

# Methods for quality assessment in general practice

Jan-Joost Rethans, Steinar Westin\* and Richard Hays\*\*

Rethans J-J, Westin S and Hays R. Methods for quality assessment in general practice. *Family Practice* 1996; 13: 468–476.

**Background and objectives.** There is now a wide variety of methods available to general practitioners who want to engage in quality assessment, quality assurance, or quality improvement activities in their practices. These methods require some kind of performance review, or at least the collection of some performance-related data. As in traditional research, the choice of methods depends on what research questions one wants to address. This paper elaborates on some key concepts related to the choice of methods, making a distinction between whether any method actually covers performance (what a doctor does in daily practice) or competence (what a doctor is capable of doing) as well as a distinction between whether a method is direct (patient–doctor contact is observable) or is indirect.

**Methods.** An overview frame will be presented of the methods most commonly used for data collection within quality assessment. These methods are discussed on their validity, reliability, feasibility and acceptability. Direct methods aimed at recording performance are assumed to hold the highest validity, but practical, economic and logistic factors may favour less ambitious methods for audit or quality improvement activities.

**Conclusions.** One crucial element in all methods is creating a set of empirical data, as a basis for comparisons, reflection, dialogue and discussions among colleagues.

**Keywords.** General practice, quality of care, methods, quality assessment.

## Introduction

Quality assurance in general practice is defined as “a continuous process of planned activities, based on performance review and setting explicit targets for good clinical practice with the aim of improving the actual quality of patient care”.<sup>1</sup> An important part of this definition is the statement “based on performance review”. This requires that, for quality improvement, data have to be collected about performance. When practitioners or researchers wish to collect these data for quality assessment they face several problems which have to be solved before the process of quality improvement can be successful. Among issues to be considered are: what sort of data have to be collected and by which methods? If one does not address these problems in the right way, there is a risk of at least partial failure of the quality improvement project involved.

But how precisely does one decide what sort of data and which methods have to be used?

Based on experience within the Centre for Quality of Care Research of the Department of General Practice in Maastricht (The Netherlands), the Department of Community Medicine and General Practice in Trondheim (Norway) and the Centre of General Practice and Rural Health in Townsville (Australia), this paper aims to give readers a frame which will assist in making these judgements in a rational and logical way. Since many methods will be reviewed here, we have included as many relevant papers as seems necessary without making it a formal review paper.

After first emphasizing that the proper research question is the most important and first step to make, an overview frame will be presented of the methods most commonly used for data collection within quality assessment. The frame takes a doctor-orientated rather than a patient-orientated approach, because the focus of this paper is what general practitioners (GPs) do. However, there may be aspects of a patient-orientated approach (e.g. WONCA-COOP charts) which add to the value of quality assessment of particular conditions. The frame presented is based upon two distinctions: the distinctions between competence and performance and between direct and indirect methods. The validity, reliability,

---

Received 26 April 1996; Accepted 17 May 1996.

Centre for Research on Quality of Care in General Practice, University of Limburg, PO Box 616, 6200 MD Maastricht, The Netherlands, \*Department of Community Medicine and General Practice, University Medical Center, Trondheim, Norway and \*\*Center for General Practice and Rural Health, The University of Queensland, Townsville, Australia.

feasibility and acceptability of the methods will be discussed. Although many methods may also be appropriate in other health care settings, the paper will primarily focus on quality assessment within general practice.

## The proper question

Suppose you are an 'expert' in quality assessment methods and you are consulted by a GP of a group practice of 10 000 patients (five GPs) in an inner city, or a large group of GPs in a rural area. This practice/group recently decided that they would like to start a quality improvement project. The question this practice would like you to answer is: "what is the best method to gather data about the quality of care given to our angina pectoris patients?" Will it be possible to answer this question in a relatively straightforward way?

Confronted with this question several issues arise. What does "quality of care given to patients with angina" mean? Is the practice interested in the actual functional status of these patients? Is the practice interested in the quality of prescriptions given to the angina patients, or in quality of referral letters, the quality of the content of the consultations or the quality of advice about leisure activities of these patients? As long as the practice has not made a clear decision about what quality they precisely are interested in, it will be impossible to advise on what is the best method to gather data about the quality of care. For example, to assess the quality of referral letters other methods have to be used (copies of referral letters) than to assess quality of prescriptions for angina patients (data from pharmacies).

Hence the first step in deciding what sort of data and which methods for quality assessment has to be taken is to define the proper research question. There is always a clear and sound link between what one intends to measure and the method one would like to use. There is no 'one and only ideal method'. All methods for quality assessment have their strengths and weaknesses and each one has a particular domain which fits best the problem to be addressed.

## Competence or performance

The next step in the decision to select a proper method for assessing quality of care is to think and decide about whether data will be collected in a competence or in a performance setting. Competence has been defined as "what a doctor is capable of doing" and performance as "what a doctor does in his day-to-day practice".<sup>2</sup> These concepts refer to the setting of the focus of interest and research has shown that these concepts indeed refer to different data sets with different qualities.<sup>3</sup> In general, performance measures are more appropriate

for experienced GPs, whereas competence measures may be more acceptable in undergraduate and vocational training.

Referring again to the practice with the interest in quality of care for their angina pectoris patients, suppose that the group decides to focus on the quality of prescriptions for these patients. Taking into account the concepts of competence and performance the next question to them would be: is the interest in the quality of prescriptions primarily based on real performance-based data (for example, for feedback reasons) or is it on what sort of prescriptions the doctors know of, the ranking of these prescriptions in terms of first and secondly preferable drugs or related issues? Performance-based data have to be gathered by performance-based methods, for example copies of original prescriptions, whereas data on knowledge about the prescriptions could easily be gathered by sending the participating GPs a short knowledge test on anti-angina prescriptions or by asking them to answer written case vignettes (sometimes referred to as patient management problems), both methods which are competence orientated. Now the practice could reply that they are interested in both performance and competence. In that case a combination of methods will have to be used. Certainly there is a tendency now to prefer performance-based data, because research on feedback, for example, has shown that performance-based data are better than competence data in changing physicians' behaviour.<sup>4</sup> However, it is stressed here that the actual decision about what setting to use originates from the primary research question.

## Direct and indirect methods

Methods for quality assessment can be divided into direct and indirect methods.<sup>2</sup> With a direct method the researcher can see or hear a physician dealing with patients or with examiners. Use of video recording is an example of a direct method. With indirect methods direct observation of patients and doctors is not done, but assessment is done by, for example, the use of written case vignettes. So, again referring to the "angina pectoris practice", suppose this practice has decided that they would like to have performance-based data upon the quality of prescriptions. The following question then to answer would be: do you want to assess it by a direct or an indirect method?

The answer to this question is not as easy as the one on competence or performance. In general one could argue that direct methods are to be preferred to indirect methods, since they are more valid with regard to what is going on during a consultation with a patient. However, also in general terms, direct methods are logistically more difficult to administer and therefore their use may be limited to relatively small groups.

Again the most important issue will be what the primary research question is. If the practice is predominantly interested in actual prescribing behaviour with a specific case of an angina pectoris patient, sending standardized patients presenting as angina pectoris patients into the practices would be a good, although expensive choice.<sup>5</sup> Copies of prescriptions would do just as well as a first step, and then, when reasons for actual prescribing behaviour have become more clear, standardized patients could act as the ultimate feedback test method.

So, to answer whether a direct or an indirect method is to be used, issues of validity, reliability, feasibility and acceptability have to be valued.

In an attempt to visualize the questions to be solved in selecting the proper method, Table 1 shows an overview of the most commonly used methods within quality assessment. The overview makes two distinctions: between competence and performance and between direct and indirect methods. In the following we will comment on most of these methods, firstly the competence methods, then the performance methods.

## Competence methods

### *Written examinations*

Of all methods used to assess competence this method probably is the mostly used and best documented one.<sup>6</sup> Several formats exist, ranging from multiple-choice questions to simple correct/incorrect answer options on literature-based statements all having in common that they primarily assess knowledge.<sup>7,8</sup> Such methods are easy to administer to large groups of GPs, are reliable if they contain enough questions and are also most often acceptable for GPs. If one intends to measure only knowledge, the methods are also valid. However one has to recall that 'to know how' and 'to do' are different issues, so basic knowledge tests have little place in quality assessment of experienced GPs.<sup>9</sup>

### *Interviews*

Interviews obtain information on opinions or perceptions about how someone would perform something. The interviews can be made with the GPs, their practice staff or with patients. However, evidence exists that what GPs say they would do not necessarily corresponds to that they would do in real life.<sup>10,11</sup> This implies that interviews with physicians and their staff with the purpose of finding out what is going on are valid with regard to competence but not necessarily with regard to performance. If the interviewer asks about opinions of medical staff, then the result may be more valid for performance. There is ample literature on interviews with patients; however, it is still uncertain what interviews with patients actually assess.<sup>12</sup> This type of research has so far mostly focused on factors related to patient-satisfaction. Since most results show that

patients are satisfied with 80–90% of the care given, regardless of its actual quality, one may wonder if answers given indeed reflect reality or dependency of patients on their physicians. Recent research in Australia has discovered that patient perceptions of quality relate more to issues of access, cost, duration of consultations and the quality of explanations provided than to real understanding of how accurate the diagnosis and management is.<sup>13</sup> However, efforts should be made within general practice to gather information from patients (questionnaires, focus groups, interviews) as consumer opinion is important.<sup>12</sup> One drawback of all interviews is that they are difficult to administer to large groups of physicians or patients.

### *Written cases (vignettes)*

In an attempt to overcome simple knowledge assessments this format aims to assess higher level skills, such as 'problem-solving skills'. A wide range of methods exists here: written simulations,<sup>14</sup> Patient Management Problems,<sup>15,16</sup> and more recently the Simulation of Initial Medical Problem-solving (SIMP),<sup>17</sup> 'key-feature cases'<sup>18</sup> and extended matching questions<sup>19</sup> offer some promise that at least applied knowledge can be tested. Computer simulations for general practice have been developed, but few studies with these still exist.<sup>20</sup> Although written vignettes are more valid than other written tests, they still assess mainly knowledge and thus competence. For large groups of physicians these instruments are acceptable, fairly simple to organize and they have reasonable reliability and validity with regard to quality assessment of competence.

### *Oral examinations (with and without patients)*

This competence method has been applauded because it allows candidates to be tested 'in-depth'.<sup>21</sup> However the method has been criticized since reliability tends to be much lower than with other methods.<sup>22</sup> An advantage is that real patients could be used, thereby strengthening its validity as a direct method; however, the setting remains competence-orientated. If large groups of GPs have to be tested the method is not easy to administer.

### *Surveys*

Much of what has been stated under interviews can be repeated here if surveys are viewed upon as 'written interviews'. When assessing opinions they are valid, but not when they address performance. However, the method is extremely useful for assessing large groups of GPs.

### *Audio/video in test situations*

There is a large experience with both these methods, firstly only for educational purposes,<sup>23</sup> but also for formative assessment of clinical competence,<sup>24</sup> as well

as for selective research purposes. In addition to audio tapes, video tapes offer the possibility to assess the body-language of candidates, and—with some limitations—the performance of physical examinations. Usually the tapes are observed with other assessment-instruments to assess quality of (part of) consultations.<sup>25</sup> An advantage of these methods is that tapes can be kept and used for later assessments. Research has shown that inter-observer reliability needs not to be a problem given enough observer training.<sup>25</sup> If one aims to assess competence in dealing with patients this method is one of the best, also for feedback purposes. The method is acceptable, but demands for technically equipped consultation rooms pose some logistical problems for the number of physicians that can participate.

#### *Observation with patient in a test situation*

See 'Oral examinations'. Also sometimes instead of patients, mannequins ('puppets') may be used for assessment of specific practical skills.

#### *Standardized patients in a test*

Since the start of the concept 'standardized patient', this method has been the subject of much interest, mostly because of its validity.<sup>26</sup> First used only for educational purposes, the method is nowadays used also for selective purposes, although still only in educational settings. Assessment of consultations with standardized patients can be either by external observers or by the standardized patients themselves. If the patients do the assessment, this is less costly compared with academic paid observers. For assessment of quality of care within a competence setting this method is one of the best, since it is direct, valid (for competence) and reliable.<sup>27</sup> If external observers, rather than the standardized patients themselves, assess the quality, specially equipped test-rooms are necessary.

## Performance methods

#### *Self-registration (recording) in practice*

The assumption underlying this method is that doctors are able to record their own activities in a reliable and valid way using specifically designed instruments. However, no conclusive studies exist which confirm this.<sup>28</sup> The method seems attractive because the logistics are easy.<sup>29</sup> What is often forgotten, however, is that GPs should be trained in order to be able to record in a reliable way, and experience indicates this is seldomly done properly. Therefore, care should be taken before data collected on content of consultations with this method are considered entirely valid and reliable. Another issue is whether physicians who record their own activities still perform in the way they always do, or if they behave more as if in a competence situation. If the recording is done only with respect to the

number of visits of special types of patients (for instance diabetics), and not so much with the content of a consultation is it better to speak of "practice activity analysis".

#### *Clinical notes (chart review)*

This is probably among the most frequently used methods to assess quality of health care, at least in hospital settings.<sup>30</sup> Some national working bodies for general practice have agreed on standards for record keeping, 'minimum data base sets', that includes the patient's identity (name, address, telephone number, marital status, occupation, the patient's medical problem list, relevant medical history and medication taken be recorded somewhere in the notes, as well as details about the last consultation).<sup>31</sup> A scoring system for analysing records has been validated and found to be reliable.<sup>32</sup> However, research has shown that data in clinical records are best to be divided in two parts. The first is data concerning patients' identity and problem list. These data can be regarded as valid and reliable, most items being present in records up to 100%. The second is data about the real content of the consultation, or 'what is going on'. In one study data in records only reflected about 32% of the content of consultations: so more is actually done than recorded, which has implications for the validity of the findings.<sup>33</sup> However, given this drawback, chart review is still a very useful method to compare GPs on their recording abilities, and as such the method reflects performance or at least performance of recording. The method is acceptable to participating GPs and the logistics will be facilitated by the introduction of personal computers in doctors' offices.<sup>32,34</sup>

#### *(Copies of) prescriptions*

This method is easy to administer, as all that is needed is either a photocopying machine in the practice or 'double-sheet prescription leaflets'. Prescriptions can even be photocopied by hand.<sup>35</sup> There seems to be no problems with acceptability, reliability and validity. However, the strength of this method is also its drawback: it limits the audit only to prescriptions and tells nothing about other aspects of performance. Only when the purpose of a quality assessment project is precisely in the area of prescription might this method be the one of choice. The methodology gives the possibility for a quantitative data collection, but data might also be used for qualitative feedback purposes, for instance within vocational training.

#### *(Copies of) referral letters*

A referral letter can be viewed upon as a (provisional) outcome of a consultation between a patient and a GP, and as such referrals are well suited for quality of care research, both from a quantitative and qualitative aspect.<sup>36</sup> One of the largest quantitative studies with

referrals is a European study in which 1548 GPs of 15 different countries participated to register 44 000 referrals.<sup>37</sup> Although referrals have been researched extensively, there is controversy with regard to their reliability and validity; that is to what extent letters can be viewed upon as quality indicators of doctors' performance. In other words, what quality of GPs do referral letters reflect?<sup>38,39</sup> There appear to be no studies which explore how many referral letters are necessary to form an opinion about the quality of referral letters of an individual GP. In conclusion, the method of referral letters assesses performance and as such it is very useful for feedback purposes, for example in peer groups. It is also an acceptable and feasible method, but care should be taken not to interpret referral letters beyond the actual referral process.

#### *Data from pharmacy*

For general practice this is a relatively new and attractive method. It depends on well-defined geographical limits to the pharmacy service area, which is easily achieved in rural general practice, for instance in Norway.<sup>40</sup> However, data from pharmacies reflect number and content of prescriptions only (see earlier for comments), and as such it is merely a more reliable system than copies of prescriptions (at least if the pharmacies work with computerized systems). In the future this method may be valuable for epidemiological surveys of side-effects of drugs.<sup>41</sup> Use of data from pharmacies has been used to detect risks of potential side-effects of drugs.<sup>42</sup>

#### *Data from insurance agencies*

Figures from official bodies are often taken as valid figures about the system of health care. Whether these data are suitable for quality of care assessment of GPs is highly questionable, since most data on patient encounters, diagnoses, time spent etc. have been collected for administrative and economic purposes, rather than for clinical reviews. Reliability, validity and also acceptability needs to be researched before such data can be employed for quality assessment. Again, careful thinking should be done on what the purpose of a quality assessment project in general practice is before using data from insurance bodies.

#### *Telephone-traffic registration*

Patients often complain about problems in finding the doctor when needed. So there are good reasons for investigating to what extent the GPs can actually be reached by telephone. Modern databased teletraffic registration allows for automatic recordings of all incoming telephones to any one practice, their numbers, at what times, and whether or not they are answered, undoubtedly a performance-based recording. In Norway, some pilot projects revealed a much lower

answering rate than the doctors presumed, allowing for marked improvements.<sup>43</sup>

#### *Critical incident review*

Basically with this technique a (large) number of physicians is asked to identify, from their own experience, incidents which appeared to have either a positive or a negative effect on the quality of care delivered to the patient.<sup>44,45</sup> Usually these descriptions are then classified into categories of what constitutes good and bad care. The method is especially worthwhile when one wishes to define the domain of a medical speciality, for example for examination purposes. Also it is easy with this method to reach a large number of physicians ensuring both reliability and validity. A drawback of the method is that no attempt is made to describe adequate care, only the extremes (good and bad care) are given. The method may, however, be useful as a method for defining tracer variables of quality.<sup>46,47</sup>

#### *Data from hospitals (e.g. from labs)*

As with data from insurance bodies, these data are in general collected for other purposes than for quality assessment of general practice. This implies that, when routine data are being used, both validity and reliability of these data for general practice needs to be considered. However, if data are collected within a specific service for general practice, for example a hospital laboratory which serves the GPs in its area, then such data can be regarded as valid and reliable. Research has shown that feedback of laboratory data counts can actually reduce the number of tests requested by GPs.<sup>4</sup> It should be kept in mind, however, that these data reflect only a part of the domain of physicians' performance. If co-operation with hospitals exists, the method is feasible for large numbers of physicians and well suited for audit activities.

#### *Practice Activity Analysis*

If activities (e.g. number of consultations, consultation time, number of requests for contraceptive medication) with regard to consultations are collected and analysed in a systematic way, one speaks of 'practice activity analysis'.<sup>48</sup> The method closely resembles what is described under 'self registration in practice'. If conducted properly, reliability and validity of Practice Activity Analysis is better than for self-registration, as the GPs themselves are not the only data recorders. Current registration networks can be regarded as a special form of Practice Activity Analysis, with high reliability and validity.<sup>49</sup> Evidence exists that these networks offer an excellent opportunity for quality assessment projects.<sup>50</sup> The method is acceptable and a large number of GPs can participate. It may well be that this is one of the methods which is very suitable for large scale quality assessment projects. However, it should be kept in mind that only indirect data

about the actual consultations can be collected in this way.

#### *Trained practice surveyors*

Models of standards assessment in practices in Canada, Australia, The Netherlands and the UK involve sending trained practice surveyors to inspect practices.<sup>51-53</sup> Surveyors are usually other GPs, but may be practice nurses or practice managers. The most common method is for trained surveyors to visit by arrangement with the practice and then combine a number of performance measures as part of a profile assessment. Individual measures include direct observation of staff and facilities, reviewing practice activity analyses and patient surveys, reviewing patient records, and discussing clinical management issues with the doctors. This method is acceptable to volunteer practices, but may be so resource intensive that its application is restricted to a relatively small number of practices and practitioners.

#### *Audio/video in practice*

Essentially this method is the same as described under competence methods, but now in a performance situation. For assessment purposes this method is very attractive since the method is direct and records performance. Most experience with the method in real practice is limited to trainees in vocational training,<sup>24,54</sup> although small-scale studies with qualified GPs have also been reported.<sup>55</sup> Again, reliability of the method is good provided that the observers are well trained. An essential issue for quality assessment is whether doctors, if their consultations are taped, do the same as when they are not taped. That is, are taped consultations valid for everyday performance? With an experimental design, Pringle and co-workers<sup>55</sup> concluded that video tapes can be valid, provided that the participating doctors do not select the consultations, but allow an external reviewer to select a sample from a larger number of recorded consultations. Formal written consent is required for this type of assessment, and is usually given. This method suits well when details about content of consultations are to be assessed for quality purposes. A problem with the method is that it is difficult to compare GPs, except with regard to communication skills, since each doctor meets different patients. Concerning logistics several problems have still to be solved, for example when it is advisable, for ethical reasons, to record performance of physical examinations. This method is particularly suitable for a small number of participants, as in peer review or educational groups.

#### *Observation in practice*

Two kinds of this method exist. First, GPs may be observed during consultations by observers (for example, colleagues) present in the same room, and secondly by letting colleagues observe the practice facilities

(management) of a surgery.<sup>56-58</sup> GPs seem to welcome this type of quality assessment, and this may be because the method is direct and addresses aspects of real practice. Although the method is performance based, there may be questions whether physicians who are aware of being visited and audited indeed show their real performance. It may be better to speak of auditing "competence behaviour in a performance setting". Reliability issues are as for video-recording; when trained observers are used reliability is good. Since observation demands observers, either peers or trained laymen, and is time-consuming, it is a method with considerable logistical problems. Nevertheless, it has found widespread application in assessment activities in Britain and Australia, particularly in vocational training.<sup>56,59</sup>

#### *Standardized patients in practice*

Preparing and sending standardized patients into GPs' offices is a new and powerful direct method for quality assessment, reliable, valid and, with proper preparations, also acceptable.<sup>3,60</sup> Since the participating GPs are unaware of when they are assessed (upon a general consent to participate), the validity of this method with regard to performance is very good. Recently it has been shown that the method can be introduced even into a 'list system' of general practice.<sup>61</sup> Logistically the method is not easy initially, but once a routine is established, barriers can be overcome. The method is not suitable for quality assessment of large groups of physicians, and is best used in groups of up to 30 physicians.<sup>60</sup>

TABLE 1 *Methods for assessment of quality of care*

Competence 'what a doctor is capable of doing'	Performance 'what a doctor does in daily practice'
Indirect methods	
Written exams	Self registration in practice
Interviews	Clinical notes (chart review)
Written cases (vignettes)	(Copies of) prescriptions
Oral exams (without patients)	(Copies of) referral letters
Surveys	Data from pharmacies
	Data from insurance agencies
	Telephone-traffic registration
	Critical incident review
	Data from hospitals (e.g. labs)
	Practice activity analysis
	Trained practice surveyors
Direct methods	
Oral exams (with patients)	
Audio/video in test situation	Audio/video in practice
Observation with patient in a test	Observation in practice
Standardized patient in a test	Standardized patient in practice

## Discussion

Although textbooks on methodology of quality improvement exist, no book—at least to our knowledge—offers the GP reader a comprehensive and yet fairly extensive table of methods. This paper attempts to fill this gap. Our two-dimensional table and our brief comments could serve as a map, by which the merits and weaknesses of most methods within quality assessment could be judged and discussed.

Although extensive, Table 1 does not pretend to be an exhaustive summation of methods. Most methods mentioned originate from a 'doctor-orientated' approach. In the context of quality assessment, one could also choose to apply 'patient-orientated' methods, such as use of patients' outcome instruments (e.g. the WONCA-COOP charts, SF-35). Also some methods exist which constitute a combination of methods. For example, within the assessment of quality of competence the method of Objective Structured Clinical Examination (OSCE) is a commonly used one.<sup>62</sup> However, this OSCE is a combination of 'observation with a patient in a test', 'standardized patients in a test' and other direct competence methods.

The social context within which quality improvement activities take place is crucial to its outcomes and its acceptability among those participating.<sup>1</sup> Most GPs are seldomly in that 'ideal' situation sketched in our initial examples, with blank sheets and totally open minds. Usually there are some local initiatives or research projects in which they are invited to take part, there are resources provided for certain kinds of activities, or there are educational or recertification requirements which should be met. The underlying purpose could be sanctionary, 'weeding out the bad apples', or part of a process of continuing quality improvement, 'polishing all the apples'.<sup>63</sup>

In our experience, creating an active participation based on supportive peer reviews and discussions, rather than external 'control', provides the best environment for learning and unbiased self-reflection. This may not in itself lead to major changes in doctors' behaviours, but it allows for them. Giving credits for quality assessment activities within postgraduate and CME training programmes seems to be one effective way of promoting quality improving activities, and is implemented in both Norway and The Netherlands, and even earlier in Australia.<sup>64</sup> The recent fundholding reform in Britain has linked funding to various quality activities, whose merits have been debated.<sup>64</sup> In any case, this trend calls for a more widespread knowledge of quality assessment methods, to which we hope this paper can contribute.

## References

<sup>1</sup> Marwick J, Grol R, Borgiel A. *Quality assurance for family doctors*. Jolimont, Australia: World Organization of National

- Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA), 1992.
- <sup>2</sup> Rethans JJ, van Leeuwen Y, Drop M, van der Vleuten C, Sturmans F. Competence and performance: two different concepts in the assessment of quality of medical care. *Fam Pract* 1990; 7: 168–174.
  - <sup>3</sup> Rethans JJ, Sturmans F, Drop R, Vleuten vd C, Hobus P. Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. *Br Med J* 1991; 303: 1377–1380.
  - <sup>4</sup> Winkens RAG, Pop P, Bugter Maessen AMA *et al.* Randomised controlled trial of routine individual feedback to improve rationality and reduce number of test requests. *Lancet* 1995; 345: 498–502.
  - <sup>5</sup> Saebö L, Rethans JJ, Johannessen T. Assessment of performance of general practitioners with angina pectoris. A study with standardized (simulated) patients in actual practice. (Submitted for publication.)
  - <sup>6</sup> Newble D, Jolly B, Wakeford R (eds). *The certification and recertification of doctors. Issues in the assessment of clinical competence*. Cambridge: Cambridge University Press, 1994.
  - <sup>7</sup> Norcini J, Swanson D, Grosso L, Shea J, Webster G. Reliability, validity and efficiency of multiple choice questions and patient management problem item formats in the assessment of physician competence. *Med Educ* 1985; 19: 238–247.
  - <sup>8</sup> Vleuten vd C, Verwijnen M, Wijnen W. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teacher* 1996 (in press).
  - <sup>9</sup> Miller GE. The assessment of clinical skills. *Acad Med* 1990; 65: S63–S67.
  - <sup>10</sup> Jones TV, Gerrity MS, Earp J. Written case simulations: do they predict physicians' behaviour? *J Clin Epidemiol* 1990; 43: 805–815.
  - <sup>11</sup> Rethans JJ, Boven van CPA. Simulated patients in general practice: a different look at the consultation. *Br Med J* 1987; 294: 809–812.
  - <sup>12</sup> Wensing M, Grol R, Smits A. Quality judgement by patients on general practice care: a literature analysis. *Soc Sci Med* 1994; 38: 45–53.
  - <sup>13</sup> Anonymous. Consumer Health Forum, Integrating consumers views about quality in general practice. AGPS, Canberra, March 1996.
  - <sup>14</sup> McGuire C. Written methods for assessing clinical competence. In Hart IR, Harden RM (eds). *Further developments in assessing clinical competence*. Montreal: Can-Heal Publications, 1987: 44–58.
  - <sup>15</sup> McGuire C, Solomon C. *Construction and use of written simulations*. Chicago: The Psychological Corporation, 1976.
  - <sup>16</sup> Westin S, Jacobsen G, Østensen AI. Patient management problems in general practice: Using a medical journal for self-assessment exercises. *Scand J Prim Health Care* 1983; 1: 63–71.
  - <sup>17</sup> De Graaf E, Post G, Drop M. Validation of a new measure of clinical problem-solving. *Med Educ* 1987; 21: 213–218.
  - <sup>18</sup> Bordage G, Page G. An alternative approach to PMPs; The 'key-features' concept. In IR Hart, RM Harden (eds). *Further developments in assessing clinical competence*. Montreal: Can-Heal Publications, 1987: 59–72.
  - <sup>19</sup> Case SM, Swanson DB. Extended-matching items: A practical alternative to free-response questions. *Teaching and Learning in Medicine* 1993; 5: 107–115.
  - <sup>20</sup> Kock CA de, Stoffers HEJH, Root op't JMH. De ontwikkeling van een computergestuurde casusgerichte toets voor het praktisch medisch onderwijs in de huisartsgeneeskunde (The development of computer based testing for assessment of competence in general practice). In Pols J *et al.* *Gezond Onderwijs 4*. Bohn Stafleu van Loghum, Houten/Zaventem, 1995: 103–109.

- 21 Muzzin L, Hart L. Oral examinations. In Neufeld V, Norman G (eds). *Assessing clinical competence*. New York: Springer, 1985: 71–93.
- 22 Vleuten vd C, Newble D. Methods of assessment in certification. In Newble D, Jolly B, Wakeford R (eds). *The certification and recertification of doctors. Issues in the assessment of clinical competence*. Cambridge: Cambridge University Press, 1994: 105–125.
- 23 Premi J. An assessment of 15 years experience in using videotape review in a family practice residency. *Acad Med* 1991; **66**: 56–57.
- 24 Hays RB, Jones BF, Adkins PB, McKain PJ. Analysis of videotaped consultations to certify competence. *Med J Aust* 1990; **152**: 609–111.
- 25 Thiel v J, Kraan H, Vleuten vd C. Reliability and feasibility of measuring medical interviewing skills: the revised Maastricht History-taking and Advice Checklist. *Med Educ* 1991; **25**: 224–229.
- 26 Barrows HS. The programmed patient: a technique for appraising student performance in clinical neurology. *J Med Educ* 1964; **39**: 802–804.
- 27 Vleuten vd CPM, Swanson DB. Assessment of skills with standardized patients: state of the art. *Teaching and Learning in Medicine* 1990; **2**: 58–76.
- 28 Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991; **66**: 762–769.
- 29 Anonymous. The Birmingham Research Unit of the Royal College of General Practitioners. Self evaluation in general practice. *J R Coll Gen Pract* 1977; **27**: 265–270.
- 30 Shaw CD. Criterion based audit. *Br Med J* 1990; **300**: 649–651.
- 31 Royal College of General Practitioners. *Evidence to the Royal Commission on the National Health Service*. Policy statement 1. London: RCGP, 1985.
- 32 Del Mar C, Lowe JB, Adkins P, Arnold E. What is the quality of general practitioner records in Australia? *Australian Family Physician: REASON* 1996; **25** (supplement): S21–S25.
- 33 Rethans JJ, Martin E, Metsemakers J. To what extent do clinical notes by general practitioners reflect actual medical performance? A study using simulated patients. *Br J Gen Pract* 1994; **44**: 153–156.
- 34 Grimsmo A, Grimstad S, Lilleholt O, Snoen S, Storset B. Informasjon til kvalitetssikring og selvevaluering i allmennpraksis (Evaluating one's own practice through information from computer-based journals). *Tidsskr Nor Laegeforen* 1994; **114**: 1983–1987.
- 35 Rokstad K *et al.* Can drug treatment be improved by feedback on prescribing profiles combined with therapeutic recommendations? *J Clin Epid* 1995; **48**: 1061–1068.
- 36 Hoftvedt BO, Mjell J. Referrals: Peer review as continuing medical education. *Teaching and Learning in Medicine* 1993; **5**: 234–237.
- 37 Fleming DM. *The European study of referrals from primary to secondary care*. Amsterdam: Thesis Publications, 1993.
- 38 Westerman RF, Hull FM, Bezemer PD. A study of communication between general practitioners and specialists. *Br J Gen Pract* 1990; **40**: 445–449.
- 39 Coulter A, Noone A, Goldacre M. General practitioners' referrals to specialist outpatient clinics. *Br Med J* 1989; **299**: 722–724.
- 40 Krokstad S, Gjelsvik PA, Mjelva E. Reduksjon i forskrivning av B-preparater i allmennpraksis (Reduction in prescription of benzodiazepines and strong analgetics psychotropic in general practice). *Tidsskr Nor Laegeforen* 1995; **115**: 3634–3637.
- 41 Petri JL. *The prescription drug history in pharmaco-epidemiology*. Maastricht: Universitaire Pers Maastricht, 1992 (Thesis).
- 42 Petri H, Leufkens H, Silkens R, van Hessen P, Urquhart J. Rapid method for estimating the risk of acutely controversial side effect of prescription drugs. *J Clin Epidem* 1990; **43**: 433–439.
- 43 Nordby K. *Telefontilgjengelighet i allmennpraksis (Telephone availability in general practice)*. Oslo: The Norwegian Medical Association, 1994.
- 44 Norman GR. Defining competence: a methodological review. In Neufeld V, Norman G (eds). *Assessing clinical competence*. New York: Springer, 1985: 15–33.
- 45 Diamond MR, Kamien M, Sim MG, Davis J. A critical incident study of general practice trainees in their basic general practice term. *Med J Aust* 1995; **162**: 321–324.
- 46 Bradley CP. Turning anecdotes into data: the critical incident technique. *Fam Pract* 1992; **9**: 98–103.
- 47 Kekki P. Use of tracer activity for health centre evaluation. *World Health Forum* 1990; **11**: 447–450.
- 48 Crombie D, Fleming D. *Practice activity analysis*. Occasional paper 41. London: Royal College of General Practitioners, 1988.
- 49 Metsemakers JFM, Höppener P, Knottnerus JA, Kocken RJJ, Limonard CBG. Computerized health information in the Netherlands: a Registration Network of Family Practices. *Br J Gen Pract* 1992; **442**: 102–106.
- 50 Hoppener P, Knottnerus JA, Grol R, Metsemakers JFM. Computerisation of general practice and quality control. Blood glucose regulation in type 2 diabetics investigated in the Registration Network Family Practices. *Fam Pract* 1992; **9**: 353–356.
- 51 Hays R (ed.). *Standards Development Unit. Practice Surveyor Manual*. Sydney: RACGP, 1995.
- 52 Bradley N, Watkin S. Survey of equipment in general practice. *Br Med J* 1989; **299**: 435–436.
- 53 Van den Hombergh P, Grol R, Smits A, van den Bosch W. Visitatie van huisartspraktijken (A practice visit to assess the organisation of the general practitioner). *Huisarts Wet* 1995; **38**: 169–174.
- 54 Campbell LM, Howie JGR, Murray TS. Use of videotaped consultations in summative assessment of trainees in general practice. *Br J Gen Pract* 1995; **45**: 137–141.
- 55 Pringle M, Stewart-Evans C. Does awareness of being video recorded affect doctors' consultation behavior? *Br J Gen Pract* 1990; **40**: 45–48.
- 56 Royal College of General Practitioners. *What sort of doctor? Report from General Practice 23*. London: Royal College of General Practitioners, 1985.
- 57 Lundetrae N, Finckenhagen M, Hanstad AM, Maartmann-Moe K. Gjensidig praksisbesøk—en ny og god kursform (Exchange practice visits—a new and good way of training in general practice). *Tidsskr Nor Laegeforen* 1992; **112**: 3437–3438.
- 58 Fraser RC, McKinley RK, Mulholland H. Consultation competence in general practice: testing the reliability of the Leicester assessment package. *Br J Gen Pract* 1994; **44**: 293–296.
- 59 Hays RB, Peterson L. An evaluation of visiting clinical teachers in general practice training. *Ed Gen Pract* 1996; **7**: 54–58.
- 60 Rethans JJ, Sturmans F, Drop R, Vleuten vd C. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. *Br J Gen Pract* 1991; **41**: 97–99.
- 61 Saebø L, Rethans JJ, Johannessen T, Westin S. Standardiserte pasienter i allmennpraksis. En ny metode for kvalitetssikring i Norge (Standardised patients in general practice. A



new method for quality assessment in Norway). *Tidsskr Nor Lægefor* 1995; 115: 3117–3119.

- <sup>62</sup> Harden R, Gleeson F. Assessment of clinical competence using an objective structured clinical examination. *Med Educ* 1979; 13: 41–54.

<sup>63</sup> Hays RB, Bridges-Webb C, Booth B. Quality assurance in general practice. *Med Educ* 1993; 27: 175–180.

- <sup>64</sup> Baker R. Problem solving with audit in general practice. *Br Med J* 1990; 300: 378–380.