*Medical Teacher, Vol. 28, No. 2, 2006, pp. 117–128*

Taylor & Francis
Taylor & Francis Group

**BEME GUIDE**

# Systematic review of the literature on assessment, feedback and physicians' clinical performance*: BEME Guide No. 7

JON VELOSKI[1], JAMES R. BOEX[2], MARGARET J. GRASBERGER[1], ADAM EVANS[1] & DANIEL B. WOLFSON[3]

*[1]Jefferson Medical College, Thomas Jefferson University, Philadelphia, USA; [2]Northeastern Ohio Universities College of Medicine, USA; [3]American Board of Internal Medicine Foundation, Philadelphia, USA*

ABSTRACT Review date: *1966 to April 2003.*

Background and context: *There is a basis for the assumption that feedback can be used to enhance physicians' performance. Nevertheless, the findings of empirical studies of the impact of feedback on clinical performance have been equivocal.*

Objectives: *To summarize evidence related to the impact of assessment and feedback on physicians' clinical performance.*

Search strategy: *The authors searched the literature from 1966 to 2003 using MEDLINE, HealthSTAR, the Science Citation Index and eight other electronic databases. A total of 3702 citations were identified.*

Inclusion and exclusion criteria: *Empirical studies were selected involving the baseline measurement of physicians' performance and follow-up measurement after they received summaries of their performance.*

Data extraction: *Data were extracted on research design, sample, dependent and independent variables using a written protocol.*

Data synthesis: *A group of 220 studies involving primary data collection was identified. However, only 41 met all selection criteria and evaluated the independent effect of feedback on physician performance. Of these, 32 (74%) demonstrated a positive impact. Feedback was more likely to be effective when provided by an authoritative source over an extended period of time. Another subset of 132 studies examined the effect of feedback combined with other interventions such as educational programmes, practice guidelines and reminders. Of these, 106 studies (77%) demonstrated a positive impact. Two additional subsets of 29 feedback studies involving resident physicians in training and 18 studies examining proxy measures of physician performance across clinical sites or groups of patients were reviewed. The majority of these two subsets also reported that feedback had positive effects on performance.*

Headline results: *Feedback can change physicians' clinical performance when provided systematically over multiple years by an authoritative, credible source.*

Conclusions: *The effects of formal assessment and feedback on physician performance are influenced by the source and duration of feedback. Other factors, such as physicians' active involvement in the process, the amount of information reported, the timing and amount of feedback, and other concurrent interventions, such as education, guidelines, reminder systems and incentives, also appear to be important. However, the independent contributions of these interventions have not been well documented in controlled studies. It is recommended that the designers of future theoretical as well as practical studies of feedback separate the effects of feedback from other concurrent interventions.*

## Context

### *Feedback and clinical performance*

The findings of research on the quality of healthcare over many decades have led policy-makers and leaders in the medical profession to search for new ways to assure that all physicians recognize accepted professional standards and translate research findings into practice (James, 2001). This challenge was traditionally addressed within the context of continuing medical education lectures. However, a wide array of interventions has been considered beyond traditional educational conferences and materials. These include educational outreach visits, local consensus processes, local opinion leaders, marketing, patient-mediated interventions, reminder systems and feedback (Mazmanian & Davis, 2002).

Feedback to physicians has been used in combination with chart reviews and medical audits (Gonnella *et al.*, 1970). In the recent Cochrane review of the impact of feedback on performance (O'Brien *et al.*, 2003), feedback was defined as ''any summary of clinical performance of health care over a specified period of time''. Medical educators and healthcare managers have reasoned for decades that physicians would welcome such information about the congruence between their own practice patterns gleaned from medical record

*Correspondence*: Jon Veloski, Director of Medical Education Research, Center for Research in Medical Education and Health Care, Jefferson Medical College, Thomas Jefferson University, 1025 Walnut St. Suite 119, Philadelphia, PA 19107, USA. Email: jon.veloski@jefferson.edu

*This BEME Systematic Review, complete with tables, figures and appendices, is published as a BEME Guide: *Systematic Review of the Literature on Assessment, Feedback, and Physicians' Clinical Performance: BEME Guide No. 7.* Veloski *et al.* Dundee, Association for Medical Education in Europe (2006). ISBN 1-903934-32-X. (http://www.amee.org)

ISSN 0142–159X print/ISSN 1466–187X online/06/020117–12 © 2006 Taylor & Francis
DOI: 10.1080/01421590600622665

reviews and accepted professional practice standards. This method has been extended to the analysis of insurance plan billing records (Woodward *et al.*, 1989; Woodward *et al.*, 1990; Hanchak *et al.*, 1996) and other approaches such as profiling and physician report cards (Dranove *et al.*, 2002).

Multiple reviews of the impact of feedback on physician performance have been published in recent decades. While some of these reviews have been conducted within the context of continuing medical education as a means of influencing physicians' behaviour (Mazmanian & Davis, 2002), others have addressed the entire spectrum of interventions designed to shape providers' behaviour (Grimshaw *et al.*, 2001). This spectrum includes not only lectures and conferences, but also educational materials, educational outreach visits, local consensus processes, local opinion leaders, marketing, patient-mediated interventions and reminders (Grimshaw *et al.*, 2001).

We evaluated the 54 reviews listed in Appendix 1 (on BEME website http://www.bemecollaboration.org). Although these reviews provided some information on the potential impact of feedback on physician performance, we found that their practical value was limited. Therefore, we recognized that we would need to address the following four issues in the design of the BEME review to differentiate it from these predecessors and to provide useful information regarding the impact of feedback on physicians' clinical performance.

### Physicians versus other healthcare providers

Medical students, resident physicians in training, dentists, physicians' assistants, nurses and other healthcare providers have been the research subjects in a large number of studies of the impact of feedback on professional performance. Consequently, nearly all reviews, including several of the most rigorous systematic reviews, have grouped studies of experienced licensed physicians with studies of other providers. However, given the roles the various members of the healthcare team play, it is unclear that it is appropriate to combine the findings of studies of the impact of feedback on different professions.

The present BEME review was designed to include empirical studies of only licensed physicians, both MDs and DOs. This included studies of practising physicians, as well as house staff at different levels of training.

### Construct of physician performance

Previous reviews have considered multiple measures of performance but have not combined them in a logical structure. Studies of feedback have addressed diverse aspects of clinical performance such as prescribing behaviour, compliance with practice guidelines, patient outcomes, innovations in practice, quality, economy, nurse–doctor collaboration, preventive care, immunization, cancer screening, smoking cessation and referrals to specialists (Grimshaw *et al.*, 2001). In the development of the BEME review we recognized the need to organize these types of performance and outcome measures within some set of constructs, to define physician performance, and to interpret the impact of feedback on measures of performance.

### Other interventions as moderating variables

A number of the published reviews related to feedback have concluded that the effect of feedback is moderated by variables such as education, reminders and academic detailing (O'Brien *et al.*, 2003). We recognized that it would be important to identify the use of these other interventions designed to influence physicians' performance and to evaluate their independent and interaction effects in relation to the use of feedback.

### Consideration of empirical studies beyond randomized controlled trials

Consistent with the recommendations of the BEME Collaboration, this review was not limited to RCTs. We considered all empirical studies, including non-randomized groups, observational and qualitative studies.

In summary, four characteristics of this BEME review distinguish it from other published reviews of the impact of feedback on clinical performance. We limited its scope to studies of physicians. We focused attention on measures of physicians' clinical performance and identified the presence of these moderating variables. Finally, we considered all empirical studies including quasi-experimental designs.

### Best evidence medical education (BEME)

The Best Evidence Medical Education (BEME) Collaboration (Harden *et al.*, 1999) involves an international group of individuals, universities and professional organisations (e.g. AMEE, AAMC, ABIM) committed to moving the medical profession from opinion-based education to evidence-based education. BEME's goal is to provide leaders with the latest findings from scientifically grounded educational research. This will enable teachers and administrators to make informed decisions on the kinds of evidence-based education initiatives that boost learner performance on cognitive and clinical measures. BEME rejects the legacy of medical education in which decisions have been made based on pseudoscience, anecdotes and flawed comparison groups rather than empirical evidence. The BEME approach contends that in no other scientific field are personal experiences relied on to make policy choices, and in no other field is the research base so limited.

BEME scholarship ''involves professional judgment by the teacher (or administrator) about his/her teaching taking into account a number of factors—the *QUESTS* dimensions: the *Quality* of the research evidence available—how reliable is the evidence? the *Utility* of the evidence—can the methods be transferred and adopted without modification? the *Extent* of the evidence, the *Strength* of the evidence, the *Target* or outcomes measured—how valid is the evidence?, and the *Setting* or context—how relevant is the evidence?'' (Harden *et al.*, 1999).

The international BEME Collaboration has three broad purposes. First, to produce systematic reviews of medical education research studies that capture the best evidence available and also meet users' needs. Second, to disseminate information worldwide to all stakeholders to make decisions concerning medical education on grounds of the best available evidence. Third, to create a culture of best evidence

medical education among teachers, administrators, educational institutions, and national and international organizations.

## This report

This BEME review was conceived by the American Board of Internal Medicine, which selected the Center for Research in Medical Education and Health Care at the Jefferson Medical College of Thomas Jefferson University in Philadelphia, Pennsylvania, USA to conduct the review. Jon Veloski, Director of Medical Education Research at the Center, accepted the charge to review and synthesize existing evidence in educational and healthcare science that addresses a specific question: ''What are the features and characteristics of feedback that influence physicians' clinical performance?'' This report presents the methodological scope and detail of the study, its principal findings, and a discussion about what the findings mean for evidence-based medical education today and tomorrow.

Four sections follow. The *Methods* section describes the protocol for the review and its implementation. The *Results* section summarizes the research reports included in the systematic review and the features of feedback that evidence shows to lead to effective performance. A *Discussion* section interprets our findings, acknowledges the review's limitations, critiques the quality of published research on the impact of feedback on physician performance, and outlines a research agenda. Finally, a brief *Conclusion* describes the essence of the review.

## Methods

### *Development of protocol*

We began to develop the protocol for the systematic review in July 2002 based on our general understanding of the review question and a set of 34 related articles readily available in the files of Jefferson's Center for Research in Medical Education and Health Care. Included in these articles were several published reviews addressing methods of changing physicians' clinical behaviour and the use of feedback with healthcare providers. We developed a draft data collection form and brief analysis plan during the development of the protocol.

### *Review question and objectives*

After completing the protocol in November 2002, we began the formal search for articles related to the question, ''What is the impact of assessment and feedback on physician performance?'' Although we believed that this review question and related objectives had been specified clearly in the protocol, we proceeded to test this assumption by coding a sample of 100 articles, using the coding form and codebook drafted in the protocol.

We found that this process of testing the coding form and associated procedures strengthened our understanding of the operational definitions of *assessment*, *performance* and *feedback*. We were also able to identify a set of *moderating variables*, or conditions, that affect the relationship between feedback and physicians' performance. The following three sections summarize the definitions of *clinical performance* and *feedback* that we used throughout the review as well as the nature of the *moderating variables* that we expected to uncover in the review.

### *Assessment and physicians' clinical performance*

Figure 1 shows a comprehensive model of the relationship between physicians' competence and performance that was presented in the protocol as a conceptual base for the review. The definitions of assessment and physician performance as well as the interpretation of studies paid close attention to the impact of other variables on physician performance. These included factors in the local micro system as well as variables in the larger macro system (Gonnella *et al.*, 1993). For example, as will be discussed later, we set aside a group of studies in which the unit of analysis was not individual physicians. In these studies the dependent variable was a clinical process or patient outcome that might be a function not only of the performance of individual physicians, but also of other providers, environmental factors or the patients themselves.

Figure 2 summarizes the Value Compass, which we used to organize the process and outcome variables that were used to assess physician performance in studies of feedback. We employed the Value Compass as a formal construct to define physician performance and interpret the impact of feedback on measures of performance along four dimensions: clinical processes, clinical outcomes, patient satisfaction, and costs. The Value Compass was developed to measure multiple dimensions of patient care (Nelson *et al.*, 1996) and has been adapted to assess the impact of variables upon multiple selected healthcare-related factor sets.

Examples of these clinical process assessments, which have high levels of content validity, included chart review as either self-assessment or within a local practice or group of practices; medical record audit by a third-party; analysis of computerized medical records; analysis of insurance plan billing records (Woodward *et al.*, 1989; Woodward *et al.*, 1990; Hanchak *et al.*, 1996); direct observation of clinicians by supervisors (Gray, 1996), peers (Ramsey *et al.*, 1993), or trained observers.

The assessment of clinical outcomes beyond the normal processes of care involved *survival*, *mortality*, *morbidity* and patient *functioning* and *well-being* (Stewart & Ware, 1992). Patient satisfaction has been assessed by measures such as reports of standardized patients (physician blinded) (Norman *et al.*, 1985; Peabody *et al.*, 2000), patient satisfaction with specific visits (Ware & Hays, 1988; Rubin *et al.*, 1993) and global patient satisfaction ratings (Tu & Reschovsky, 2002).

Cost assessments included accounting costs, inpatient length of stay (in days), and other administrative and/or managerial costs.

We had initially planned in the protocol that we would not limit the review to assessment of physician performance in clinical settings with real patients. We considered reviewing studies involving a variety of measures of physician competence such as written examinations, simulations and other tools outside the clinical arena. During the review we realized that the widely accepted definition of performance used in other reviews of feedback was based on studies with real patients in clinical settings. There were few credible feedback

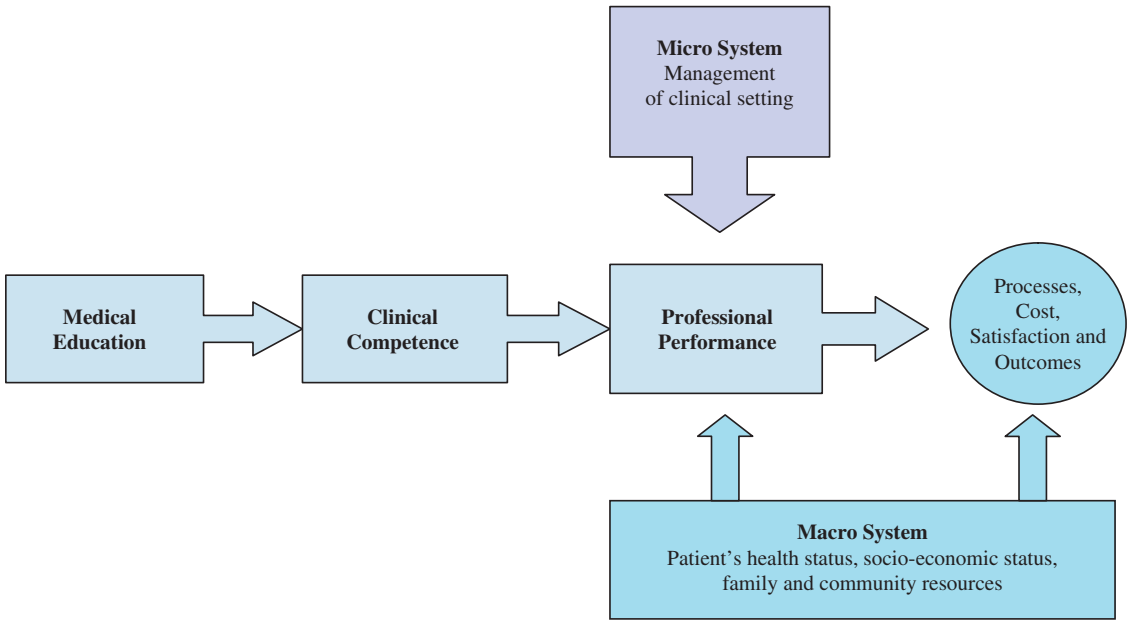


**Figure 1.** Model of physicians' clinical competence, professional performance and other variables affecting processes, costs, satisfaction and outcomes. *Source*: Adapted from: Gonnella *et al.*, *Assessment Measures in Medical School, Residency and Practice: The Connections* (New York, Springer, 1993).

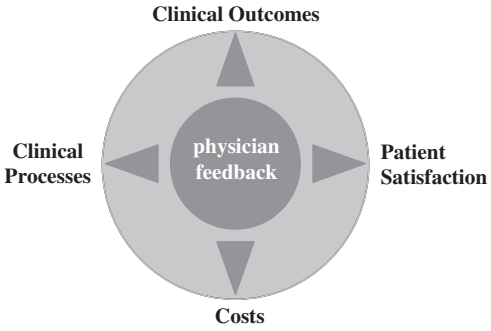


**Figure 2.** Value Compass model for physician feedback. *Source*: Nelson *et al.* (1996).

studies based on vignettes (Chang *et al.*, 1999; Peabody *et al.*, 2000), standardized patients (Norman *et al.*, 1985), and computer simulations (Issenberg *et al.*, 1999).

In summary, the definition of assessment and physicians' clinical performance used for the review was narrowed to performance in clinical settings rather than competence demonstrated on tests and simulations. It was understood that assessments would be based on measures of either processes or outcomes in these clinical settings.

*Feedback*

The term *feedback*, which originated in electronics and process controls, has a connotation that is objective and non-threatening. When a sensor in a system determines that the output of a process deviates from a target value, a controller communicates a feedback signal to adjust the system to produce the desired output. Although the concept addressed in this review is similar, the results of physician performance assessment involve communication on a much more complex level.

Feedback can be framed to communicate an individual physician's status in relation to a standard of behaviour or professional practice. This standard can be defined based on professional judgment (e.g. consensus panel), a local standard set by the individuals being studied (Norman *et al.*, 1985), or statistical norms (local norms, regional/national, historical). While the feedback given to a physician or group based on performance assessment can take the form of objective reports intended to guide professional decision-making, it can also be packaged as administrative recommendations (Hanchak & Schlackman, 1995), compensation guidelines (Hanchak *et al.*, 1996), evaluations (Nash *et al.*, 1993), report cards, rankings (Nash *et al.*, 1997), criticisms, warnings, sanctions or denials (Bindman *et al.*, 1998).

Feedback can also be communicated as praise, rewards, special recognition, or professional achievement. It was important to clarify the range of feedback to be reviewed to assure the accuracy of search terms and coding during the review.

*Moderating variables*

The third and final issue was to identify any moderating variables that could affect the relationship between feedback and performance based on the findings of studies of mixed methods of changing provider behaviour. Examples include the following:

- Was the feedback provided at the level of individual physicians or to a group?
- What was the source of feedback?
- What was the role of public disclosure?
- What was the organizational context of the medical practice, e.g. were the physicians independent practitioners or salaried employees?

- Were there other, concurrent interventions involving the redesign of systems, dissemination of protocols, guidelines, reminder systems or educational programmes?

The most important moderating variables were identified, encoded and evaluated as part of the review (Davis *et al.*, 1995, 1999; Oxman *et al.*, 1995; Grimshaw *et al.*, 2001).

*Search strategy and management of citations and abstracts*

The sources for the review included 11 electronic databases, reference lists in key published reviews of the literature, reference lists in primary sources, and citations from the contents of key, relevant journals. The MEDLINE search strategy is summarized in Table 1 (on BEME website http://www.bemecollaboration.org).

We expected to find the most relevant publications in MEDLINE. Therefore, we performed the preliminary searches during development of the protocol using OVID MEDLINE to identify search terms and to characterize the nature of the articles to be reviewed. Subject headings of acceptable articles were examined for additional terms.

We examined the MEDLINE records of relevant articles found in the reference lists of other reviews or research studies to locate additional subject headings and key phrases. These were added to the search strategy. MEDLINE records were substituted wherever duplicate citations were found in searches of other databases because of their standardization and level of detail.

The other 10 databases used in the search are not specific to medicine and healthcare and were more likely to index articles using the term 'physicians' rather than synonyms such as interns, residents or the names of individual specialties. Therefore, we were able to use simpler search strategies for these databases (see Table 2 on BEME website http://www.bemecollaboration.org).

Using the OVID multi-file searching and its capabilities to remove duplicates, we entered a very simple search strategy that simultaneously searched CINAHL, ERIC, and PsycINFO. We examined the records of the usable articles retrieved for additional subject headings and reran the search individually in the other databases.

*Reference lists and journal contents*

When the reviewers noted additional citations that appeared relevant to the review while reviewing articles, MEDLINE records were added to the database. Notable among these was the update in June 2003 of the Cochrane Library's Audit and Feedback review (O'Brien *et al.*, 2003).

Searches of the contents of *Evaluation & the Health Professions, Medical Care, Medical Education, Medical Teacher* and *Teaching and Learning in Medicine* were performed during summer and autumn of 2003.

*Management of citations and abstracts*

We developed a local database using *Reference Manager*$^{TM}$ to maintain citations and abstracts and to review tracking data. This enabled us to use OVID's *Direct Export Feature* to download the complete record of each citation retrieved by our search into this local Reference Manager database. Ovid's *Direct Export Feature* is optimized for MEDLINE, but it does not always parse the details of citation records from other databases into the correct fields when downloading to Reference Manager. Therefore, when duplicate citations were available we used either the one from MEDLINE or from the database with the best bibliographic control.

Our initial plan had been to keep both relevant and irrelevant citations in the Reference Manager database. We knew that the software had the capability to identify and purge duplicates when they appeared in later searches. We hoped this would eliminate the need to screen the same citations visually more than once. However, the software was sluggish as the size of the database grew to thousands of records. We retained only relevant citations in the database.

Duplicate citations presented problems throughout the search. As the results of new searches were downloaded, their citations were compared with the Reference Manager database to identify duplicates. Following the recent National Library of Medicine addition of HealthSTAR and Bioethicsline citations to MEDLINE, we encountered many duplicate records in MEDLINE, which Reference Manager was usually able to identify and remove. However, we were unable to intercept some duplicates that were discovered during the process of study selection described below.

**Study selection**

*Primary screen applied to results of electronic search*

Beginning in April 2003, the Information Scientist (MG) visually screened the title and abstract of the citations identified by the electronic searches and classified each as *Include*, *Exclude* or *Uncertain* using the screening criteria defined in the protocol and summarized in Table 3 (on BEME website http://www.bemecollaboration.org).

Additional notes were added that reflected the main points of the inclusion criteria to qualify these decisions for secondary screening. These notes enabled other reviewers to examine the citations in batches.

The annotation of records as they were reviewed in the database was laborious because the database had to re-index the database each time it was updated. A brief report was printed for each citation. The full citation was reviewed in the database, and the printed list was annotated. These annotations were subsequently added to the citations in the database in batches.

When new features of articles that would have implications for the inclusion or exclusion of other citations were identified, other citations in the *Reference Manager* database matching these features were retrieved and re-evaluated.

Some non-English-language articles were retrieved in the search and were screened when an abstract was available in English. None were selected for inclusion in the review.

*Secondary screen by lead member of Topic Review Group*

Beginning in May 2003, the lead member of the Topic Review Group (TRG) screened the titles and abstracts of all citations coded as *Include* or *Uncertain*. He also reviewed a 5% random sample of 100 citations that had been *Excluded*.

The findings of this secondary screen were discussed with the Information Scientist as well as key members of the Group (DW, JRB, BB, JS), and minor changes were made to the primary screening criteria.

During the early phases of the review all citations that passed the secondary screen by the lead member of the TRG were assigned to members of the Group for data extraction as described in the next section. However, the process of data extraction uncovered three types of studies. Although these studies appeared to meet all selection criteria, the following three features set them apart:

- *Feedback confounded with other interventions*—There was a large number of published studies in which the independent variable associated with the use of feedback was confounded with other interventions such as implementation of practice guidelines, educational programmes, financial incentives or opinion leaders. Thus, the intervention comprised feedback plus some other activity that could have a significant impact on the outcome. Although it was recognized that these studies would be suitable for systematic review because the use of feedback was confounded with one or more other interventions, we decided to identify them separately for review.
- *House staff*—Careful review of the articles for the studies that had been screened suggested ambiguity around the issue of supervision of house staff and the confounding effect of the educational environment on the impact of feedback on house staff. It was decided that these studies would be earmarked for a separate review.
- *Unit of analysis not physicians*—The third group included studies in which the unit of analysis was either hospitals, clinical sites, patients or large organizations involving multiple types of providers. Even when the unit was a physician group practice, many studies did not provide details on the provider mix and the specific magnitude of the effects on the performance of the physicians.

Correspondingly, the procedures for the secondary screen were modified. In addition to screening the titles and abstracts, the Lead Member of the TRG screened all articles to identify any with issues related to confounding, house staff and unit of analysis. Any articles with these characteristics were referred to another member of the Group for confirmation. The remaining articles were referred to a member for data extraction.

## Extraction and analysis of data

### *Review procedures*

We developed a four-page structured coding form with 34 forced-choice items (Appendix 2 on BEME website http://www.bemecollaboration.org). The first set of items involved the source of the citation, review tracking and the reviewer's global rating of the study's quality and overall evaluation of the effect of feedback on performance. The second set of items was related to the design and sample of the study, including a classification of the type of empirical study, and the specialty, nationality and number of physicians studied. A third group of items was related to the characteristics of the outcome variable such as the type of physician performance measured and the source of the data. A final set of items probed the quality of the feedback, including the physicians' involvement, the source of the standards used to judge performance, the volume of data communicated, the mode of communication, the timing, and the use of other interventions designed to influence physician performance.

During May 2003 a sample of representative articles that had passed the screening by the Information Scientist and the Lead Member of the TRG was used to pre-test the coding form and instructions to reviewers. Although there were no changes to the overall structure of the form, the wording of many of the items was edited for clarification, and several new items were added.

After screening, one of the investigators or a member of the TRG read each article and completed a coding form. A second form was completed independently by another investigator or member of the TRG for the articles that met all selection criteria.

The TRG Lead Reviewer reviewed the coding of all articles. Minor discrepancies, such as disagreements on study design, were resolved by discussion with the primary reviewer leading to consensus. Major discrepancies, such as disagreements on the dependent variable, were referred to another member of the TRG.

### *Classification of studies*

During the process of reviewing and selecting articles, we set aside the three groups of studies described above that involved feedback but did not meet all selection criteria for the review. These included studies of feedback combined with other interventions, resident physicians, and studies in which the unit of analysis was either hospitals, clinical sites, patients or large organizations involving multiple types of providers.

### *Data analysis*

The review forms data were entered into an Excel spreadsheet together with the data that had been collected in *Reference Manager*. We used STATA version 8.0 (College Station, TX, USA) to compute descriptive statistics. Chi-square was used to test the significance of differences in nominal variables. Independent *t*-tests were used for numerical variables.

## Quality assurance and coding accuracy

During the early phases of the review two steps were taken to enhance coding accuracy. First, the four-page coding form was pre-tested by Mr Veloski, Ms Grasberger and Dr Boex and revised accordingly throughout the development period for the BEME protocol. This procedure was repeated with other members of the TRG during the first three months of data collection. Second, a set of coding procedures with definitions of many of the terms used on the form was developed and revised in accordance with changes in the form. These procedures were used by members of the TRG when reviewing articles.

Later, throughout the implementation of the review, additional quality assurance measures were implemented. Mr Adam Evans, a medical student in Jefferson's class of 2006, was recruited as a research assistant during the

2003–04 academic year. Under Mr Veloski's supervision, Mr Evans independently reviewed and coded the articles that had been reviewed. Mr Veloski supervised the student during the review. Any articles with more than three discrepancies in coding were referred to a member of the TRG for a third independent review. Mr Veloski, Ms Grasberger and Dr Boex handled the largest share of the volume, but other members of the TRG were involved to resolve difficult cases.

## Results

The search identified 3702 citations (Figure 3). Screening of the titles and abstracts reduced this to 638 after excluding 3064 ineligible articles (Figure 4). Examples of the latter include commentaries, descriptive reviews of small, arbitrary samples of studies, non-empirical descriptive reports, studies of medical students or non-physician providers, and studies of feedback on teaching skills to faculty in medical education programmes.

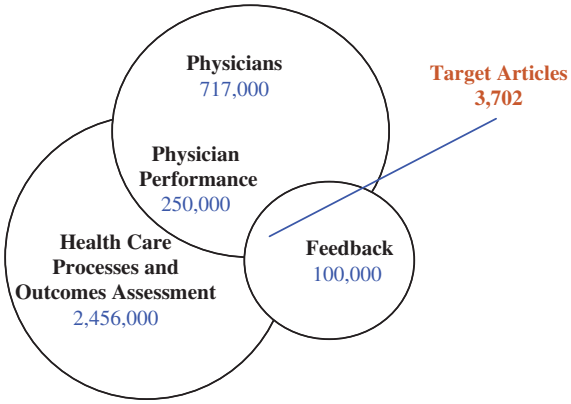


**Figure 3.** Search strategy and approximate number of articles: impact of feedback on physician performance.

The remaining 638 citations were reviewed. A total of 369 were excluded for reasons such as incomplete reporting of methods and results, the absence of baseline assessments, or the absence of follow-up measurements. There were 90 published studies among the 369 in which the term 'feedback' was used to describe the prospective communication of information to physicians related to their management of specific patients. This type of intervention is closer to the definition of reminders (Grimshaw *et al.*, 2001). Examples included the communication of cost information when the physician ordered a diagnostic test, the cost of medications prescribed, recommendations for management options based on a patient's diagnosis, and other types of reminders based on management protocols. We chose to exclude these studies by adhering to the conventional definition of feedback, which is consistent with common usage and with the major studies and reviews related to physicians' clinical performance (Jamtvedt *et al.*, 2003).

There were 132 studies in which feedback was combined with one or more other interventions such as educational programmes, practice guidelines and reminders. Another 29 studies involved resident physicians in training. Finally, there were 18 studies based on units of analysis other than physicians. These included feedback and control groups defined as geographic regions, hospitals, units within hospitals and primary care sites.

## Studies meeting all selection criteria

### *Descriptive analysis*

A set of 41 studies met all selection criteria (Appendix 3 on BEME website http://www.bemecollaboration.org). Although the earliest had been published in 1977 (Wennberg *et al.*, 1977), one-third appeared after 1996. Four of the 41 were published in *Medical Care*, four in the *British Medical Journal* and two in the *Journal of the American Medical Association*. The remainder were dispersed across

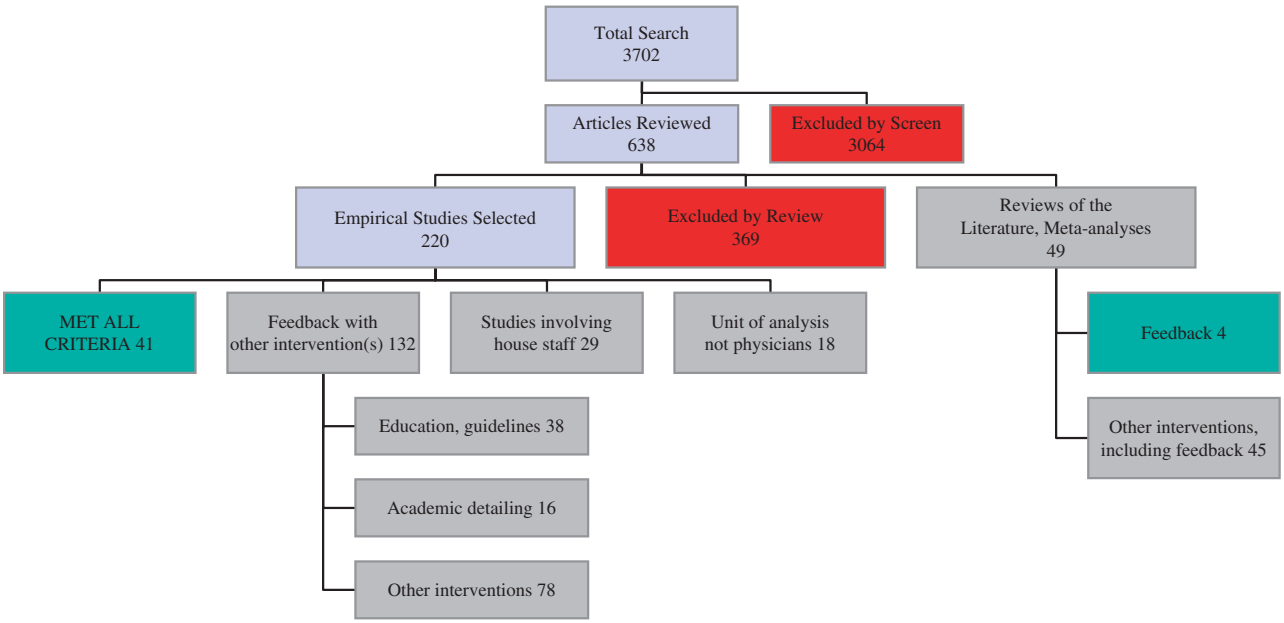


**Figure 4.** Disposition of 3702 articles systematic review of the literature on feedback and physician performance.

29 other journals. Over one-third of the studies were conducted in the United States (15), followed by Canada (7), the Netherlands (7), the United Kingdom (5), Australia (4), and Denmark (3).

Nearly all (83%) of the 41 studies involved either randomized controlled trials (17), crossover designs (10) or time-series analysis of a single cohort (7). The remaining seven studies included cross-sectional studies and parallel, non-randomized controls.

The majority (25) of studies involved only primary care physicians.

The most common source of data included medical records (13), automated medical records (12) and billing records (9). More than half of the studies (26) analysed only clinical processes, such as test-ordering behaviour. Two analysed both clinical processes and costs, and three studies used clinical processes and patient satisfaction. Only five studies considered patient health outcomes. Two examined patient satisfaction as an indicator of performance.

The feedback was provided by a department or other local administrative unit in nearly half (19) of the studies, while the next most common source was a research team (12). In the remaining studies, the feedback was provided by a professional group or government agency.

*Characteristics of feedback associated with positive effects*

Overall, 29 (70%) studies reported positive effects of feedback on performance. Positive effects were associated with two major characteristics. These included the source of the feedback and the duration of the follow-up cycle of monitoring performance and providing feedback.

The source of feedback included research teams, administrative units, professional groups and government. Of the 12 studies involving feedback provided by research teams, only half demonstrated positive effects. On the other hand, 83% of the studies of feedback provided by either administrative units or professional groups showed positive effects ($\text{chi}^2(1) = 4.9$, $p < 0.03$).

The studies that yielded significant positive effects ($t(1,37) = 2.05$, $p < 0.05$) also tended to be longer in duration. The life span of these studies averaged about two years (mean: 2.6, median and mode: 2), while the studies that did not show an effect lasted less than one year (mean: 0.9, median and mode: 0.5).

*Characteristics of feedback that had little or no effect*

Some characteristics of the feedback did not seem to be related to changes in physician performance. These included the extent of the physicians' involvement in the design of the feedback systems, whether local statistical norms or published professional standards were used for comparison in reporting performance, the relative amount of detail of feedback, whether reports of the physicians' performance were made public, and whether the feedback was written or verbal. These unexpected findings could be related to the ambiguity of these independent variables and the lack of a standard nomenclature to describe and quantify their features. For example, a statement in the methods section that team meetings were held to involve the physicians in the design of the feedback system would not routinely specify whether all physicians attended all meetings and whether all participated in the design. Likewise, as will be discussed later, the form and content of feedback reports were not described consistently across all studies.

**Studies of feedback with other interventions**

As shown in Figure 4, a total of 132 studies probed the impact of feedback when it was combined with other interventions intended to influence physician behaviour (Appendix 4 on BEME website http://www.bemecollaboration.org). However, these studies did not meet all selection criteria and were not subjected to systematic review for two reasons. First, the design and analysis failed to isolate the independent effect of feedback from the effects of the other interventions. Furthermore, as will be described below, there was inconsistency in the nomenclature used to identify similar methods and variation in the description of the details of what sometimes appeared to be similar methods. The studies were subdivided into the following three sets: education and guidelines, educational outreach visits and other interventions.

*Education and guidelines*

The first included 39 studies in which feedback was coupled with education, including group activities such as lectures, meetings and seminars or the dissemination of written materials such as self-study manuals, protocols, guidelines or written disease management practice recommendations. Overall, 24 (63%) reported a positive impact of feedback in combination with education.

*Educational outreach visits*

The second set of 16 studies combined feedback with educational outreach visits. These activities have also been referred to as academic detailing, personalized educational visits, targeted outreach visits, coaching, face-to-face instruction, technical assistance, streamlining or personal visits by pharmacists (Grimshaw *et al.*, 2001). Overall, 12 (75%) of these studies reported positive results.

*Other interventions*

The final set of 77 studies included a wide variety of other methods described variously as algorithms, care management, reminders, local opinion leaders, patient-mediated interventions, surveillance, benchmarking, group consensus processes, multidisciplinary teamwork, critical care maps, financial incentives and token rewards. Also included were multi-faceted interventions such as total quality management, process improvement facilitation and continuous quality improvement. The number of studies for each method was too small to support further analysis, and the exact procedures for most methods were not clearly described in the published articles. Overall, 62 (81%) of the studies reported positive effects of feedback in combination with the other intervention.

### Studies involving house staff

As shown in Figure 4 there were 29 studies that involved house staff in training or a mixture of faculty and house staff (see Appendix 5 on BEME website http://www.bemecollaboration.org). A total of 18 (62%) reported positive effects of feedback. The majority of the studies appeared to be experimental in nature. No studies were limited to house staff at a specific level of training that made it possible to make assumptions about level of supervision. The majority (93%) were conducted for one year or less.

### Studies in which the unit of analysis was not physicians

As shown in Figure 4, another subset of 18 studies involved units of analysis such as geographic regions, hospitals, units within hospitals, primary care sites or populations of patients (Appendix 6 on BEME website http://www.bemecollaboration.org). These studies used proxy measures of physician performance such as diabetes care, preventive care, pain management and choice of medications or surgical procedures. Although these indicators were linked to physicians' decisions, it was not always clear how the effect of the physician's actions was isolated from the effect of other factors in the healthcare units. Furthermore, it was not always clear in studies over extended periods of time whether the original cohort of physicians that received the feedback was still intact throughout the study. Nevertheless, 14 (78%) of the studies reported positive effects of feedback on the performance of the healthcare units being studied.

## Discussion

*What do the findings mean?*

The continuing widespread interest in the connection between feedback and physicians' clinical performance is affirmed by the large number of citations between 1966 and 2003. Among more than 3000 citations screened, we found 638 articles addressing the impact of feedback on physician performance.

Ultimately, we eliminated the descriptive reports, commentaries, brief reviews and studies of other health professionals and narrowed these to 220 studies containing empirical data on feedback and performance. However, only 41 of these met all selection criteria for the review in that they addressed the specific effects of feedback on physician performance. The remaining related studies involved feedback to house staff in training, the impact of feedback to physicians on large healthcare entities, or the effect of feedback combined with other interventions designed to change physicians' behaviour.

Of the 41 studies that met all selection criteria, nearly three-quarters were positive. Furthermore, about three-quarters of the remaining empirical studies that were relevant but did not meet all selection criteria reported positive effects. For example, three-quarters of the 132 studies involving feedback combined with other interventions (educational programmes, guidelines, academic detailing, local opinion leaders, reminders and a variety of multi-faceted interventions) yielded positive results. Of 29 studies involving house staff, nearly two-thirds were positive, and over three-quarters of the studies of the impact of feedback on units of analysis such as hospitals or groups of patients reported positive results.

The variation in methods, nomenclature and sampling presented formidable challenges when we tried to integrate the findings across diverse studies. The variation in outcome variables precluded any systematic analysis of effect sizes. However, the overall findings have implications for those considering the use of feedback to influence physicians. The results indicate that physicians recognize the source and purpose of the feedback. They are more likely to be influenced by a source that will continue to monitor the physicians' performance. The finding that physicians are more responsive to feedback provided over an extended period of time may be related to the nature of professional practice. Physicians are faced with a continuous stream of data and information from their practice environment as well as the broader medical community. Repetition of feedback may be needed to get their attention. It may also establish the commitment of the organization providing the feedback.

Given the finding that physicians are more likely to respond to feedback provided by certain sources over an extended period of time, an important question that follows relates to the form, content and medium of the most successful feedback. For example, one might expect stronger effects for feedback presented in carefully designed graphic reports. The clinical content of the ideal report would adjust the assessments of a physician's performance based on important clinical features of the physician's patient population, including demographics, disease stage, and comorbidities. Unfortunately, the published studies did not always provide sufficient detailed descriptions or examples of the feedback reports that would have made it possible to address questions related to format and content. Many studies referred to the feedback reports as tables, graphs, profiles or lists including individual and peer data grouped for comparison. Although the studies provided descriptions of the clinical issues related to the topic under study, the description of methods did not usually include specific information about the layout and content of the feedback reports.

There are several features of this systematic review that set it apart from previous reviews of feedback and performance. First, we excluded studies involving non-physician providers because the goal of this study was to focus attention on practising physicians. We scrutinized the studies involving feedback to house staff. We reasoned that the effect of feedback on trainees would be impossible to gauge because it would be difficult to isolate the trainees' independent decision-making from the contributions of other trainees, attending physicians and consultants who all work together as a team in academic environments.(Sicotte *et al.*, 1996). We concentrated on studies of the effect of feedback alone and did not directly address a large number of studies involving feedback in combination with a broad spectrum of other interventions. Unfortunately, the lack of a standard nomenclature for these approaches interfered with systematic analysis. Finally, we separated the studies that appeared to address the relationship between feedback and physician performance but actually involved studies of entire hospitals or patient populations rather than physicians as individuals.

The contributions of physician performance were not distinguished from the effects of other forces in the environment, and it was not always clear that the same physicians were involved in care at different points in time.

Another important feature of this review is the inclusion of non-randomized designs in addition to randomized clinical trials. We decided to consider studies using these quasi-experimental research designs because we recognized the importance of the external validity of the studies, or the degree to which the experimental conditions mirrored the real world. For example, we believe that a crossover design in which feedback is provided in varying degrees to an intact group of practising physicians over a period of several years has greater validity than a six-month experiment in which house staff are randomized into either a feedback or no feedback group. While non-randomized studies are subject to threats to validity such as the Hawthorne Effect, the finding that feedback had a greater effect over an extended period of time argues against this threat.

The international scope of this review calls attention to the challenges of synthesizing studies of physician performance across national boundaries with different models of health-care delivery. Although the number of studies analysed was too small to analyse international differences, it is important to consider the findings reported in the Cochrane Review that effects were greater when baseline compliance was low (Jamtvedt *et al.*, 2003). This factor needs to be examined more closely when attempting to make inferences regarding physicians' clinical behaviour in different healthcare settings.

The findings of this study support the use of feedback to influence the clinical performance of physicians. The results indicate that large, systematic efforts at feedback supported by an authoritative source and sustained over time have a greater chance of success than short-term interventions introduced as research studies.

The process of developing and implementing the protocol for the review and related findings has implications for readers of future studies of the impact of feedback on physician performance. When reading the literature, readers must ask the following four questions: How was physician performance measured? What was the source of the feedback and did the physicians view it as important? How long did the feedback continue? Did the study isolate the effects of feedback from the effects of other, concurrent interventions?

Finally, these results have implications for the designers of future theoretical as well as practical studies of feedback alone. More carefully controlled studies that separate the effects of feedback from other concurrent interventions will better inform policy-makers and the profession about the moderating effects of some important characteristics of feedback. Well-designed studies are needed to examine the effects of characteristics of feedback and the process of communication with physicians. Relevant characteristics include the length, format and content of feedback reports. Related characteristics include the level of physicians' involvement in the design of the feedback system, the nature of standards used to judge performance, and the mode of communicating the feedback to the physicians.

### Limitations of the review

This review of over three decades of literature has several limitations. The first is related to the lack of a standardized nomenclature to describe the two basic constructs of feedback and physicians' clinical performance. The Methods section of this review describes the steps that we took to develop search criteria to match these two constructs. The large number of citations screened suggests that the bulk of the relevant literature was identified. Nevertheless, the large volume of false hits indicates that the fit between the search criteria and the relevant literature was less than perfect.

The second limitation is related to the articles themselves. In general, the uneven quality of technical writing presented barriers when individual articles were reviewed. The lack of standardized nomenclature made it difficult to compare studies, and impossible to compare effects quantitatively across studies. The uneven level of detail in reporting methods and results interfered with the process of data extraction. Finally, the wide array of outcomes measures made it impossible to compare effect sizes across studies.

### Research agenda

This review has the potential to guide future investigators as they design studies related to feedback and physician performance. Consequently, editors and peer reviews may wish to consider the findings of this review as they evaluate future research on feedback and physicians' clinical performance.

One important finding of this study is that feedback can have a greater impact when it is provided over a period of several years. Therefore, the most effective new studies of feedback will involve interventions over an extended period of time. The most effective studies will involve experienced, licensed physicians who have completed their medical education.

The remarkably small number of studies that met all selection criteria prevented us from examining the effects of characteristics of feedback such as format and content. This indicates a need for more carefully designed, long-term studies with clearly defined samples of physicians that have completed their education. The most appropriate research designs will isolate the effects of feedback from other, concurrent interventions and their interactions. For example, there is reason to believe that combinations of feedback with educational programmes, dissemination of guidelines or educational outreach activities may strengthen the effect of feedback. There is a need for rigorous studies that isolate the independent and interaction effects of these interventions.

### Summary/Conclusions

The use of feedback to enhance physicians' clinical performance has been debated in the literature for over 30 years. We located 3702 articles by searching 11 electronic databases, by performing Internet searches and by checking the reference lists of published articles. Screening and initial attempts to review eliminated over 80% of these articles, which were either non-empirical, involved non-physicians or used the term feedback to refer to other types of interventions such as prospective reminder systems.

Of 220 remaining empirical studies on feedback, nearly two-thirds were set aside because they involved a combination of feedback and other interventions. Examples of the other interventions included educational programmes, workshops, guidelines, academic outreach visits, reminders, patient education and local opinion leaders. The designs of these studies and the reports of results made it impossible to isolate the effect of feedback from the other interventions. However, about three-quarters reported positive effects for feedback in combination with the other intervention. Another group of studies set aside involved resident physicians in training. The training level of the residents was not usually specified, and it was impossible to gauge the extent to which the residents' performance was influenced by other residents and fellows at higher levels of training and by attending physicians. Finally, we set aside another small set of studies that involved studies of hospitals or group practices in which the unit of analysis was not physicians and the outcome variable could be affected by a number of different factors in the environment in addition to physician performance.

Eventually we selected and systematically analysed 41 studies that involved baseline measurement of physicians' clinical performance, feedback and follow-up assessment. The physicians' performance, which was usually measured by reviewing medical charts or billing records, was frequently related to clinical processes or test ordering. We found that feedback provided by administrative units or professional groups was more effective than feedback provided by research teams. Furthermore, we noted that the studies lasting two years or more were more likely to have a positive effect. The results of this review of published studies indicate that feedback does have a positive effect on physicians' clinical performance. The most effective feedback is provided by a credible, authoritative source over a number of years.

**Funding/support**

This review was supported by the American Board of Internal Medicine Foundation.

**Acknowledgements**

The following individuals participated in the research study at various stages: Barbara Barzansky, PhD, American Medical Association; Daniel Duffy, MD, American Board of Internal Medicine Foundation; Brian Leas, American Board of Internal Medicine Foundation; Frank T. Leone, MD, Jefferson Medical College; Paul E. Mazmanian, PhD, School of Medicine, Virginia Commonwealth University; David B. Nash, MD, MBA, Jefferson Medical College; and Joan Sargeant MEd, Dalhousie University.

The authors would like to express their sincere appreciation to Ms Pat Lilley at the BEME Collaboration for her assistance and support throughout the review. Special gratitude is owed to Dr Barry Issenberg, Dr William McGaghie, and the other members of their team who completed the first BEME review on simulations in 2005. Their report served as an inspiration and also provided a template for this report.

The authors are especially grateful to Ms Bethany Brooks, who assisted with the formatting, editing and proofreading of multiple drafts of this report. Nevertheless, the authors assume full responsibility for the final version.

**Notes on contributors**

JON VELOSKI, MS, is Director of Medical Education Research in the Center for Research in Medical Education and Health Care at the Jefferson Medical College of Thomas Jefferson University.

JAMES BOEX, PhD, is Director of the Office of Health Services at Northeastern Ohio Universities College of Medicine.

MARGARET GRASBERGER, MS, is Senior Librarian, Information Services in the Scott Memorial Library, Jefferson Medical College of Thomas Jefferson University.

ADAM EVANS is a medical student at the Jefferson Medical College of Thomas Jefferson University.

DANIEL B. WOLFSON, MHSA, is Executive Vice President of the American Board of Internal Medicine Foundation.

**References**

BINDMAN, A.B., GRUMBACH, K., VRANIZAN, K., JAFFE, D. & OSMOND, D. (1998) Selection and exclusion of primary care physicians by managed care organizations, *Journal of the American Medical Association*, 279, pp. 675–679.

CHANG, K., SAUEREISEN, S., DLUTOWSKI, M., VELOSKI, J.J. & NASH, D.B. (1999) A cost-effective method to evaluate practice variation among primary care physicians, *Evaluation & the Health Professions*, 22, pp. 184–196.

DAVIS, D.A., THOMSON-O'BRIAN, M.A., FREEMANTLE, N., WOLF, F.M., MAZMANIAN, P. & TAYLOR-VAISEY, A. (1999) Impact of formal continuing medical education. Do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *Journal of the American Medical Association*, 282, pp. 867–874.

DAVIS, D.A., THOMSON, M.A., OXMAN, A.D. & HAYNES, R.B. (1995) Changing physician performance: a systematic review of the effect of continuing medical education strategies, *Journal of the American Medical Association*, 274, pp. 700–705.

DRANOVE, D., KESSLER, D., MCCLELLAN, M. & SATTERTHWAITE, M. (2002) *Is More Information Better? The Effects of 'Report Cards' on Health Care Providers*. NBER Working Paper No. 8697 (New York, National Bureau of Economic Research).

GONNELLA, J.S., GORAN, M.J., WILLIAMSON, J.W. & COTSONAS, N.J. (1970) Evaluation of patient care: an approach, *Journal of the American Medical Association*, 214, pp. 2040–2043.

GONNELLA, J.S., HOJAT, M., ERDMANN, J.B. & VELOSKI, J.J. (1993) Assessment measures in medical school, residency, and practice: the connections, *Academic Medicine*, 68, pp. S3–S106.

GRAY, J.D. (1996) Global rating scales in residency education. [Review]. *Academic Medicine*, 71(Suppl.), pp. S55–63.

GRIMSHAW, J.M., SHIRRAN, L., THOMAS, R., MOWATT, G., FRASER, C., BERO, L., *et al.* (2001) Changing provider behavior: an overview of systematic reviews of interventions, *Medical Care*, 39, pp. II-2–II-45.

HANCHAK, N.A. & SCHLACKMAN, N. (1995) The measurement of physician performance, *Quality Management in Health Care*, 4, pp. 1–12.

HANCHAK, N.A., SCHLACKMAN, N. & HARMON-WEISS, S. (1996) U.S. Healthcare's quality based compensation model, *Health Care Finance Review*, 17, pp. 143–159.

HARDEN, R.M., GRANT, J., BUCKLEY, G. & HART, I.R. (1999) BEME Guide No. 1: Best evidence medical education, *Medical Teacher*, 21, pp. 553–562.

ISSENBERG, S.B., MCGAGHIE, W.C., HART, I.R., MAYER, J.W., FELNER, J.M., PETRUSA, E.R., *et al.* (1999) Simulation technology for health care professional skills training and assessment, *Journal of the American Medical Association*, 282, pp. 861–866.

JAMES, B.C. (2001) Making it easy to do it right, *New England Journal of Medicine*, 345, pp. 991–993.

Jamtvedt, G., Young, J.M., Kristoffersen, D.T., Thomson O'Brien, M.A. & Oxman, A.D. (2003) Audit and feedback: effects on professional practice and health care outcomes. [Review] [85 refs]. *Cochrane Database of Systematic Reviews*. CD000259.

Mazmanian, P. & Davis, D.A. (2002) Continuing medical education and the physician as a learner: guide to the evidence, *Journal of the American Medical Association*, 288, pp. 1057–1060.

Nash, D.B., Markson, L.E., Howell, S. & Hildreth, E.A. (1993) Evaluating the competence of physicians in practice: from peer review to performance assessment, *Academic Medicine*, 68, pp. S19–S22.

Nash, I.S., Nash, D.B. & Fuster, V. (1997) Do cardiologists do it better? *Journal of the American College of Cardiology*, 29, pp. 475–478.

Nelson, E.C., Mohr, J.E., Batalden, P.B. & Plume, S.K. (1996) Improving health care, Part 1: The clinical value compass, *Journal on Quality Improvement*, 22, pp. 243–258.

Norman, G.R., Neufeld, V.R., Walsh, A., Woodward, C.A. & McConvey, G.A. (1985) Measuring physicians' performances by using simulated patients, *Journal of Medical Education*, 60, pp. 925–934.

O'Brien, M.A., Oxman, A.D., Davis, D.A., Haynes, R.B., Freemantle, N. & Harvey, E. (2003) Audit and feedback: effects on professional practice and health care outcomes, *The Cochrane Library*, pp. 1–31.

Oxman, A.D., Thomson, M.A., Davis, D.A. & Haynes, R.B. (1995) No magic bullets: a systematic review of 102 trials of interventions to improve professional practice, *Canadian Medical Association Journal*, 153, pp. 1423–1431.

Peabody, J.W., Luck, J., Glassman, P., Dresselhaus, T.R. & Lee, M. (2000) Comparison of vignettes, standardized patients, and chart abstraction, *Journal of the American Medical Association*, 283, pp. 1715–1722.

Ramsey, P.G., Wenrich, M.D., Carline, J.D., Inui, T.S., Larson, E.B. & LoGerfo, J.P. (1993) Use of peer ratings to evaluate physician performance, *Journal of the American Medical Association*, 269, pp. 1655–1660.

Rubin, H.R., Gandek, B., Rogers, W.H., Kosinski, M., McHorney, C.A. & Ware Jr, J.E. (1993) Patients' ratings of outpatient visits in different practice settings. Results from the Medical Outcomes Study, *Journal of the American Medical Association*, 270, pp. 835–840.

Sicotte, C., Pineault, R., Tilquin, C. & Contandriopoulos, A.P. (1996) The diluting effect of medical work groups on feedback efficacy in changing physician's practice, *Journal of Behavioral Medicine*, 19, pp. 367–383.

Stewart, A.L. & Ware Jr, J.E. (1992) *Measuring Functioning and Well-being*, 2nd edn (Durham, NC, Duke University Press). Ref Type: Serial (Book, Monograph).

Tu, H.T. & Reschovsky, J.D. (2002) Assessments of medical care by enrollees in for-profit and nonprofit health maintenance organizations, *New England Journal of Medicine*, 346, pp. 1288–1293.

Ware Jr, J.E. & Hays, R.D. (1988) Methods for measuring patient satisfaction with specific medical encounters, *Medical Care*, 26, pp. 393–402.

Wennberg, J.E., Blowers, L., Parker, R. & Gittelsohn, A.M. (1977) Changes in tonsillectomy rates associated with feedback and review, *Pediatrics*, 59, pp. 821–826.

Woodward, C.A., Cohen, M., Ferrier, B.M., Goldsmith, C. & Keane, D. (1989) Correlates of certification in family medicine in the billing patterns of Ontario general practitioners, *Canadian Medical Association Journal*, 141, pp. 897–904.

Woodward, C.A., Ferrier, B.M., Cohen, M. & Goldsmith, C. (1990) A comparison of the practice patterns of general practitioners and family physicians graduating from McMaster and other Ontario medical schools, *Teaching and Learning in Medicine*, 2, pp. 79–88.