

## Multisource feedback: a method of assessing surgical practice

Claudio Violato, Jocelyn Lockyer, Herta Fidler

New methods are needed for assessing surgeons' performance across a wide range of competencies. Violato and colleagues describe the development of a programme based on feedback from medical colleagues, coworkers, and patients for the assessment of surgeons throughout Alberta, Canada

Department of  
Community Health  
Sciences, Faculty of  
Medicine,  
University of  
Calgary, 3330  
University Drive  
NW, Calgary, AB  
T2N 4N1, Canada

Claudio Violato  
*professor*

Jocelyn Lockyer  
*associate professor*

Office of  
Continuing Medical  
Education and  
Professional  
Development,  
University of  
Calgary

Herta Fidler  
*coordinator—evaluation,  
research and special  
projects*

Correspondence to:  
C Violato  
violato@ucalgary.ca

BMJ 2003;326:546–8

The assessment and maintenance of competence of physicians has received worldwide attention,<sup>1–4</sup> partly in response to concerns about poor performance by physicians and the safety of patients<sup>5 6</sup> and partly as a result of demands for accountability to patients and funding agencies.<sup>2–4</sup> New approaches to quality improvement have resulted, as have initiatives focusing on identifying and assessing poor performance.<sup>7–9</sup>

Throughout the Western world, thinking about competence has shifted. Medical expertise and clinical decision making are increasingly recognised as only components of competence. Communication skills, interpersonal skills, collegiality, professionalism, and a demonstrated ability to continuously improve must also be considered when assessing physicians.<sup>2–4 7 8 10 11</sup>

Multisource feedback, using questionnaire data from patients, medical colleagues, and coworkers, is gaining acceptance and credibility as a means of providing primary care physicians with quality improvement data as part of an overall strategy of maintaining competence and certification.<sup>1 7 8</sup> Work with Canadian, American, and Scottish generalist physicians shows that this method is reliable, valid, and feasible.<sup>7 8 12–15</sup> Research in both industry and medicine shows that multisource feedback systems (or 360° feedback) can result in individual improvement and the adoption of new practices.<sup>12 16–18</sup>

The College of Physicians and Surgeons of Alberta, the statutory medical registration body for the province of Alberta, adopted a performance appraisal or multisource feedback system for all physicians in its jurisdiction—the physician achievement review program. This system focuses on quality improvement and operates entirely separately from the complaints and disciplinary procedures. Medical colleagues, coworkers (for example, nurses, pharmacists, and psychologists), patients, and the physician (self) all provide survey based data, which are summarised by item and category and compared with the physician's specialty group. The instruments for family physicians were psychometrically tested and adopted.<sup>7 8 19</sup> As part of its overall goal of ensuring that all physicians in the province participate in a multisource feedback process every five years, the college asked a committee of surgeons and social scientists to design and test instru-

### Summary points

The general competencies of generalist physicians (family physicians and internists) can be assessed by medical peers, coworkers, and patients

Valid and reliable multisource feedback questionnaires are a feasible means of assessing the competencies of practising surgeons in communication, interpersonal skills, collegiality, and professionalism

These quality improvement data can be used to supplement information provided through traditional sources of hospital surgical outcome data

Many surgeons in this study used the feedback to contemplate or initiate changes to their practice

ments that could be used for the surgical specialties. This paper describes the development and evaluation of a multisource feedback system for surgeons designed to assess a broad range of competencies.

### Methods

#### Development of the instrument

The committee of surgeons from the major surgical disciplines developed questionnaires that could be used for all surgical specialties. Their work was based on

- The generic performance template previously developed for family physicians (medical knowledge and skills, attitudes and behaviour, professional responsibilities, practice improvement activities, administrative skills, and personal health)<sup>8</sup>
- Copies of the instruments being used for family physicians<sup>19</sup>
- The seven roles or competencies that the Royal College of Physicians and Surgeons of Canada had identified as integral to specialty practice (medical

expert-clinical decision maker, communicator, health advocate, manager, professional, collaborator, and scientist-scholar).<sup>11</sup>

- The committee was asked to develop instrument for physician colleagues, coworkers, and patients. The self assessment instrument would use the items in the medical colleague instrument, rewritten in the first person.
- The medical colleague and self assessment questionnaires consisted of 34 items that rated the physician on a five point scale (1=among the worst; 5=among the best) or unable to assess. The items examined communication, diagnostic and treatment skills, medical records, transfer and coordination of care, respect for patients, collaboration, professionalism, ability to assess the medical literature, continuing learning, and stress management. The 19 item coworker questionnaire used the same scale and focused on communication, collaboration, respect for patients and colleagues, accessibility, and support for colleague and coworker learning. The 39 item patient questionnaire asked patients for their level of agreement with statements about selected aspects of care and used a five point scale (1=strongly disagree; 5=strongly agree). The items focused on communication, respect, the office and office staff, and information received.

### Testing the instrument

We selected a proportionate stratified (by surgical specialty) random sample of 252 surgeons. We invited up to 25 surgeons from each of vascular surgery, obstetrics and gynaecology, plastic surgery, otolaryngology, orthopaedics, general surgery, cardiovascular and thoracic surgery, neurosurgery, ophthalmology, urology, and general practice surgery to participate. Some specialties contributed fewer than 25 surgeons, and all were included.

In this type of study, generalisability (with a goal of  $Ep^2 > 0.70$ ) is a key consideration.<sup>7, 8, 12-14</sup> The generalisability coefficient ( $Ep^2$ ) is calculated to determine what modifications can be made to an instrument, by examining both the numbers of items and the numbers of raters needed to achieve data stability. Adding items and adding observers will both increase generalisability. Instruments that are too short will decrease content validity, whereas instruments that are too long produce redundancy and inefficiency. Similarly, it can be difficult to find sufficient raters able to assess someone, and quality of data is reduced. On the basis of our previous generalisability analyses of data stability ( $Ep^2 > 0.70$ ),<sup>7, 8</sup> we asked surgeons to identify eight coworkers and

eight medical colleagues to whom the survey would be sent. We instructed the surgeons to ask 25 consecutive patients to complete surveys and place them in sealed envelopes. Each surgeon completed a self assessment survey.

We enhanced content validity (sampling of appropriate content and skills) by using a table of specifications based on the list of core competency areas provided by the College of Physicians and Surgeons of Alberta and asking the working group of surgeons to ensure that each competency was covered within the instruments. The surgeon committee's endorsement of the items confirmed face validity (appearance). We did exploratory factor analyses for each instrument to ensure that the items grouped into factors consistent with the competencies identified as critical for this quality improvement initiative. We used principal component analyses using varimax factor rotation to identify internal relations in the ratings and extract the factors (that is, the group of items on each instrument that were most closely correlated with other items on the instrument). These became the factors used to develop summary scores (subscales) to provide surgeons with aggregate data. We confirmed the number of scales for each instrument on empirical grounds (eigenvalues were greater than 1) and assessed for concordance with previous empirical work.<sup>7, 8</sup> We used Cronbach's  $\alpha$  to determine internal consistency reliability. We conducted a three month follow up survey to assess whether surgeons had contemplated or initiated changes to their practice on the basis of the multisource feedback.

### Results

We received feedback about the draft instruments from 99 (15.6%) of the 635 surgeons in the province. We made relatively few adjustments.

A total of 201 surgeons provided data for the study. Participants comprised 25 general surgeons, 25 orthopaedic surgeons, 24 obstetricians and gynaecologists, 24 otolaryngologists, 24 ophthalmologists, 20 plastic surgeons, 20 urologists, 15 cardiovascular and thoracic surgeons, 13 neurosurgeons, 6 general practice surgeons, and 5 vascular surgeons. This represented 31.7% of the surgeons in the province. The table presents the response rates and percentage of the possible total for each of the instruments. The response rates for all instruments exceeded 80%. Response rates for each surgeon by type of rater were similarly high (table). For most (67 of 92) of the items on the coworker, patient, and medical colleague instruments,

Response rate and results of factor analysis

	Self	Medical colleague	Coworker	Patient
No of surgeons for whom data provided (of 201 possible)	194	198	197	185
No of respondents	194	1441	1418	4185
Percentage of total possible respondents	96.5	89.6	88.2	83.2
Mean (SD) raters per surgeon	1 (0)	7.27 (0.90)	7.2 (0.98)	22.63 (4.28)
Mean (SD) rating	4.07 (0.73)	4.50 (0.64)	4.5 (0.67)	4.5 (0.75)
Cronbach's $\alpha$ (SE)	0.97 (0.16)	0.98 (0.18)	0.95 (0.16)	0.93 (0.21)
Factors (% variance)	Clinical performance (52.7) Patient care (4.6) Communication and humanistic (4.2) Professional development (3.6)	Patient care (57.4) Communication and humanistic (4.8) Clinical performance (3.6) Professional development (3.2)	Humanistic and psychosocial (58.9) Coworker collegiality (6.9) Communication (5.1)	Information provision (57.0) Humanistic (6.8) Office staff (3.9) Personal communication (3.2) Physical office (2.6)

less than 20% of respondents reported being unable to assess the physician on that item. The factors derived from the exploratory factor analyses (table) were consistent with the intent of each of the instruments and the overall areas identified for assessment. The eigenvalues for each of the factors were greater than 1 and accounted for 69.0% of the total variance for the medical colleague instrument, 65.1% of the total variance for self, 69.8% of the total variance for coworkers, and 73.7% of the total variance for the patient instrument.

The mean ratings on all of the instruments were between 4.0 and 5.0. Overall, the surgeons rated themselves less highly than their medical colleagues, coworkers, and patients rated them.

All of the Cronbach's  $\alpha$  reliability indices were  $>0.90$ , indicating internally consistent instruments. In the three month follow up survey 144 (71.6%) of the surgeons contemplated or initiated change on the basis of the multisource feedback provided to them (range of changes 1-30; mean (SD) 12.6 (3.3)). These changes focused on communication with patients and colleagues, collaboration, office systems, and stress management.

## Discussion

Our results indicate that multisource feedback is feasible for assessing surgeon competencies for quality improvement purposes. Recruitment and response rates were high, consistent with the mandatory nature of the programme, although participation was not enforced during the development stage. Relatively few surgeons reported difficulty acquiring sufficient patient surveys or identifying sufficient numbers of coworkers and medical colleagues.

The factor analysis indicated that the instruments had theoretically meaningful and cohesive factors consistent with the overall intent of the competency areas determined by the College of Physicians and Surgeons of Alberta and consistent with previous research.<sup>7-8</sup> The high Cronbach's  $\alpha$  levels confirmed the reliability of the instruments.

These results indicate that multisource feedback systems can be used to assess key competencies such as communication skills, interpersonal skills, collegiality, medical expertise, and ability to continually learn and improve, which medical organisations and the public believe need attention.<sup>2-4 10 11</sup> Moreover, the feedback from the assessment provoked contemplation or initiation of change in many surgeons. Research on the relation between multisource feedback ratings and direct observation of surgeons' performance or results from objective structured clinical examinations, for example, could be used to confirm the validity of our method. Meanwhile, procedural competence or surgical outcomes that are routinely monitored in hospitals by annual appointment procedures, morbidity and mortality reviews, and critical incident investigations should be used in conjunction with our multisource feedback techniques to enhance performance.

Contributors: All authors contributed to the concept of the study. CV and JL wrote the paper, with contributions from HF. HF supervised the collection and preparation of the data. CV did the data analysis and is the guarantor. Tina Vonhof, project coordinator, managed the data collection. John Swinarski,

assistant registrar, College of Physicians and Surgeons of Alberta, provided ongoing direction and support for the study. Ray Lewkonja provided editorial advice.

Funding: Contract with the College of Physicians and Surgeons of Alberta.

Competing interests: None declared.

- 1 Southgate L, Hays RB, Norcini J, Mulholland H, Ayers B, Woolliscroft J, et al. Setting performance standards for medical practice: a theoretical framework. *Med Educ* 2001;35:474-81.
- 2 Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;287:226-35.
- 3 Maudsley RF, Wilson DR, Neufeld VR, Hennen BK, DeVillier MR, Wakefield J. Educating future physicians for Ontario: phase II. *Acad Med* 2000;75:113-26.
- 4 Levine AM. Medical professionalism in the new millennium: a physician charter. *Ann Intern Med* 2002;136:243-6.
- 5 Wilson RM, Harrison BT, Gibberd RW, Hamilton JW. An analysis of adverse events from the quality in Australian health care study. *Med J Aust* 1999;170:411-5.
- 6 Kohn L, Corrigan J, Donaldson M, eds. *To err is human: building a safer health care system*. Washington, DC: Institute of Medicine, 1999.
- 7 Violato C, Marini A, Toews J, Lockyer J, Fidler H. Using peers, consulting physicians, patients, co-workers and self to assess physicians. *Acad Med* 1997;70:57-63.
- 8 Hall W, Violato C, Lewkonja R, Lockyer J, Fidler H, Toews J. Assessment of physician performance in Alberta: the physician achievement review project. *CMAJ* 1999;161:52-7.
- 9 Southgate L, Cox J, David T, Hatch D, Howes A, Johnson N, et al. The assessment of poorly performing doctors: the development of the assessment programmes for the General Medical Council's performance procedures. *Med Educ* 2001;35(suppl 1):2-8.
- 10 Accreditation Council for Graduate Medical Education Outcome Project. [www.acgme.org/Outcome/](http://www.acgme.org/Outcome/) (accessed 30 Sept 2002).
- 11 Societal Needs Working Group. Skills for the new millennium: CanMEDS 2000 project. *Ann R Coll Physicians Surg Can* 1996;29:206-16.
- 12 Lipner RS, Blank LL, Leas BF, Fortna GS. The value of patient and peer ratings in recertification. *Acad Med* 2002;77(10 suppl):S64-6.
- 13 Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *JAMA* 1993;269:1655-60.
- 14 Wenrich MD, Carline ID, Giles LM, Ramsey PG. Ratings of the performances of practicing internists by hospital-based registered nurses. *Acad Med* 1993;68:680-7.
- 15 Griffin E, Sanders C, Craven D, King J. A computerized 360° feedback tool for personal and organizational development in general practice. *Health Informatics Journal* 2000;6:71-80.
- 16 Bracken DW, Timmreck CW, Church AH, eds. *The handbook of multisource feedback: the comprehensive resource for designing and implementing MSF processes*. San Francisco, CA: Jossey-Bass, 2001.
- 17 Tornow WW, London M, CCL Associates. *Maximizing the value of 360-degree feedback*. San Francisco, CA: Jossey-Bass Publishers, 1998.
- 18 Fidler H, Lockyer J, Toews J, Violato C. Changing physicians' practices: the effect of individual feedback. *Acad Med* 1999;74:702-14.
- 19 College of Physicians and Surgeons of Alberta. PAR Program. [www.par-program.org](http://www.par-program.org) (accessed 30 Sept 2002).

(Accepted 10 December 2002)

## Endpiece

### Consultant reappraisal 2500 years ago

Zoroaster sought to give an official status to primitive Iranian medicine ... Three classes of physicians are specified ... those who heal with sacred spells, those who heal with vegetable juices, and those who heal with the knife. The medical spellbinders were reckoned the ablest because they could heal without the intervention of drugs or the knife. To qualify in surgery, the candidate was required to exhibit his skill upon three heretics in succession. If all three of his patients died, he was disqualified from practice; if they lived, he was admitted to practice, but punished if a subsequent patient died from malpractice.

Garrison FH.

Persian medicine and medicine in Persia: a geometrical survey. *Bull Hist Med* 1933;1:144

Jeremy Hugh Baron,  
honorary professorial lecturer,  
Mount Sinai School of Medicine, New York