

Assessment of the performance of general practitioners by the use of standardized (simulated) patients

JAN-JOOST RETHANS

FERD STURMANS

RIET DROP

CEES VAN DER VLEUTEN

SUMMARY. A study was undertaken whereby a set of standardized (simulated) patients visited general practitioners without being detected, in a health care system where doctors had fixed patient lists. Thirty nine general practitioners were each visited during normal surgery hours by four standardized patients who were designed to be indistinguishable from real patients. The objective of the study was to see whether the actual performance of general practitioners, as assessed by standardized patients, met predetermined consensus standards of care for actual practice. The patients presented standardized accounts of headache, diarrhoea, shoulder pain and diabetes. The mean group scores of the doctors on the predefined standards of care for the different complaints ranged from 33 to 68%. The results show that standardized patients may be the method of choice in the assessment of the quality of actual care of doctors. It is hypothesized that the substandard scores of the doctors do not reflect inadequate competence, but are a result of the difference between competence and performance.

Introduction

IN the current debate about the quality of the performance of general practitioners, problems arise in defining the methods for assessing quality of practice.¹ Good methods are those which possess high validity and high reliability, but in real practice the feasibility of a method is also an important aspect. Traditional assessment methods have relied on written tests and clinical examinations but doubts have been cast on their validity and reliability.² Audit of medical records has been criticized for its low reliability in assessing several aspects of a consultation.^{3,4} Audio- and video-taping of consultations are methods which have both high validity and reliability. A disadvantage of these methods is that the researcher cannot control which patients enter the surgery room, making it difficult to compare performance between doctors. This is not a problem with the standardized (or simulated) patient method and therefore this method has been described as the best one for assessing the management of patients by doctors.⁴ With the help of standardized patients it has recently been shown that doctors in real

doctors in real practice performed more actions categorized as essential for good quality care than they said they would in an open ended questionnaire.⁵ Experience with the use of standardized patients who report data which can be considered reliable and valid is, however, highly limited and requires further testing.^{4,6}

Problems arise not only in defining the methods of assessing the performance quality of doctors, but even more in defining the level of quality of care, in other words, deciding what is 'competence', what is 'good' and what is 'bad'. One of the most common ways to define a level of quality for performance is to set standards for actual health care. In most instances a group of experts in general practice determines a set of standards for a particular medical problem. This procedure, however, carries the risk of setting 'armchair standards', or standards that have no basis in actual practice.⁷ It has, for example, been shown that a group of doctors who were first asked to assess a practical standard for a particular medical complaint, performed at only 56% of their own standard during actual consultations.⁴ The department of general practice at the University of Limburg has undertaken a study to determine whether the actual performance of general practitioners, as assessed by standardized patients, meets predetermined standards for actual practice.

Method

From among 24 nationally accepted sets of standards of care in general practice in the Netherlands, eight medical problems were identified which were all common in general practice, presented a diagnostic challenge and could be presented by a standardized patient.⁸ The 24 sets of standards had been determined by a consensus procedure with several stages. During the stages, experts in general practice as well as general practitioners 'in the field' commented on the sets of standards and tested them in practice. It was stressed that the standards should be practicable for actual practice and that they should not reflect an 'academic view' of real practice. The standards were divided into three categories: essential actions (considered to be necessary for good quality care), intermediate actions (not essential for good care but not harmful either) and superfluous actions. The eight problems which were selected for the study were translated into standardized roles for standardized patients to use. A panel of three general practitioners independently ranked these roles with respect to face validity. The four cases which ranked highest and on which there was agreement among the panel were finally chosen for this study. Figure 1 shows the most important features of the four cases. The facts which the doctor collected about history, the results of physical and laboratory examination, instructions given to the patient, treatment, and follow up were to be reported by the standardized patients and scored according to the sets of standards. The number of items on each standard ranged from 26 to 36. The standards used and their preparation process have been published in detail before.⁸ Figure 2 represents an example of one of the standards used.

Twelve standardized patients (six women and six men) were selected. Three standardized patients were allocated to each case study; these were of the same sex and approximately the same

Jan-Joost Rethans, MD, assistant professor, Centre for Quality Assurance in Research in General Practice, Universities of Nijmegen and Maastricht, The Netherlands; Ferd Sturmans, professor, Department of Epidemiology; Riet Drop, professor, Department of Medical Sociology; Cees van der Vleuten, PhD, associate professor, Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands.

Submitted: 16 May 1990; accepted: 4 September 1990.

Case 1.	A 30-year-old man visits his general practitioner with a headache. He is to say that he is an assistant accountant and has had the headache since he started his new job, three months ago, at an accounts office. The pressure to succeed at his job is quite considerable. In addition he has had to move away from his girlfriend, who now lives 200 kilometres away and whom he can meet only at the weekends.
Case 2.	A 26-year-old woman presents to her general practitioner with acute diarrhoea of three days onset and asks for something to stop her complaint. If the doctor asks questions about her job situation, she is to say that she works at a butcher's shop.
Case 3.	A 40-year-old woman presents to her general practitioner with pain in one shoulder, which she has had for four days. She is newly divorced and has a daughter aged 15 years. Should the doctor ask, she is to answer that the pain started after she had painted several doors in her new apartment.
Case 4.	A divorced man, 59 years old, with a diabetes type 2, visits his general practitioner for a new (repeat) prescription for his oral antidiabetic medication.

Figure 1. Features of the four standardized medical complaints.

<i>History</i>	
1.	Essential. Ascertain presence of diarrhoea.
2.	Essential. Ascertain presence of nausea, vomiting.
3.	Essential. Ascertain presence of abdominal pain.
4.	Essential. Ascertain course and duration of the complaints.
5.	Essential. Ask for possible causes such as contact with illness, travel, special foods etc.
6.	Essential. Ask whether any other complaints are present.
7.	Intermediate. Check medication being taken.
8.	Intermediate. Check eating habits history.
9.	Intermediate. Check for fever.
10.	Essential. Ascertain stool consistency and frequency.
11.	Essential. Ascertain nature of pain.
12.	Essential. Ascertain location of pain, shifts in location.
13.	Superfluous. Ask for additional information.
<i>Physical examination</i>	
14.	Essential. Examine abdomen: inspection, percussion, auscultation, palpation.
15.	Superfluous. Carry out further examinations.
<i>Laboratory tests</i>	
16.	Essential. Take stool culture.
17.	Superfluous. Carry out other laboratory tests.
<i>Guidance and advice</i>	
18.	Essential. Present the diagnosis.
19.	Essential. Discuss the prognosis.
20.	Essential. Give advice about diet.
21.	Superfluous. Give other guidance and advice.
<i>Medication</i>	
22.	Intermediate. Prescribe antiemetics.
23.	Intermediate. Prescribe absorbents.
24.	Superfluous. Prescribe other medication.
<i>Return visit</i>	
25.	Essential. Explain that patient should return in two or three days if complaints persist.
26.	Superfluous. Request return visit.

Figure 2. The standard of care for the standardized patient presenting with diarrhoea; the actions taken by the general practitioner are divided into essential actions, intermediate actions and superfluous actions.

age. The training of standardized patients for playing their role and for reporting reliable and valid facts about the doctor's performance has been described in another paper.⁹ Before the actual visits all standardized patients signed a written consent to keep all medical and personal information about the general practitioners in the project strictly for research purposes. The doctors selected for the study were visited by the standardized patients during a four month period starting at least 12 months after they agreed to participate.

Results

Of the 442 doctors asked to participate 137 (31%) agreed to be visited. Of these 137 doctors, 39 were selected. The criteria for selection have been outlined in a previous paper.⁹ The personal and practice characteristics did not differ from national characteristics.

For the subcategories 'essential', 'intermediate' and 'superfluous' actions, Table 1 shows the mean number of actions which were actually performed by the general practitioners for each of the four medical complaints. To some extent scores can be standardized as the percentage of the potential total maximum score in each category of action for the standard. Since the number of superfluous actions can be potentially infinite no percentages were calculated for this category. For each of the complaints, there was a clear distinction in the adherence to the standards between the essential and intermediate actions. We have also analysed which elements of the standards (for example, history or physical examination) were adhered to and which were omitted, but no consistent pattern was found for the four complaints. From Table 1 it can be seen that a much smaller proportion of the maximum scores were obtained by the doctors for the diabetic case than for the other cases.

Discussion

Since only 31% of doctors approached agreed to participate, the results might reflect performance of the more competent doctors in the province of the medical school. Although the personal and practice characteristics of the participating doctors did not deviate from national characteristics, we cannot exclude the possibility that the performance of non-participating doctors would be different, possibly lower, than the results in this study.

The results showed that in actual practice doctors met only between 33–68% of established national consensus standards for essential actions. Because these standards have been developed for practical use, one would expect that doctors in actual practice would meet 90–100% of the standards. Since 10 visits could be considered the minimum number to be representative of the individual doctor's usual level of performance, the results in this study do not allow us to look at individual doctors,¹⁰ but are only applicable to the participating doctors as a group. The fact that no consistent pattern was found regarding which elements of the standards were adhered to and which were omitted, might be the result of content specificity and of the fact that only four cases were used. An essential question in the current process of standard setting and especially with regard to the substandard scores of the participating doctors is whether this result reflects inadequate performance. In this whole project 137 doctors agreed to be visited over a period of three years. The fact that these doctors were not afraid of being audited with the very direct method of standardized patients, shows that they had faith in their own methods of dealing with patients. The conclusion from this study — that physicians in practice perform at a level considerably below the standards set by their peers — is not new.^{4,11} However, this has never before been demonstrated in actual practice with a method as direct as the

Table 1. Number of actions scored (mean, range and interquartile range) for the four medical problems as reported by standardized patients for consultations with 39 general practitioners.

	Number of actions scored			Mean no. of actions scored as a percentage of maximum no. from standard
	Mean (n = 39)	Range	Inter-quartile range (quart 1 to quart 3)	
<i>Headache case</i>				
Essential actions (maximum 13)	8.9	5-12	8-10	68
Intermediate actions (maximum 12)	3.8	0-8	3-5	32
Superfluous actions	2.6	0-7	1-4	
<i>Diarrhoea case</i>				
Essential actions (maximum 15)	8.5	5-14	7-10	57
Intermediate actions (maximum 5)	2.4	1-4	1-3	48
Superfluous actions	1.8	0-8	0-2	
<i>Shoulder pain case</i>				
Essential actions (maximum 19)	12	4-16	10-14	63
Intermediate actions (maximum 6)	1.7	0-3	1-2	28
Superfluous actions	1.8	0-6	1-3	
<i>Diabetic case</i>				
Essential actions (maximum 21)	6.9	2-14	5-8	33
Intermediate actions (maximum 6)	0.2	0-2	0-0	3
Superfluous actions	2.0	0-7	1-3	

n = number of consultations.

use of standardized patients with four visits per doctor. As already mentioned in the introduction, whether standards are external (as in this study) or internal does not seem to affect the results obtained: doctors still performed below the standards set.

This observation of substandard performance of doctors has in the past led to actions by various professional organizations. These actions were primarily directed towards mandatory attendance at continuing medical education. The underlying assumption of these directives is that poor performance is a reflection of inadequate knowledge and/or skills, which would be remediable by additional instructions. Studies about the effects of postgraduate education on the behaviour of doctors in actual practice often produce conflicting evidence.¹² It has, for instance, been shown that additional postgraduate education does not seem to change the practice behaviour of doctors.¹³ Recently it has also been shown in a study of postgraduate teaching of funduscopy to general practitioners that there was no measureable learning effect, although the doctors were very enthusiastic.¹⁴

An alternative explanation to that of inadequate competence is that physicians, when not under supervision, do not perform at the level they are capable of.¹⁵ The willingness of all participating doctors to allow their actual behaviour to be scrutinized, however, suggests that the results of this study do not reflect inadequate performance *per se*. The subjective opinions of the standardized patients after their visits support this view. At the first visits in actual practice some patients were almost embarrassed because they felt that the doctors were really interested in them and did their best to help them. The results of this study stress the need for more studies which investigate the relation-

ship between actual behaviour and maximum competence of practising doctors.

Another explanation for the substandard scores of the doctors could be that the standards involved, notwithstanding the consensus procedure used, do not adequately reflect the longitudinal relationship between patients and their general practitioners. It is known that in actual practice doctors show efficient performance and do only what is necessary at that particular moment.⁶

The large difference between the mean score for the chronic diabetic case and those for the other cases could be because doctors do not ask all their questions at the first consultation, but spread their questions over several visits. This hypothesis can be tested by visiting doctors several times with the same standardized patient. Our department of general practice is currently involved in such a study. The results of the present study stress the need to establish standards which take account of longitudinal relationships in general practice.

The finding that doctors perform below predetermined standards does not prove that doctors are incompetent; it should at least be tested against the hypothesis that standards for actual care are still not realistic.

It can be concluded that the standardized patient method is feasible and is the most direct method for assessing the performance of practising physicians. The data gathered with this method about actual practice can give a new stimulus to the debate about what constitutes 'good' and 'bad' in general practice.

References

- Mok J. Assessment in general practice *J R Coll Gen Pract* 1988; **38**: 344-345.
- Maguire P. Assessing clinical competence. *Br Med J* 1989; **298**: 4-5.
- Fessel JW, van Brunt EE. Assessing quality of care from the medical record. *N Engl J Med* 1972; **286**: 134-138.
- Norman GR, Neufeld VR, Walsh A, et al. Measuring physicians' performances by using simulated patients. *J Med Educ* 1985; **60**: 925-34.
- McClure CL, Gall EP, Meredith KE, et al. Assessing clinical judgement with standardized patients. *J Fam Pract* 1985; **20**: 457-464.
- Rethans JJE, van Boven CPA. Simulated patients in general practice: a different look at the consultation. *Br Med J* 1987; **294**: 809-812.
- Neufeld VR, Norman GR. *Assessing clinical competence*. New York: Springer, 1985.
- Grol R, Mesker P, Schellevis F (eds). *Peer review in general practice*. Nijmegen: University Department of General Practice, 1988: 87-89; 97-98; 109-108; 117-119.
- Rethans J-J, Drop R, Sturmans F, van der Vleuten C. A method for introducing standardized (simulated) patients into general practice consultations. *Br J Gen Pract* 1991; **41**: 94-96.
- van der Vleuten CPM. Assessment of clinical skills with standardized patients: state of the art. In: *Naar een rationeel systeem voor toetsing van studieprestaties in probleemgestuurd onderwijs (thesis)*. Maastricht University, 1989: 93-140.
- Gonella JS, Goran MJ. Evaluation of patient care. *JAMA* 1970; **214**: 2040-2043.
- Haynes R, Davis D, McKibbon A, Tugwell P. A critical appraisal of the efficacy of continuing medical education. *JAMA* 1984; **251**: 61-64.
- Sibley JC, Sackett DL, Neufeld V, et al. A randomized trial of continuing medical education. *N Engl J Med* 1982; **306**: 511-515.
- Reeders K, Hiemstra RJ, Bender W. Teaching funduscopy to GPs in an OSCE setting: failures and successes. In: Bender W, Hiemstra RJ, Scherpier AJJA, Zwierstra RP (eds). *Teaching and assessing clinical competence*. Groningen: Boekwerk Publications, 1990.
- Rethans J-J, Leeuwen VY, Drop R, et al. Performance and competence: two different constructs in the assessment of quality of medical care. *Fam Pract* 1990; **7**: 168-174.

Address for correspondence

Dr J-J Rethans, Department of General Practice, University of Limburg, Postbus 616, 6200 MD Maastricht, The Netherlands.