-25.

Under the Health Professions procedural code of the Regulated Health Professions Act, 1991 (RHPA), the College of Medical Radiation Technologists of Ontario (CMRTO) is required to provide a means to assess how people in the profession actually perform in practice. As the first step in satisfying the practice assessment requirement of the Quality Assurance program in 2004, the CMRTO retained Edumetrics Ltd. to undertake the development of a multi-source feedback (MSF) program specific to the profession of medical radiation technology.

In a comprehensive study led by Dr. Claudio Violato, a professor in the Department of Community Health Sciences, Faculty of Medicine at the University of Calgary and an expert in the development of program evaluations, psychometrics, research methods and statistics, the CMRTO developed a 'multi-source' assessment process that provides a means for peers, colleagues, patients and the assessed person to complete a survey focused on the standards of practice. The assessed person receives a summary assessment report or feedback about his or her performance. The multi-source feedback assessment provides a formative evaluation, providing continuous reaction to the performance of a medical radiation technologist (MRT). It compares the MRT's clinical performance to that of other MRTs. The CMRTO multi-source practice assessment approach was piloted in 2005.

A number of strategies were used to recruit 327 MRTs to take part in the pilot study. Recruitment strategies included consultation with MRTs respecting the standards of practice as the basis of the MSF project, a detailed communication strategy and telephone recruitment. Although the CMRTO

staff experienced challenges in recruiting MRTs to take part in the pilot study and dealing with the resultant feedback, the pilot study was considered a success.

Sargeant, J., K. Mann, et al. (2008). "Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use." *Advances in Health Sciences Education* 13(3): 275-88.  
INTRODUCTION: Receiving negative performance feedback can elicit negative emotional reactions which can interfere with feedback acceptance and use. This study investigated emotional responses of family physicians' participating in a multi-source feedback (MSF) program, sources of these emotions, and their influence upon feedback acceptance and use. METHODS: The authors interviewed 28 volunteer family physician participants in a pilot study of MSF, purposefully recruited to represent the range of scores. The study was conducted in 2003-2004 at Dalhousie University. RESULTS: Participants' emotional reactions to feedback appeared to be elicited in response to an internal comparison of their feedback with self-perceptions of performance. Those agreeing with their feedback; i.e., perceiving it as generally consistent with or higher than self-perceptions responded positively, while those disagreeing with their feedback; i.e., seeing it as generally inconsistent with or lower than self-perceptions, generally responded with distress. For the latter group, these feelings were often strong and long-lasting. Some eventually accepted their feedback and used it for change following a long period of reflection. Others did not and described an equally long reflective period but one which focused on and questioned MSF procedures rather than addressed feedback use. Participants suggested providing facilitated reflection on feedback to enhance assimilation of troubling emotions and interpretation and use of feedback. CONCLUSIONS: Negative feedback can evoke negative feelings and interfere with its acceptance. To overcome this, helpful interventions may include raising awareness of the influence of emotions, assisting recipients to focus their feedback on performance tasks, and providing facilitated reflection on feedback.

Sargeant, J., K. Mann, et al. (2008). "'Directed' self-assessment: practice and feedback within a social context." *Journal of Continuing Education in the Health Professions* 28(1): 47-54.  
INTRODUCTION: Accurate self-assessment appears to be difficult and, some would propose, even impossible. Recent reviews suggest that peer assessment may be more accurate and that multisource feedback (MSF) may inform self-assessment. We had conducted a series of studies of family physicians in an MSF program including assessments from patients, medical colleagues, and coworkers and self-assessment. Using this body of research, this article explores self-assessment within the social context of multisource feedback and investigates the influence of feedback from peers and others upon self-assessment. METHODS: This is a review article in which we synthesized findings of the series of studies with respect to self-assessment, used conclusions to propose a model for self-assessment within a social context, and suggest practical and research implications. RESULTS: Physicians compared peers' and others' assessment feedback with global self-perceptions of performance. Negative feedback, especially from medical colleagues, that was inconsistent with self-perceptions was not readily reconciled with self-assessments. Multiple internal and environmental factors influenced reconciliation and assimilation of negative feedback. Reflection upon feedback and self-perceptions appeared to be instrumental to reconciliation, and reflection could be facilitated. DISCUSSION: We propose a model of "directed" self-assessment to facilitate the integration of external feedback, especially negative feedback, with self-perceptions and enable its use for practice improvement. Implications for education and research include increasing understanding of ways physicians assimilate external feedback and of the role of educators as facilitators of "directed" self-assessment and self-learning to assist physicians in integrating external feedback.

Sargeant, J. M., K. V. Mann, et al. (2003). "Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: A pilot study." *Academic Medicine* 78(10 SUPPL.): S42-S44.

Purpose. To describe responses of family physicians, their medical colleagues, and coworker raters to a multisource feedback assessment process. Method. Data collection tools included multisource feedback self-assessment and medical colleague, coworker, and patient rating forms; and program evaluation physician and rater questionnaires. Results. The pilot study included 142 physicians and their raters, with 113 (80%) physicians completing evaluations. Positive correlations were found between familiarity scores and medical colleague and coworker mean ratings. Peer medical colleagues were significantly more familiar with physicians than were consultants. Consultants were unable to rate items most frequently. Physicians disagreed with colleague feedback more frequently. Agreement was positively correlated with scores. Conclusions. Familiarity, ability to observe

physicians appropriately to rate them, and physicians' responses to feedback are factors to consider when multisource feedback is used.

Saucier, S. D. (2004). "360-degree feedback systems: development and effectiveness." *Magma Connexion/Medical Group Management Association* 4(9): 24-5.

Schmalenberg, C., M. Kramer, et al. (2008). "Clinically competent peers and support for education: structures and practices that work... article 3 in a series of 8." *Critical Care Nurse* 28(4): 54.  
Structures that foster clinical competency include annual reviews, educational support, recognition, review sessions, and best-practice teams.

Schonrock-Adema, J., M. Heijne-Penninga, et al. (2007). "Assessment of professional behaviour in undergraduate medical education: Peer assessment enhances performance." *Medical Education* 41(9): 836-842.

Objectives: To examine whether peer assessment can enhance scores on professional behaviour, with the expectation that students who assess peers score more highly on professional behaviour than students who do not assess peers. Methods: Undergraduate medical students in their first and second trimesters were randomly assigned to conditions with or without peer assessment. Of the total group of 336 students, 278 students participated in the first trimester, distributed over 31 tutorial groups, 17 of which assessed peers. The second trimester involved 272 students distributed over 32 groups, 15 of which assessed peers. Professional behaviour was rated by tutors on 3 dimensions: Task Performance; Aspects of Communication, and Personal Performance. The rating scale ranged from 1 (poor) to 10 (excellent). Data were analysed using multivariate repeated measures multilevel analysis. Results: Assessment scores were found to have generally increased in the second trimester, especially the personal performance scores of students who assessed peers. In addition, female students were found to have significantly higher scores than male students. Conclusions: In undergraduate medical education, peer assessment has a positive influence on professional behaviour. However, the results imply that peer assessment is only effective after students have become adjusted to the complex learning environment.

Schulpen, T. W. J. and K. M. J. Lombarts (2007). "Quality improvement of paediatric care in the Netherlands." *Archives of Disease in Childhood* 92(7): 633-636.

The development of the quality improvement programme of the Paediatric Association of the Netherlands is described within the setting of the national programme of the Dutch government. The programme is based on four pillars: site visits by peers (*visitatie*), continuous medical and professional education, development of clinical (evidence based) guidelines and patient safety with complication registration. The site visits by peers play a central role in assessing the quality improvement activities in hospital based paediatric care. The self assessment approach and the confidential character of the visits are well received by the surveyed specialists. Recent inclusion of quality criteria in the legally required 5 yearly medical specialist recertification process has boosted the care for quality, which could serve as example for other countries.

Sen Gupta, T. K., R. B. Muray, et al. (2008). "Rural internships for final year students: clinical experience, education and workforce." *Rural & Remote Health* 8(1): 827.

INTRODUCTION: The James Cook University School of Medicine is the only complete medical school in northern Australia, and it has a mission to prepare graduates to meet the unique needs of the region with a particular emphasis on rural, remote, Indigenous and tropical health. Eight-week 'rural internships' have been undertaken by all sixth-year medical students at James Cook University since 2005. Each student had previously completed at least 12 weeks of structured rural placements in years 2 and 4, as well as other core teaching in rural health including the year 2 subject, 'Rural, Remote, Indigenous and Tropical Health'. Students worked in rural hospitals across northern Australia developing and practising clinical skills under the supervision of senior staff. Students undertook full-time inpatient and outpatient responsibilities under supervision, being rostered for after-hours work with appropriate support. Assessment involved a learning portfolio, including multi-source feedback from peers, supervisors and patients, and a population health project and a telephone referral exercise. METHODS: This article describes the development, implementation and assessment of the first years of the program, from 2005 to 2007. Evaluation included student questionnaires, site visits and interviews, and follow-up teleconferences with preceptors. RESULTS: The rural internship provides senior medical students with valuable experience by active participation in the healthcare team. Students reported a rich and varied clinical experience. Students accept limited supervised

responsibility and further their ability and confidence to undertake the role of the intern. Importantly, they proved not to be a burden to the system. This rotation therefore appears to meet educational needs without compromising the local workforce (and indeed may add to it). Students felt welcomed by their communities and enjoyed the social and cultural aspects of their attachment, as well as the clinical aspects and the opportunity to further their understanding of rural communities, rural health care and the healthcare team. Preparation of the students, the preceptors and the communities emerged as a key element of success. CONCLUSION: This model extends and enhances the traditional apprenticeship model by its rural focus and distributed nature, and involvement of the entire student cohort. In addition, the contribution to patient care by senior students and junior doctors enables a consultant-registrar-resident model, in which experienced rural doctors function as consultants providing advice, support and tuition rather than predominantly face-to-face patient care. This approach also provides a means to address an emerging paradox: rural preceptors and communities want to teach students, appreciating the long-term workforce implications, but are increasingly constrained by resources, particularly time. Similar innovative approaches should be explored in other settings.

Shumway, J. M., R. M. Harden, et al. (2003). "AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician." *Medical Teacher* 25(6): 569-84.

Two important features of contemporary medical education are recognized. The first is an emphasis on assessment as a tool to ensure quality in training programmes, to motivate students and to direct what they learn. The second is a move to outcome-based education where the learning outcomes are defined and decisions about the curriculum are based on these. These two trends are closely related. If teachers are to do a better job of assessing their students, they need an understanding of the assessment process, an appreciation of the learning outcomes to be assessed and a recognition of the most appropriate tools to assess each outcome. Assessment tools selected should be valid, reliable, practical and have an appropriate impact on student learning. The preferred assessment tool will vary with the outcome to be assessed. It is likely to be some form of written test, a performance test such as an OSCE in which the student's competence can be tested in a simulated situation, and a test of the student's behaviour over time in clinical practice, based on tutors' reports and students' portfolios. An assessment profile can be produced for each student which highlights the learning outcomes the student has achieved at the required standard and other outcomes where this is not the case. For educational as well as economic reasons, there should be collaboration across the continuum of education in test development as it relates to the assessment of learning outcomes and in the implementation of a competence-based approach to assessment.

Siegal, E. C., E. J. Angelakis, et al. (2008). "Can peer review contribute to earlier detection of breast cancer? A quality initiative to learn from false-negative mammograms." *Breast Journal* 14(4): 330-334. Although Mammography Quality Standards Act requires tracking true positives, tracking false negatives is not required. We describe a peer review process implemented at Lahey Clinic to identify false-negative mammograms. We defined a false-negative mammogram as one which was read as negative within 12 months of a cancer diagnosis, and in which two of three radiologists correctly identified the site of cancer. Reviewing radiologists were blinded to each other and to computer-aided design (CAD), but were aware that somewhere in the mammogram was cancer. 25/64, 983, or 0.038% were classified as misses. The false-negative rate of any one radiologist averaged <0.1% without outliers. Of the false negatives, 60% were in heterogeneously dense tissue, 72% were asymmetries or masses rather than calcifications, and 24% were correctly identified by CAD in two views. We use these data for quality assurance, educational purposes, and to help identify patterns of undetected cancers to aid in earlier and improved detection of breast cancers.

Sood, R., A. Sood, et al. (2007). "Non-evidence-based variables affecting physicians' test-ordering tendencies: A systematic review." *Netherlands Journal of Medicine* 65(5): 167-177.

Background: The concept of evidence-based medicine (EBM) was introduced in 1992. Incorporation of EBM into physicians' practices, however, has been slow. Test-ordering tendencies are still based on variables that are not necessarily evidence-based. Methods: The literature was reviewed to identify the non-EBM variables that affect physicians' practices of test ordering. Studies of interest were limited to original research on the determinants of physicians' test-ordering tendencies. The search strategy included queries in MEDLINE (1992-2006), Web of Science (1993-2006), EMBASE (1992-2006), and PsycINFO (1992-2006); checking of reference lists; hand searching relevant journals; and personal communication with experts. Two independent reviewers abstracted information on the



design, quality, and limitations of the study. Review articles, letters, and editorials were excluded from analysis. Results: 104 original studies reporting on the variables affecting test ordering were identified. Of these, 53 studies assessing physician variables affecting test ordering were identified. Some of the recognisable physician factors included age, sex, degree of specialisation, geographic location and practice setting, individual belief systems, experience, knowledge, fear of malpractice litigation, physician regret, financial incentives, awareness of costs of tests ordered, and provision of written feedback by peers or employers. Conclusion: Despite considerable advances in our understanding of EBM and its application to patient care, several non-EBM physician variables influence physicians' test-ordering characteristics. Ongoing effort is needed to identify the modifiable non-EBM determinants of physicians' test ordering and to use appropriate tools and techniques to encourage evidence-based behaviours for test ordering.

Southgate, L., M. Campbell, et al. (2001). "The General Medical Council's Performance Procedures: The development and implementation of tests of competence with examples from general practice." *Medical Education*, Supplement 35(1): 20-28.

Objective: This paper describes the development of the tests of competence used as part of the General Medical Council's assessment of potentially seriously deficient doctors. It is illustrated by reference to tests of knowledge and clinical and practical skills created for general practice. Subjects and tests: A notional sample of 30 volunteers in 'good standing' in the specialty (reference group), 27 practitioners referred to the procedures and four practitioners not referred but who were the focus of concern over their performance. Tests were constructed using available guidelines and a specially convened working group in the specialty. Methods: Standards were set using Angoff, modified contrasting group and global judgement methods, as appropriate. Results: Tests performed highly reliably, showed evidence of construct validity, intercorrelated at appropriate levels and, at the standards employed, demonstrated good separation of reference and referred groups. Likelihood ratios for above and below standard performance based on competence were large for each test. Seven of 27 doctors referred were shown not to be deficient in both phases of the performance assessment.

Southgate, L., J. Cox, et al. (2001). "The General Medical Council's Performance Procedures: Peer review of performance in the workplace." *Medical Education*, Supplement 35(1): 9-19.

The General Medical Council procedures to assess the performance of doctors who may be seriously deficient include peer review of the doctor's practice at the workplace and tests of competence and skills. Peer reviews are conducted by three trained assessors, two from the same speciality as the doctor being assessed, with one lay assessor. The doctor completes a portfolio to describe his/her training, experience, the circumstances of practice and self rate his/her competence and familiarity in dealing with the common problems of his/her own discipline. The assessment includes a review of the doctor's medical records; discussion of cases selected from these records; observation of consultations for clinicians, or of relevant activities in non-clinicians; a tour of the doctor's workplace; interviews with at least 12 third parties (five nominated by the doctor); and structured interviews with the doctor. The content and structure of the peer review are designed to assess the doctor against the standards defined in Good Medical Practice, as applied to the doctor's speciality. The assessment methods are based on validated instruments and gather 700-1000 judgements on each doctor. Early experience of the peer review visits has confirmed their feasibility and effectiveness.

Spath, P. (2007). "Fortifying individual case review activities. Screening cases for peer review." *Hospital Peer Review* 32(9): 129-32.

Spath, P. (2007). "Meeting more explicit peer review imperatives." *Hospital Peer Review* 32(8): 111-4.

Spath, P. (2007). "The real deal on holding successful case reviews." *Hospital Peer Review* 32(10): 145-8.

Staccini, P., E. Hergon, et al. (2007). "A structured ePortfolio to handle a programme of professional competences assessment in the area of transfusion medicine." *Transfusion Clinique et Biologique* 14(3): 352-358.

In order to organize a nationwide program for the evaluation of professional practices in the area of blood transfusion, the French National Blood Transfusion Institute and the Nice-Sophia Antipolis University designed and implemented a web based service aimed at following-up and guiding the physicians involved in such an assessment program. The core component is a structured electronic

portfolio (ePortfolio), the implementation of which was based on an object-oriented environment combined with a content management system. The modelling of the global evaluation system makes it possible to describe this type of portfolio according to six axes: an axis "objectives" (competencies accreditation); an axis "target" (health care professionals); an axis "content" (numerical documents); an axis "structure" (matrix of answer defined in space and time); an axis "source" (single source peer-reviewed); an axis "level of evidence" (validation of the proof after its deposit by an identified and authenticated peer user, whole tracking of the exchanges and interactions between users and device). copyright 2007 Elsevier Masson SAS. All rights reserved.

Staccini, P. M. and P. Rouger (2008). "Modeling and using a web-based and tutored portfolio to support certification of professional competence in transfusion medicine." AMIA(Annual Symposium Proceedings/AMIA Symposium): 697-701.

In order to manage a nationwide assessment program leading to certification of professional competence in blood transfusion throughout France, the National Institute of Blood Transfusion (INTS) and the University of Nice-Sophia Antipolis designed and developed a structured and tutored web-based portfolio. The entire process of certification has been approved by the national healthcare agency (HAS). Eleven assessment programs have been written. The structure of this e-portfolio is based on a matrix of actions defined according to standards of practice. For each action, elements of proof are uploaded by the physician and peer-reviewed by an expert (a tutor) before validation. The electronic portfolio stores all the history of the actions performed by users. This tracking feature generates alerts which are e-mailed to users (physicians and tutors) according to a list of monitored events. After one year of design and development, the application is now being used routinely.

Stark, R., D. Korenstein, et al. (2008). "Impact of a 360-degree professionalism assessment on faculty comfort and skills in feedback delivery." *Journal of General Internal Medicine* 23(7): 969-72.

**BACKGROUND:** Professionalism is identified as a competency of resident education. Best approaches to teaching and evaluating professionalism are unknown, but feedback about professionalism is necessary to change practice and behavior. Faculty discomfort with professionalism may limit their delivery of feedback to residents. **OBJECTIVES:** A pilot program to implement a 360-degree evaluation of observable professionalism behaviors and determine how its use impacts faculty feedback to residents. **DESIGN:** Internal Medicine (IM) residents were evaluated during ambulatory rotations using a 360-degree assessment of professional behaviors developed by the National Board of Medical Examiners(R). Faculty used evaluation results to provide individual feedback to residents. **PATIENTS/PARTICIPANTS:** Fifteen faculty members. **MEASUREMENTS AND MAIN RESULTS:** Faculty completed pre- and post-intervention surveys. Using a 7-point Likert scale, faculty reported increased skill in giving general feedback (4.85 vs 4.36,  $p < .05$ ) and feedback about professionalism (4.71 vs 3.57,  $p < .01$ ) after the implementation of the 360-degree evaluation. They reported increased comfort giving feedback about professionalism (5.07 vs 4.35,  $p < .05$ ) but not about giving feedback in general (5.43 vs 5.50). **CONCLUSIONS:** A 360-degree professionalism evaluation instrument used to guide feedback to residents improves faculty comfort and self-assessed skill in giving feedback about professionalism.

Staton, L. J., S. M. Kraemer, et al. (2007). "'Correction:'Peer chart audits: A tool to meet Accreditation Council on Graduate Medical Education (ACGME) competency in practice-based learning and improvement." *Implementation Science* 2: 24.

**ABSTRACT:** **BACKGROUND:** The Accreditation Council on Graduate Medical Education (ACGME) supports chart audit as a method to track competency in Practice-Based Learning and Improvement. We examined whether peer chart audits performed by internal medicine residents were associated with improved documentation of foot care in patients with diabetes mellitus. **METHODS:** A retrospective electronic chart review was performed on 347 patients with diabetes mellitus cared for by internal medicine residents in a university-based continuity clinic from May 2003 to September 2004. Residents abstracted information pertaining to documentation of foot examinations (neurological, vascular, and skin) from the charts of patients followed by their physician peers. No formal feedback or education was provided. **RESULTS:** Significant improvement in the documentation of foot exams was observed over the course of the study. The percentage of patients receiving neurological, vascular, and skin exams increased by 20% (from 13% to 33%) ( $p = 0.001$ ), 26% (from 45% to 71%) ( $p < 0.001$ ), and 18% (51%-72%) ( $p = 0.005$ ), respectively. Similarly, the proportion of patients receiving a well-documented exam which includes all three components - neurological, vascular and skin foot exam - increased over time (6% to 24%,  $p < 0.001$ ). **CONCLUSION:** Peer chart audits performed by residents in the absence of formal feedback were associated with improved

documentation of the foot exam in patients with diabetes mellitus. Although this study suggests that peer chart audits may be an effective tool to improve practice-based learning and documentation of foot care in diabetic patients, evaluating the actual performance of clinical care was beyond the scope of this study and would be better addressed by a randomized controlled trial.

Stevenson, K., W. Mackey Jr, et al. (1979). "Child psychiatry perspectives. Problems in implementation of peer review in child psychiatry." *Journal of the American Academy of Child Psychiatry* 18(2): 391-400.

This paper describes the efforts of a committee of child psychiatrists to evaluate the problems and complexities of meeting PSRO requirements. A retrospective review of charts was done, using PSRO model criteria sets on children and adolescents on inpatient units at five hospitals. Problems were noted in the use of diagnoses, the care given, and the length of stay. Many charts would not have passed the initial nonphysician screening and would have required a physician review. It is anticipated that once PSRO criteria are introduced, a concurrent review might be less difficult. The results of this review suggest that to meet PSRO criteria, costs would have to go up.

Strike, C., E. Wenghofer, et al. (2007). "Physician peer assessments for compliance with methadone maintenance treatment guidelines." *Journal of Continuing Education in the Health Professions* 27(4): 208-13.

**INTRODUCTION:** Medical associations and licensing bodies face pressure to implement quality assurance programs, but evidence-based models are lacking. To improve the quality of methadone maintenance treatment (MMT), the College of Physicians and Surgeons of Ontario, Canada, conducts an innovative quality assurance program on the basis of peer assessments. Using data from this program, we assessed physician compliance with MMT guidelines and determined whether physician factors (e.g., training, years of practice), practice type, practice location, and/or caseload is associated with MMT guideline adherence. **METHODS:** Secondary analysis of methadone practice assessment data collected by the College of Physicians and Surgeons of Ontario, Canada. Assessment data from methadone prescribing physicians who completed their first year of methadone practice were analyzed. We calculated the mean percentage compliance per guideline per physician and global compliance across all guidelines per physician. Linear regression was used to assess factors associated with compliance. **RESULTS:** Data from 149 physician practices and 1,326 patient charts were analyzed. Compliance across all charts was greater than 90% for most areas of care. Compliance was less than 90% for take-home medication procedures; urine toxicology screening; screening for hepatitis B virus (HBV), hepatitis C virus (HCV), human immunodeficiency virus (HIV), tuberculosis, other sexually transmitted infections, and completion of a psychosocial assessment. Mean global compliance across all charts and guidelines per physician was 94.3% (standard deviation = 7.4%) with a range of 70% to 100%. Linear regression analysis revealed that only year of medical school graduation was a significant predictor of physician compliance. **DISCUSSION:** This is the first report of MMT peer assessments in Canada. Compliance is high. Few countries conduct similar assessment processes; none report physician-level results. We cannot quantify the contribution of peer assessment, training, or self-selection to the compliance rates, but compared to other areas of practice these rates suggest that peer assessment may exert a significant effect on compliance. A similar assessment process may in other areas of clinical practice improve physician compliance.

Tankisi, H., K. Pugdahl, et al. (2006). "Influence of peer review medical audit on pathophysiological interpretation of nerve conduction studies in polyneuropathies." *Clinical Neurophysiology* 117(5): 979-983.

**Objective:** To evaluate the possible influence of peer review medical audit on experienced physicians' pathophysiological interpretation of nerve conduction studies in polyneuropathy patients. **Methods:** Since 1992, 7 European neurophysiologists have collected samples of their patient examinations for regular review where the physicians interpret each other's cases electronically and subsequently discuss them at regular workshop meetings (i.e. a form of medical audit). Two sets of 100 polyneuropathy examinations interpreted with an interval of 4-6 years were selected. The sets contained 1456 and 1719 nerve conduction studies, each given a pathophysiological test conclusion by each individual physician. Inter-physician agreement on interpretation of demyelination and axonal loss of the nerve, as well as neuropathic and unspecific findings, was estimated using kappa statistics. **Results:** Increased agreement from set 1 to set 2 was found on interpretation of demyelination of the nerve (set 1: kappa=0.22; set 2: kappa=0.45), and of neuropathic (set 1: kappa=0.46; set 2: kappa=0.64) and unspecific findings (set 1: kappa=0.35; set 2: kappa=0.54). No

changes were found on interpretation of axonal loss (set 1: kappa=0.26; set 2: kappa=0.31) and normal findings (set 1 and set 2: kappa=0.90). Conclusions: Participation in regular peer review medical audit resulted in increased agreement on interpretation of nerve conduction studies for 6 of the 7 participants. The study further highlights the need for better definition of criteria for identification of demyelinating, and in particular, axonal peripheral neuropathies. Significance: International collaboration involving peer review medical audit may contribute to development of practice guidelines and, in turn, to increased quality of electrodiagnostic medicine.

Tate Jr, J. S. (2007). "Quality Improvement and Peer Review." *Journal of the American College of Surgeons* 205(1): 196.

The Lancet (2008). "The pitfalls and rewards of peer review." *The Lancet* 371(9611): 447.

Theman, T. W., H. M. Oetter, et al. (2009). "Revalidation of Canadian physicians." *Canadian Medical Association Journal* 180(5): 539.

Thompson, R. E. (1988). "Exploding 12 myths about quality assurance and peer review." *Medical Staff Counselor* 2(4): 39-49.

In order to accomplish effective, objective peer review and quality assurance in the midst of concerns about confidentiality and malpractice, 12 widely-held, false assumptions about the nature of peer review and quality assurance must be discarded.

Tracey, J. M. (1991). "Peer review. Part 1: Attitudes to internal and external standards." *New Zealand Medical Journal* 104(904): 8-10.

Two peer review groups of ten general practitioners each reviewed their management of headache and fatigue. Each group used internal standards which they had set for the review of one condition and external standards set by the other group for the other condition. Data for the review was collected from the general practitioners' case notes and presented to them. The following year their management was again reviewed. The attitudes of the general practitioners to the use of both internal and external standards in peer review were measured using written questionnaire. There was no statistically significant difference in attitude to the two different types of standards. The attitudes of the general practitioners to external standards from a variety of sources were also measured. External standards developed by other peer review groups of general practitioners were significantly more acceptable ( $p$  less than 0.05) than standards developed by hospital specialists or the Department of Health.

Tracey, J. M. (1991). "Peer review. Part 2: The influence of internal and external standards on outcome." *New Zealand Medical Journal* 104(906): 66-7.

Two peer review groups of ten general practitioners each reviewed their management of two conditions, headache and fatigue. Each group used internal standards which they had set for the review of one condition and external standards set by the other group for the other condition. Data for the review were collected from the general practitioners' case notes and presented to them. The following year their management was again reviewed. Group 1 improved their management of both conditions by 15%. Group 2 improved their management of the condition where internal standards were used by 12% and the condition where external standards were used by 9%. This difference was not statistically significant. Although there was a significant improvement in the management of both conditions in each group, the use of internal versus external standards made no difference to the degree of improvement following peer review.

Triggle, C. R. and D. J. Triggle (2007). "What is the future of peer review? Why is there fraud in science? Is plagiarism out of control? Why do scientists do bad things? Is it all a case of: "All that is necessary for the triumph of evil is that good men do nothing?"" *Vascular Health and Risk Management* 3(1): 39-53.

Peer review is an essential component of the process that is universally applied prior to the acceptance of a manuscript, grant or other scholarly work. Most of us willingly accept the responsibilities that come with being a reviewer but how comfortable are we with the process? Peer review is open to abuse but how should it be policed and can it be improved? A bad peer review process can inadvertently ruin an individual's career, but are there penalties for policing a reviewer who deliberately sabotages a manuscript or grant? Science has received an increasingly tainted name because of recent high profile cases of alleged scientific misconduct. Once considered the

results of work stress or a temporary mental health problem, scientific misconduct is increasingly being reported and proved to be a repeat offence. How should scientific misconduct be handled - is it a criminal offence and subject to national or international law? Similarly plagiarism is an ever-increasing concern whether at the level of the student or a university president. Are the existing laws tough enough? These issues, with appropriate examples, are dealt with in this review.

Trunkey, D. D. and R. Botney (2001). "Assessing competency: a tale of two professions." *Journal of the American College of Surgeons* 192(3): 385-95.

Vautrey, R. and R. Neal (2000). "Using peer review visits to carry out the clinical governance baseline assessment in one primary care group." *Journal of Clinical Governance* 8(3): 135-140.

This paper reports the process of the clinical governance baseline assessment undertaken by one primary care group. The method chosen in order to undertake this was peer review practice visits, facilitated by prior completion of a detailed questionnaire. This questionnaire covered a number of clinical and organisational areas including access and availability, chronic disease management, risk management, education and training, and included a simple audit of the management of patients with ischaemic heart disease. The visits provided a forum for discussion of these areas. Twenty two (out of 29) practices received a visit; most reported that the process was acceptable. The practices reported educational benefits of the process. If more practices are to engage in this in the future, then the issues of financial incentives, clinical workload, and a service requirement to engage in the process need to be addressed. The true success of the clinical governance agenda in primary care can only be judged in the longer term by measures of quality improvement.

Veltman, L. L. (2007). "Getting to havarti: moving toward patient safety in obstetrics." *Obstetrics & Gynecology* 110(5): 1146-50.

Most health care professionals who are involved in efforts to improve patient safety are aware of James Reason's "Swiss cheese" model of how accidents occur. Some elements and pressures of current obstetric practice may weaken defences and safeguards against perinatal injury. Several components of obstetric care in labor and delivery units can be used as targets for tightening the "holes" in the Swiss cheese model. These include improving communications, preparing for rare critical events through simulation training, developing protocols for administration of important medications used in labor and delivery (oxytocin, misoprostol, and magnesium sulfate), increasing the in-house presence of obstetricians, developing an effective departmental infrastructure that includes effective peer review, providing risk management education about high-risk clinical areas that have the potential to result in catastrophic injury, and staffing the unit for all contingencies during all hours, day and night. Acceptance by the obstetric medical staff is critical to the implementation of these patient safety elements.

Violato, C. and W. G. Hall (2000). "Alberta physician achievement review." *Canadian Medical Association Journal* 162(13): 1803.

Violato, C. and J. Lockyer (2006). "Self and peer assessment of pediatricians, psychiatrists and medicine specialists: implications for self-directed learning." *Advances in Health Sciences Education* 11(3): 235-44.

Self-regulation in medicine depends on accurate self-assessment. The purpose of the present study was to examine the discrepancy between self and peer assessments for a group of specialist physicians from internal medicine (IM), pediatrics, and psychiatry clinical domains (i.e., patient management, clinical assessment, professional development, and communication). Data from 304 psychiatrists, pediatricians and internal medicine specialists were used. Each physician had data from an identical self and 8 peer (38 item/4 clinical domains assessment). A total of 2306 peer assessments were available. Physicians were classified into quartiles based on mean assessment peer data and compared with self-assessment data. The analyses showed that self and peer assessment profiles were consistent across specialties and domains. Physicians assessed in the lowest and highest quartiles (i.e., <25th and >75th) by colleagues tended to rate themselves 30-40 percentile ranks higher and lower than peers, respectively. This study suggests that practicing physicians are inaccurate in assessing their own performance. These data suggest that systems to provide practicing physicians with regular and routine feedback may be appropriate if we are to ensure physicians are able to accurately assess themselves in a profession in which self-regulation is predicated upon the assumption that physicians know their capabilities and limitations.

Violato, C., J. M. Lockyer, et al. (2008). "Changes in performance: a 5-year longitudinal study of participants in a multi-source feedback programme." *Medical Education* 42(10): 1007-13.

**OBJECTIVES:** Multi-source feedback (MSF) enables performance data to be provided to doctors from patients, co-workers and medical colleagues. This study examined the evidence for the validity of MSF instruments for general practice, investigated changes in performance for doctors who participated twice, 5 years apart, and determined the association between change in performance and initial assessment and socio-demographic characteristics. **METHODS:** Data for 250 doctors included three datasets per doctor from, respectively, 25 patients, eight co-workers and eight medical colleagues, collected on two occasions. **RESULTS:** There was high internal consistency ( $\alpha > 0.90$ ) and adequate generalisability ( $E_p(2) > 0.70$ ). D study results indicate adequate generalisability coefficients for groups of eight assessors (medical colleagues, co-workers) and 25 patient surveys. Confirmatory factor analyses provided evidence for the validity of factors that were theoretically expected, meaningful and cohesive. Comparative fit indices were 0.91 for medical colleague data, 0.87 for co-worker data and 0.81 for patient data. Paired t-test analysis showed significant change between the two assessments from medical colleagues and co-workers, but not between the two patient surveys. Multiple linear regressions explained 2.1% of the variance at time 2 for medical colleagues, 21.4% of the variance for co-workers and 16.35% of the variance for patient assessments, with professionalism a key variable in all regressions. **CONCLUSIONS:** There is evidence for the construct validity of the instruments and for their stability over time. Upward changes in performance will occur, although their effect size is likely to be small to moderate.

Violato, C. and S. Saberton (2006). "Assessing medical radiation technologists in practice: a multi-source feedback system for quality assurance." *Canadian Journal of Medical Radiation Technology* 37(2): 10-17.

**Objective:** To develop and psychometrically evaluate (feasibility, reliability, validity) a questionnaire based multi-source feedback (MSF) system for quality assurance (QA) for medical radiation technologists (MRTs).

**Methods:** Surveys with 12, 14, 30 and 31 items were developed for assessment of MRTs by patients, co-workers, MRT colleagues and themselves, respectively, using 5-point scales with an "unable to assess" category. Core competencies assessed were related to legislation, standards of practice and ethics, equipment and materials, diagnostic examinations and radiation treatment, safe practice, relationship with patients, records and reporting, communications, team work and professional development. Each MRT was assessed by 25 patients, 6 MRT colleagues, and 6 co-workers. Feasibility was assessed with response rates for each instrument. Validity was assessed with a table of specifications, the percentage of participants unable to assess the MRT for each item, and exploratory factor analyses to determine which items grouped together into scales. Cronbach's alpha assessed reliability.

**Results:** A sample of 307 MRTs registered with the College of Medical Radiation Technologists of Ontario (CMRTO) volunteered to undergo MSF assessments. The number of respondents for the MRT peer questionnaire was 1,730, for the co-worker questionnaire was 1,745 and for the patient questionnaire were 7,007. The mean ratings ranged from 4 to 5 for each item on each scale. All of the instruments' full scales had very high Cronbach's alpha  $> .92$ . The factor analyses revealed a 5-factor solution (71.28% of the total variance) for the peer and self surveys, a 3-factor solution for the co-worker questionnaire (74.28% of the variance), and a 2-factor solution for the patient questionnaire (81.53% of the variance).

**Conclusion:** An MSF system employing surveys that have high reliability, validity and feasibility were developed to provide feedback to MRTs on core competencies and skills.

Weber, D. (2003). "25 Health care trends. What's hot, what's not, and what does the future hold." *Physician Executive* 29(1): 6-14.

As we usher in 2003, America's health care system remains in a chaotic state. Will managed care live or die? Will quality improvement efforts pay off? Are we ready for the next bioterrorism attack? Will the shortage of physician soon rival the shortage of nurses? To help gauge where health care stands today and what the future holds, The Physician Executive asked doctors who serve on ACPE's peer review panel to list the hottest health care trends in the U.S right now. Then, we took the list to three respected health care futurists -- Leland Kaiser, PhD, Jeff Goldsmith, PhD, and Russel Coile, MBA -- and asked them for their insights on the trends. Yes, Kaiser, Goldsmith and Coile are opinionated. Yes, they're controversial. But no matter whether you agree or disagree with their views, the three health care futurists' comments could spark discussions that will help shape U.S. health care this year and beyond. The trends are presented in no particular order.

Weigelt, J. A., K. J. Brasel, et al. (2004). "The 360-degree evaluation: increased work with little return?" *Current Surgery* 61(6): 616-26; discussion 627-8.

**OBJECTIVE:** To test a 360-degree resident evaluation tool on our trauma/critical care services to determine if multiple raters yielded equivalent information compared with traditional faculty evaluations. **DESIGN:** Prospective evaluation. **PARTICIPANTS:** Residents, nurses, faculty, and staff at an academic medical center. **METHODS:** The evaluation tool was developed based on extensive qualitative analysis of 13 major medical specialties' Residency Review Committee (RRC) criteria relative to the ACGME competencies and then revised with content specific to surgery. The evaluation contained 19 items divided into ACGME competency areas. Each item was scored on a 1 to 9 Likert scale: 1 = not meeting expectations and 9 = exceeding expectations. Residents on the trauma and surgical intensive care unit rotations evaluated themselves, and they were also evaluated by chief residents, surgical intensive care unit fellows, faculty, surgical intensive care unit nurses, trauma nurse clinicians, and nurse practitioners. Multiple analyses of variance were used to compare ratings by rater groups. **RESULTS:** Ten residents were evaluated on the trauma service from April to August 2003. Between 74 and 106 evaluations were obtained per resident per competency area. Average scores across the competencies were remarkably similar, ranging from 6.18 for practice-based learning and systems-based practice to 6.54 for professionalism. Although there was variability within rater groups, ratings were not statistically different between groups for any ACGME competency. **CONCLUSIONS:** The 360-degree evaluations provide limited new information compared with traditional faculty ratings. Follow-up studies are required to confirm this finding with larger samples of residents and surgical specialties.

Wenghofer, E. F., D. Way, et al. (2006). "Effectiveness of an enhanced peer assessment program: introducing education into regulatory assessment." *Journal of Continuing Education in the Health Professions* 26(3): 199-208.

**INTRODUCTION:** The College of Physicians and Surgeons of Ontario developed an enhanced peer assessment (EPA), the goal of which was to provide participating physicians educational value by helping them identify specific learning needs and aligning the assessment process with the principles of continuing education and professional development. In this article, we examine the educational value of the EPA and whether physicians will change their practice as a result of the recommendations received during the assessment. **METHODS:** A group of 41 randomly selected physicians (23 general or family practitioners, 7 obstetrician-gynecologists, and 11 general surgeons) agreed to participate in the EPA pilot. Nine experienced peer assessors were trained in the principles of knowledge translation and the use of practice resources (tool kits) and clinical practice guidelines. The EPA was evaluated through the use of a postassessment questionnaire and focus groups. **RESULTS:** The physicians felt that the EPA was fair and educationally valuable. Most focus group participants indicated that they implemented recommendations made by the assessor and made changes to some aspect of their practice. The physicians' suggestions for improvement included expanding the assessment beyond the current medical record review and interview format (eg, to include multisource feedback), having assessments occur at regular intervals (eg, every 5 to 10 years), and improving the administrative process by which physicians apply for educational credit for EPA activities. **CONCLUSIONS:** The EPA pilot study has demonstrated that providing detailed individualized feedback and optimizing the one-to-one interaction between assessors and physicians is a promising method for changing physician behavior. The college has started the process of aligning all its peer assessments with the principles of continuing professional development outlined in the EPA model.

Wenghofer, E. F., A. P. Williams, et al. (2006). "Physician-patient encounters: the structure of performance in family and general office practice." *Journal of Continuing Education in the Health Professions* 26(4): 285-93.

**INTRODUCTION:** The College of Physicians and Surgeons of Ontario, the regulatory authority for physicians in Ontario, Canada, conducts peer assessments of physicians' practices as part of a broad quality assurance program. Outcomes are summarized as a single score and there is no differentiation between performance in various aspects of care. In this study we test the hypothesis that physician performance is multidimensional and that dimensions can be defined in terms of physician-patient encounters. **METHODS:** Peer assessment data from 532 randomly selected family practitioners were analyzed using factor analysis to assess the dimensional structure of performance. Content validity was confirmed through consultation sessions with 130 physicians. Multiple-item measures were constructed for each dimension and reliability calculated. Analysis of variance

determined the extent to which multiple-item measure scores would vary across peer assessment outcomes. RESULTS: Six performance dimensions were confirmed: acute care, chronic conditions, continuity of care and referrals, well care and health maintenance, psychosocial care, and patient records. DISCUSSION: Physician performance is multidimensional, including types of physician-patient encounters and variation across dimensions, as demonstrated by individual practice. A conceptual framework for multidimensional performance may inform the design of meaningful evaluation and educational recommendations to meet the individual performance of practicing physicians.

Whitehouse, A., A. Hassell, et al. (2007). "360 degree assessment (multisource feedback) of UK trainee doctors: field testing of team assessment of behaviours (TAB)." *Medical Teacher* 29(2-3): 171-6.

BACKGROUND: This study was to see if the team assessment of behaviours (TAB) 360 degree assessment tool was able to identify interpersonal behaviour problems in doctors in training, to see if feedback was useful, to gauge the value of the process by those involved, and to learn lessons about implementing the process for the future. METHODS: TAB was administered to assess interpersonal behaviours of senior house officers in four hospitals in the West Midlands, UK. In addition, questionnaires were sent to all participants, some were interviewed about the whole process, and records kept of the time involved. RESULTS: One hundred and seventy-one SHO volunteers received 1378 assessments. The median number of ratings per SHO was 8 (mode 9). Sixty-four percent of SHOs received 'no concern' ratings in all four behaviours (domains) assessed. Twenty-one percent received one 'some concern' rating. Fifteen percent received more than one 'concern' rating. CONCLUSION: Assessors and trainees found the process practical, valuable and fair. Educational supervisors found it valuable, although only 23% learned something new about their trainees. Clinical tutors valued the system. Administrative staff found it time consuming. The TAB four-domain rating form with its single pass category identified specific concern about volunteer trainees' professional behaviour. Not all trainees received skilled feedback.

Whitehouse, A., M. Walzman, et al. (2002). "Pilot study of 360 degree assessment of personal skills to inform record of in training assessments for senior house officers." *Hospital Medicine* (14623935) 63(3): 172-175.

The assessment of a trainee doctor's human skills, including teamworking, communication and maintaining trust, is difficult but necessary to confirm competence and to alert trainees and trainers of potential problems before they become intractable. This study used 360° team observation reports for this purpose. The process was easy to administer and valued by trainees.

Wick, M. R. (2007). "Medicolegal liability in surgical pathology: a consideration of underlying causes and selected pertinent concepts." *Seminars in Diagnostic Pathology* 24(2): 89-97.

Malpractice actions against surgical pathologists are still relatively uncommon, but they have increased in frequency over time and are associated with sizable indemnity figures. This discussion categorizes areas of liability in surgical pathology into three groups: those that represent health system flaws (problems with specimen identification, or transportation, or both; lack of clinical information or erroneous information; sampling effects and defects; and poorly reproducible or poorly defined diagnostic or prognostic criteria), others that exist at the interface between the system and individuals (allowing clinicians to bypass pathologic review of referred specimens; acceding to clinical demands for inadvisable procedures; and working in a disruptive environment), and truly individual errors by pathologists (lapses in reasoning; deficiencies concerning continuity in the laboratory; invalid assumptions regarding recipients of surgical pathology reports; over-reliance on the results of "special" tests; and problems with peer consultation). Finally, two important topic areas are discussed that commonly enter into lawsuits filed against surgical pathologists; namely, "delay in diagnosis" of malignant neoplasms and "failure to provide adequate prognostic information." Based on a review of the pertinent literature, we conclude that the clinical courses of most common malignancies are not affected in a significant manner by delays in diagnosis. Moreover, the practice of using "personalized external validity" for supposedly prognostic tests is examined, with the resulting opinion that prognostication of tumor behavior in individual patients is not reliable using anything but anatomic staging systems. copyright 2007 Elsevier Inc. All rights reserved.

Wilkinson, J. R., J. G. Crossley, et al. (2008). "Implementing workplace-based assessment across the medical specialties in the United Kingdom." *Medical Education* 42(4): 364-373.



**OBJECTIVES:** To evaluate the reliability and feasibility of assessing the performance of medical specialist registrars (SpRs) using three methods: the mini-clinical evaluation exercise (mini-CEX), directly observed procedural skills (DOPS) and multi-source feedback (MSF) to help inform annual decisions about the outcome of SpR training. **METHODS:** We conducted a feasibility study and generalisability analysis based on the application of these assessment methods and the resulting data. A total of 230 SpRs (from 17 specialties) in 58 UK hospitals took part from 2003 to 2004. Main outcome measures included: time taken for each assessment, and variance component analysis of mean scores and derivation of 95% confidence intervals for individual doctors' scores based on the standard error of measurement. Responses to direct questions on questionnaires were analysed, as were the themes emerging from open-comment responses. **RESULTS:** The methods can provide reliable scores with appropriate sampling. In our sample, all trainees who completed the number of assessments recommended by the Royal Colleges of Physicians had scores that were 95% certain to be better than unsatisfactory. The mean time taken to complete the mini-CEX (including feedback) was 25 minutes. The DOPS required the duration of the procedure being assessed plus an additional third of this time for feedback. The mean time required for each rater to complete his or her MSF form was 6 minutes. **CONCLUSIONS:** This is the first attempt to evaluate the use of comprehensive workplace assessment across the medical specialties in the UK. The methods are feasible to conduct and can make reliable distinctions between doctors' performances. With adaptation, they may be appropriate for assessing the workplace performance of other grades and specialties of doctor. This may be helpful in informing foundation assessment.

Williams, B. W. (2006). "The prevalence and special educational requirements of dyscompetent physicians." *Journal of Continuing Education in the Health Professions* 26(3): 173-91. Underperformance among physicians is not well studied or defined; yet, the identification and remediation of physicians who are not performing up to acceptable standards is central to quality care and patient safety. Methods for estimating the prevalence of dyscompetence include evaluating available data on medical errors, malpractice claims, disciplinary actions, quality control studies, medical record review studies, and in-stream assessments of physician performance. These data provide a range of estimates from 0.6% to 50%, depending on the method. A reasonable estimate of dyscompetence appears to be 6% to 12%. Age-related cognitive decline, impairment due to substance use disorders, and other psychiatric illness can contribute to underperformance, diminishing physicians' insight into their level of performance as well as their ability to benefit from an educational experience. Currently, dyscompetent physicians in the United States are identified through either the legal system or peer review. The primary method of resolving issues of underperformance in physicians is through continuing medical education (CME). Although a number of specialized assessment and education programs exist in the United States, these programs are largely underutilized. Similar programs exist in Canada and have provided evidence of the efficacy of a more specialized and individualized educational approach for underperforming physicians. Current specialty programs focused on this population employ individual assessments of knowledge and performance, individually designed educational programs, long-term plans for maintenance of educational activity, and repeated assessment of performance level. Noting that few CME programs offer these requirements, a number of changes to current medical quality assurance programs that might foster such educational requirements for underperforming physicians are provided.

Williams, J. R., K. Mechler, et al. (2008). "Innovative peer review model for rural physicians: system design and implementation." *Journal of Rural Health* 24(3): 311-5. **CONTEXT:** The peer review process in small rural hospitals is complicated by limited numbers of physicians, conflict of interest, issues related to appropriate utilization of new technology, possibility for conflicting recommendations, and need for external expertise. **PURPOSE:** The purpose of this project was to design, test, and implement a virtual peer review system for small rural hospitals in Texas. We sought to define the characteristics of a virtual peer review system in the context of rural health care, and to explore the benefits from peer review administration within a rural network supported by a university. **METHODS:** Physicians from small rural hospitals participated in pilot testing of the system. Policies and procedures reflecting the innovative character of the new peer review process were developed based on legal/regulatory requirements and desired educational focus of the process. An information technology system to support the virtual peer review was selected, tested, and deployed. **FINDINGS:** The system tests suggested feasibility of the procedures, reliability of the communication lines, and functional anonymity of the hospitals and physicians participating in the virtual peer review. Participating institutions and individual physicians expressed satisfaction with the reliability and user friendliness of the system as demonstrated during the pilot

tests. CONCLUSIONS: Hospital licensing and accreditation require a process to monitor and evaluate the care of patients. Utilizing means of virtual communication is a viable option for small rural hospitals. This process is dependable, user-friendly and provides functional anonymity to participating hospitals and physicians. The peer review system has successfully functioned since 2004.

Wood, J., J. Collins, et al. (2004). "Patient, faculty, and self-assessment of radiology resident performance: a 360-degree method of measuring professionalism and interpersonal/communication skills." *Academic Radiology* 11(8): 931-9.

RATIONALE AND OBJECTIVES: To develop and test the reliability, validity, and feasibility of a 360-degree evaluation to measure radiology resident competence in professionalism and interpersonal/communication skills. MATERIALS AND METHODS: An evaluation form with 10 Likert-type items related to professionalism and interpersonal/communication skills was completed by a resident, supervising radiologist and patient after resident-patient interactions related to breast biopsy procedures. Residents were also evaluated by faculty, using an end-of-rotation global rating form. Residents, faculty, and technologists were queried regarding their reaction to the assessments after a 7-month period. RESULTS: Fifty-six complete 360-degree data sets (range, 2-14 per resident) and seven rotational evaluations for seven residents were analyzed and compared. Internal consistency reliability estimates were 0.85, 0.86, and 0.87 for resident, patient, and faculty 360-degree evaluations, respectively. Correlations between resident-versus-patient, resident-versus-faculty, and patient-versus-faculty ratings for the 56 interactions were -0.06 ( $P = .64$ ), 0.31 ( $P < .02$ ), and 0.45 ( $P < .0006$ ), respectively. Pearson correlation coefficients approached significant correlation (0.70) between the faculty global rating and patient 360-degree scores ( $P = .08$ ) but not with faculty 360-degree scores. Residents and faculty felt that completing the 360-degree forms was easy, but the requirement for faculty presence during the consent process was burdensome. CONCLUSION: Results from this pilot study suggest that self, faculty, and patient evaluations of resident performance constitutes a valid and reliable assessment of resident competence. Additional data are needed to determine whether the 360-degree assessment should be incorporated into residency programs and how frequently the assessment should be performed. Requiring only a specified number of assessments per rotation would make the process less burdensome for residents and faculty.

Wood, L., A. Hassell, et al. (2006). "A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design." *Medical Teacher* 28(7): e185-91.

Multi-source feedback (MSF) has become the accepted mechanism of ensuring the appropriate professional behaviour of doctors. It is part of the mandatory assessment of doctors in training and is to be utilized as part of the revalidation of trained doctors. There is significant variation in the models of MSF currently used within the National Health Service and new models of MSF are being designed by various specialties. No single model has been recognized as the 'gold standard'. However, there is a large published literature concerning MSF, both in the context of health systems and, more extensively, within industry. This published literature is reviewed, drawing attention to aspects of MSF systems in which there is consensus on effective approaches as well as other aspects in which there is doubt about the optimum approach. In the light of the review 10 principles key in the development of effective MSF models have been produced.

Wood, L., D. Wall, et al. (2006). "'Team observation': a six-year study of the development and use of multi-source feedback (360-degree assessment) in obstetrics and gynaecology training in the UK." *Medical Teacher* 28(7): e177-84.

Multi-source feedback, or 360-degree assessment, is an important part of the assessment of people in the workplace, in both health and industry. Almost all published work concentrates on content validity and generalizability. However, an assessment system needs construct validity, and has to have practicability and acceptability, without sacrificing fitness for purpose, content validity or inter-rater reliability. This was a six-year study of the first UK-wide hospital-based multi-source feedback system, in the specialty of obstetrics and gynaecology. This paper describes the development of the assessment tool, its use and the analyses of the results in several areas. These are picking up poor performance, congratulating good behaviour, construct validity, the number of domains to be measured, and the minimum number of raters. The study demonstrated that the Team Observation system in reality only measured a very limited number of attributes, and that the main construct under scrutiny is interpersonal behaviour. The system can identify those who may have a problem, using less than 10 raters, and yet the process can be a positive experience for the large majority of people who have been assessed.

Zbar, R. I. S., L. D. Taylor, et al. (2009). "The disruptive physician: righteous Maverick or dangerous Pariah?" *Plastic & Reconstructive Surgery* 123(1): 409-15.

**BACKGROUND:** "Disruptive physician" is a term appearing more frequently in many hospital bylaws. It has significant negative implications that can lead to loss of privileges for plastic surgeons. **METHODS:** Exploring the various definitions of disruptive physician reveals palpable differences between those of the Joint Commission and the American Medical Association. These discrepancies expose plastic surgeons to potential harm when actively addressing quality issues in the hospital environment. **RESULTS:** The disruptive label can be inappropriately leveraged by hospital administrators against plastic surgeons who confront quality issues. Moreover, the term disruptive is open to subjective interpretation. Challenging the disruptive label in court reveals only that the justice system is concerned that the actual process leading to the disruptive charge is followed appropriately as outlined within the organizational bylaws; the courts are not interested in the actual quality issues and generally will not second-guess the judgment of peer review panels or hospital administrators. **CONCLUSIONS:** Plastic surgeons would benefit from familiarizing themselves with these issues. Hospitals should be required to use root cause analysis when dealing with quality issues raised by members of the medical staff. Furthermore, findings from root cause analysis should be privileged from legal discovery in all jurisdictions to permit honest exploration of quality issues. When a conflict does arise, consideration of mediation should be given to resolve disputes.