Assessing the Competence of Practicing Physicians in New Zealand, Canada, and the United Kingdom: Progress and Problems

Ian St George, MD, FRACP, FRNZCGP; Tiina Kaigas, MD; Pauline McAvoy, MB, ChB, FRNZCGP, MRCGP

Members of the public expect practicing physicians to be competent. They expect poorly performing physicians to be identified and either helped or removed from practice. “Maintenance of professional standards” by continuing education does not identify the poorly performing physician; assessment of clinical performance is necessary for that. Assessment may be responsive—ie, following a complaint—or periodic, either for all physicians or for an identified high-risk group. A thorough review using a range of tools is appropriate for a responsive assessment but is not practical for periodic assessment for all. A single, valid, reliable, and practical screening tool has yet to be devised to identify physicians whose practice is suboptimal. Further, articulate commentators are concerned about the harm that too-intensive scrutiny of professional performance may cause. We conclude that high performance by all physicians throughout their careers cannot be fully ensured, but it is nonetheless the responsibility of licensing bodies to use reasonable methods to determine whether performance remains acceptable. Such methods should be shown scientifically to be accurate, valid, and reliable for practicing physicians. Such an approach is likely to encourage the agreement and cooperation of the profession. To do less risks losing the trust of the public.

(Fam Med 2004;36(3):172-7.)

Most members of the public assume that the performance of physicians is assessed regularly. Eighty percent of New Zealanders believe that the Medical Council of New Zealand (MCNZ), the national licensing body, is responsible for ensuring the competence of physicians.1 Several public (lay) members on the council of the College of Physicians and Surgeons of Ontario (Canada) were surprised to find, when they began to participate in council meetings, that there was no requirement for physicians to revalidate their competence after licensure. They had assumed there was a regular reassessment process in place for physicians, similar to that for airline pilots. During their consultation on proposals for revalidation of all physicians’ licenses, the General Medical Council (GMC) of the United Kingdom (UK) also learned of the public’s assumption that regular review of physicians was already in place.

The principal purpose of most medical regulatory jurisdictions will be similar to that of the New Zealand Medical Practitioners Act 1995: “to protect the health and safety of members of the public by prescribing or providing for mechanisms to ensure that medical practitioners are competent to practice medicine.”2 Those mechanisms have traditionally included licensing of physicians, supervision of training physicians (ie, residency training), and interval recertification of practicing physicians. Many jurisdictions have recently added the capacity to review the competence and performance of practicing physicians, rather than simply assessing their knowledge.

There is a clear intent by legislators that poorly performing physicians will be identified and then reeducated or prevented from practicing. Identifying the poor performers has never seemed a collegial activity for physicians. Yet, it is just what the public and the politicians expect of a self-regulating profession. While it may not be reasonable to assume that isolated episodes
of medical error necessarily signify incompetence or negligence, we nonetheless believe that the public and politicians are justified in expecting the profession to identify and help (or if that is not possible, remove) physicians who perform their clinical medicine poorly.

This paper describes selected current approaches to performance review of physicians. Each approach attempts to ensure the continuing competence of physicians.

Models for Continuing Competence

Learning Model

Many “maintenance of professional standards” (MOPS) programs are learning systems used for the purpose of recertification. Most are based on a continuous quality improvement (CQI) concept. CQI was borrowed by medicine after development in the manufacturing industry, in which people working alongside each other observed and oversaw each other’s work. The theoretical model shows a Gaussian curve of competence shifting to the right as a result of CQI. But how applicable to the complex and often isolated professional activities of clinical practice is such a simple model? And, if the Gaussian curve of clinical competence shifts at all, do those represented by the “low competence” tail learn and improve?

MOPS programs usually reward attendance at formal continuing medical education (CME) activities, self-assessment of learning needs, patient feedback, academic activities, and audits. However, without enabling or practice-reinforcing strategies, CME participation has little effect on changing physician behavior, assessing one’s own learning needs can be quite inaccurate, popularity with one’s patients does not necessarily signify clinical competence—we all know genial fools—and we have no evidence that publishing research, participating in research activities, or teaching have any effect on clinical competence. Audits of clinical activities are useful if they are thorough, heed denominator issues, and are followed by constructive feedback, but they frequently are not.

Thus, MOPS programs are comprised of a range of activities, many of which are at best associated with clinical competence. They do not necessarily verify it. Most importantly, MOPS does nothing to identify poorly performing physicians.

Assessment Model

In 1993, the Canadian Federation of Medical Licensing Authorities began considering its Monitoring and Enhancement of Physician Performance (MEPP) program. A three-step system was proposed: (1) screening all physicians, (2) assessment of physicians at risk or need, and (3) detailed needs assessment.

Van der Vleuten extended the metaphor of Miller’s learning pyramid and suggested that different assessment tools are appropriate for different stages of a physician’s career. Indeed, the assessment of the practicing physician emphasizes performance as well as competence and emphasizes specialty performance as well as general performance (Figure 1).

Many educational bodies have developed lists of domains of medical competence, including the US National Board of Medical Examiners. Those adopted for the purposes of competence reviews in New Zealand are shown in Table 1 as one example. The UK GMC has published Good Medical Practice, which defines the standards of practice expected of all licensed physicians and which has become the framework within which the performance of physicians is assessed.

Assessment tools have been adapted from those used in undergraduate and vocational education, then selected and refined for the specific purpose of assessing the performance of practicing physicians. These include, for example, the interview, case-based oral examinations, record reviews, peer ratings, patient satisfaction questionnaires, and observing patient encounters.

The three steps envisaged in the original MEPP discussions remain, but most regulatory jurisdictions start by assessing the performance of physicians about whom concerns have been expressed (responsive assessment). They move later to periodic assessment and screening of all physicians or at least screening of physicians at risk of poor performance.
**Responsive Assessment.** The Medical Council of New Zealand currently assesses the performance of practicing physicians only on receipt of a complaint or report of a problem. The assessment may be restricted to the domain of competence suggested by the complaint, but some jurisdictions routinely assess general competence covering all domains. The GMC in the UK introduced its Performance Assessment Procedures in 1997, and it assesses performance and competence across all domains using the framework of *Good Medical Practice.*

Poor performance is, however, as likely among physicians about whom concerns have not been expressed as it is among physicians about whom a concern has been expressed. A system confined to responsive reviews does nothing to identify the poorly performing physicians among the rest.

**Periodic Assessment for All.** Assessing all physicians at regular intervals seems an obvious solution. Further, feedback from those assessed suggests strongly that a stigma of incompetence is attached to responsive assessments (ie, assessments in response to complaints), with consequent disturbance to the self-respect, practice, and personal lives of subjects (unpublished data, Medical Council of New Zealand). This may not occur, however, with routine periodic assessments, and if all physicians were to undergo periodic assessment of performance, the stigma might be reduced.

Such an assessment, according to Norcini, should include three components: an assessment of patient outcomes, an evaluation of medical knowledge and judgment (a review of credentials), and the judgments of peers and patients. In our view, such an approach, while laudably thorough, would be prohibitively resource intensive.

A full assessment of all domains of competence for all physicians is a naive notion. Certainly we can assess all physicians for all domains of competence, reliably and validly, but the magnitude of such a venture in terms of feasibility and cost would be so enormous as to be beyond practicality.

**Screening Assessment for All.** An alternative approach would be to find a simple screening test. Such a screening assessment, as with screening for disease, should be evaluated against a set of criteria such as the UK National Screening Committee’s modification (Table 2) of Wilson and Jungner’s classic list. Various regulatory jurisdictions have tried to identify simple screening tests that might identify broader incompetence, but most have chosen a kit of screening tools.

The system in Alberta provides an example. There, the Physician Achievement Review (PAR) screens 20% of Alberta’s physicians every year, with the aim of screening every physician every 5 years. Peer ratings (eight physician colleagues and eight nonphysicians), self-assessment questionnaires, and patient questionnaires (25 patients) are used as a screening test. The domains assessed are patient care, communication and humanistic factors, clinical performance, and professional development. All the instruments have good psychometric properties and high internal consistency reliability. While Alberta physicians agree that some form of regular assessment of their competence is needed, they do not all agree that PAR does it well.

There are more than 25,000 licensed physicians in Ontario, and research is being conducted to evaluate the use of existing databases as a screening tool to assess quality of care. Indicators that appear to predict quality include physician demographics, training, practice volume, drug prescribing habits, and test ordering.

In Ontario, about 400 physicians are assessed each year by another physician practicing the same specialty of medicine. The physician reviewed is visited and given constructive feedback on office facilities, records, and patient care. When concerns are identified by this screening process, the physician is referred for an in-depth clinical assessment. Recently, a “complementary
approach” has been pilot tested, in which randomly selected physicians submit a one-page biographical sketch, a questionnaire on the practice facility, and photocopies of six specified patient records; these are assessed off site. Two hundred physicians were assessed off site; 23% resulted in a recommendation for an onsite review. Of these, 35% (ie, 16 of 200) were judged not to practice in a safe manner. Most of the physicians reviewed thought off-site review was acceptable, were satisfied, thought the report helpful, and wanted the process to continue.16

In the UK, the new revalidation program will require all physicians who wish to hold a license to submit evidence, on a 5-year cycle, of their continued fitness to practice. For the vast majority of physicians who work for the UK National Health Service, this will be achieved through annual appraisal, where evidence about performance against the headings of Good Medical Practice will be reviewed with a trained appraiser, usually a peer.17 A personal learning plan will be developed and action agreed on.18 The first physician revalidation will take place in 2005. The effect and acceptability cannot yet be assessed, but the portfolios of evidence of a pilot cohort of volunteers found that on the basis of the evidence presented, a recommendation for revalidation could be made in two-thirds of cases, with the need for additional information to be provided in the remainder.19 The quality assurance measures to be undertaken by the GMC to maintain public and professional confidence in the system have not yet been described.

The experience of all the above programs indicates that no single simple screening test has been discovered that will reliably, validly, and practically indicate poor performance. Written knowledge tests come close for reliability21 but would lack face validity for most physicians.

Screening a High-risk Group. If there is currently no single practical test that can reliably identify the poorly performing physician, can we identify a high-risk group that might be more intensively scrutinized?

In Quebec, existing databases are used to identify outliers who are subsequently assessed thoroughly.22 Prescribing, investigation, and referral patterns are routinely examined for a number of indicator conditions (eg, approach to medical treatment of angina, inappropriate use of benzodiazepines or nonsteroidal anti-inflammatory drugs in the elderly), and outliers are thus identified. These outliers then undertake a “Structured Oral Interview” consisting of a knowledge assessment using clinical cases, a skills and attitudes assessment using an objective structured clinical examination (OSCE), and a clinical reasoning assessment using script concordance tests. Outliers represent variation;23 variation exists and must be accepted in all systems. The purpose of the Quebec program is to detect unacceptable and dangerous variation.

Among Ontario general practitioners and family physicians, age, gender, membership in a professional body, and type of practice were the only predictors of poor performance (younger physicians, women, certificate holders of the College of Family Physicians of Canada, and urban physicians scored higher in the College of Physicians and Surgeons of Ontario’s Peer Assessment Committee’s grading system).24 Advancing age was shown to correlate with declining performance in longitudinal and cross-sectional studies.25

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria for Implementing a National Screening Program (UK National Screening Committee)</td>
</tr>
</tbody>
</table>

- **The condition**
  - The condition should be an important health problem.
  - The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood, and there should be a detectable risk factor, disease marker, latent period, or early symptomatic stage.
  - All the cost-effective primary prevention interventions should have been implemented as much as possible.

- **The test**
  - There should be a simple, safe, precise, and validated test.
  - The distribution of the test values in the target population should be known and a suitable cut-off level defined.

- **The treatment**
  - There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.
  - There should be an agreed-on evidence-based policy covering which individuals should be offered treatment and the appropriate treatment to be offered.
  - Clinical management of the condition and patient outcomes should be optimized by all healthcare providers prior to participation in a screening program.

- **The screening program**
  - There should be evidence from high-quality randomized controlled trials that the screening program is effective in reducing mortality or morbidity.
  - There should be evidence that the complete screening program (test, diagnostic procedures, treatment/intervention) is clinically, socially, and ethically acceptable to health professionals and the public.
  - The benefit from the screening program should outweigh the physical and psychological harm (caused by the test, diagnostic procedures, and treatment).
  - The opportunity cost of the screening program (including testing, diagnosis, and treatment) should be economically balanced in relation to expenditure on medical care as a whole.
  - There should be a plan for managing and monitoring the screening program and an agreed-on set of quality assurance standards.
  - Adequate staffing and facilities for testing, diagnosis, treatment, and program management should be available prior to the commencement of the screening program.
  - All other options for managing the condition should have been considered (eg, improving treatment, providing other services).
Because of the correlation between advanced age and lower performance, in Ontario, all physicians who turn age 70 are automatically selected for assessment, along with a random sample of all physicians. While this approach targets an at-risk group, not all would agree that this approach is acceptable. Indeed, in the New Zealand Human Rights Commission’s view, “A competence review for physicians over a certain age would be prima facie age discrimination” under part 1 of the Human Rights Act 1993 (New Zealand Human Rights Commission, May 14, 2002).

Selecting male physicians for review would face similar objections even though the Ontario experience suggests that men, as a group, perform more poorly than women. Selecting by nonmembership of professional bodies, on the other hand, may be considered an appropriate indication for review. Selecting for (rural) isolation is difficult; rurality scores have been devised for country general practitioners in New Zealand, but professional isolation is by no means always related to performance, nor are rural physicians the only ones who might practice in isolation.

Thus, while superficially attractive, the identification of a high-risk group to be singled out for review does create difficulties. Further, it would not reliably identify all of those performing poorly.

Doubts About “Ensuring” Competence

There is a groundswell of coherent, articulate, and often independent argument against the further expansion of assessments and similar activities designed to ensure that professionals are accountable. The medical philosopher Roger Neighbour wrote:

People at the top of the power pyramid—politicians, managers, regulators, the devisers of guidelines and protocols—wish, for the best reasons, to make those of us at the bottom buck our ideas up in the name of raising standards. Unfortunately, because they are too busy to understand the complexity of what we do (or too besotted with innovation), the top people resort to over-simplified rules and models, whose rigidity stifles vitality, undermines common sense, and saps motivation. On every problem, so the zeitgeist would have us believe, a solution must be imposed. But now the proliferation of solutions, each understandable in its own local context, has itself become the problem; enforcing improvement is the greatest obstacle to securing it.26

In the introduction to her 2002 Reith Lectures, Onora O’Neill asked:

Does the revolution in accountability support or possibly undermine trust?27

She expanded:

. . . beneath this admirable rhetoric, the real focus is on performance indicators chosen for ease of measurement and control rather than because they measure accurately what the quality of performance is.

She expanded:

The pursuit of ever more perfect accountability provides citizens and consumers, patients and parents with more information, more comparisons, and more complaints systems, but it also builds a culture of suspicion and low morale and may ultimately lead to professional cynicism, and then we would have grounds for public mistrust.

Plants don’t flourish when we pull them up too often to check how their roots are growing: . . . professional life may not go well if we constantly uproot [it] to demonstrate that everything is transparent and trustworthy.

One might extend her metaphor: if we feed and water the plants, if we check the parts that are obvious to ensure they are getting enough nourishment, then we can trust the roots to grow well and support even better performance above the ground. She called it intelligent accountability—“more attention to good governance and fewer fantasies about total control.”

But, others have written:

The proliferation of formal medical assessment agencies signifies that conscience and reflectivity—could they be reliably discerned—no longer offer credible guarantees of goodness in doctors.25

West recently wrote:

The (UK) proposals for ensuring physicians’ sustained good performance are wise, but their implementation should be experienced as encouraging and enabling.24

We agree. If physicians’ professionalism cannot of itself provide a sufficient guarantee of continuing good performance, then any externally applied system should facilitate rather than impose.

Conclusions

Based on our review of experiences in New Zealand, Canada, and the UK, we offer the following conclusions.

First, assessing the performance of physicians about whom concerns have been expressed cannot alone identify all poor performers, but regularly assessing the performance of all physicians in all domains of competence is impractical.

Second, while high-risk groups can be recognized from personal and practice characteristics, privacy,
human rights, and other constraints may effectively prevent their identification and assessment.

Third, screening tools for identification of poorly performing physicians are being evaluated, but there is as yet no single practical test that can accurately identify physicians whose performance is in need of more thorough assessment.

High performance by all physicians cannot be completely ensured. Nonetheless, there is a common public perception that it is the responsibility of licensing authorities to ensure that performance remains at an acceptable standard throughout a physician’s career. Authorities should increase their surveillance of performance using methods shown to be accurate, valid, and reliable for practicing physicians. Only such an approach is likely to encourage the agreement and cooperation of the profession. To do less risks losing the trust of the public.

Corresponding Author: Address correspondence to Dr St George, Medical Council of New Zealand, PO Box 11-649, Wellington, New Zealand. 64-4-3847635. Fax: 64-4-3858902. isstgeorge@mcnz.org.nz.

REFERENCES

22. Chan B. Using administrative data to evaluate quality of care. Appen-