

The effects of peer review in general practice

R. GROU, PhD

H. MOKKINK, PhD

F. SCHELLEVIS, MD

SUMMARY. *This paper describes the effects of an intensive, structured programme of peer review on the behaviour of general practitioners. Trained assessors evaluated 43 doctors before and after participating in the programme. Both the medical and non-medical performance of the general practitioners was assessed using a previously tested measuring instrument. In addition, the prescribing behaviour of the participants was compared with that of a control group. After taking part in the programme, the work of the general practitioners conformed more closely to a number of criteria for good general practice care. This was particularly true for clarifying the questions and expectations of the patients, active stimulation of patient involvement, history taking and providing information and advice to patients. The greatest change occurred among general practitioners who had previously conformed least with the established criteria. During peer review participants in the project prescribed fewer drugs, such as analgesics, tranquillizers and antibiotics, than before peer review while non-participants prescribed more. The value of these changes is discussed.*

Introduction

OVER the period 1981–86 the Department of General Practice at Nijmegen University carried out a project to study peer review.^{1,2} The project covered both medical and non-medical activities of general practitioners and aimed to motivate general practitioners to take part in peer review, to teach them the relevant skills, to make them more aware of their performance, and to correct their performance where necessary.

A total of 243 general practitioners from the area, mostly trainers for vocational and undergraduate training, participated in the project. The majority had a very positive attitude towards participation.

The participants met in groups of about 10 for one day a month (excluding holidays) over a period of between 18 months and two years. Use was made of definite standards and criteria for general practice care formulated before the start of the project and based on a literature study, research and panel discussions by experienced general practitioners and social scientists.¹ These criteria were discussed by the groups and once accepted, were used as a measure of good general practice care. At the beginning of the programme the groups were given a great deal of guidance but later the aim was to make them independent.

The supervisors of the groups followed a set of instructions so that their approach was as consistent as possible.¹ The basic structure of the programme was:

1. Introduction (three or four sessions) — creating a confidential atmosphere and teaching basic audit skills (assessment, using criteria, and giving and receiving feedback).

2. Initial review (four sessions) — teaching methods of audit, for example, evaluation of audiotaped consultations and assessment of medical performance.

3. Reinforcement (four or five sessions) — teaching variations in methods of audit and the assessment of progress in performance.

4. Final phase (one or two sessions) — preparing for audit without guidance.

This paper describes changes in the medical and non-medical performance and in the prescribing behaviour of a group of general practitioners who took part in the project. Previous results of the project have been described elsewhere.^{2,3}

Method

Assessment of performance

Data on the medical and non-medical performance of 43 of the general practitioners before they participated in peer review were available from another study.⁴ These doctors did not differ significantly from the national population of general practitioners in factors such as age or age structure of their practices. The performance of these 43 doctors was also assessed between six months and one year after peer review.

Trained, external assessors observed the participants in their surgeries for two days and scored the consultations which had been recorded on audiotape. This was done using a research version of the same peer review criteria. This research instrument has been developed and tested in earlier research and has proved to be reasonably valid.^{4,5} The criteria used in the instrument have been termed 'important' in a number of studies.⁶⁻⁸ Each general practitioner had 15–20 consultations assessed before peer review and 12–15 consultations assessed after peer review. In total about 1300 consultations were assessed.

Medical performance. In order to assess medical performance, criteria for good performance were used in the areas of history taking, examination, providing information, therapy, referral behaviour and asking patients to report back. There are 24 such 'protocols', which cover 55–60% of all problems presented in Dutch general practice. Each protocol comprises an average of 35 'obligatory' activities. For each general practitioner between 400 and 700 activities, taken from different protocols, were used in the assessment. A calculation was made to find out what percentage of the 'obligatory' activities were actually carried out. By applying factor analysis to the doctor's percentage on all the protocols, it appeared that a high percentage on a single protocol indicated a high percentage on the others. A mean percentage for a large number of activities in a series of consultations therefore reflects the medical performance of a general practitioner. The inter-reliability of the assessors was found to be extremely good when compared with other studies.⁹ After correction for the chance of consensus, the kappa value was 0.71–0.75.

Non-medical performance. The non-medical activities were assessed using detailed criteria for aspects of care such as 'the general orientation of consultations', 'clarifying the expectation and questions of the patient' and 'providing an explanation to patients'. Assessment took place at each consultation using a five point scale from 'not in agreement with the standards' (score 1) to 'in total agreement with the standards' (score 5). The inter-reliability was lower than for medical performance — mean Kendall's tau at the pre-measurement phase 0.30–0.50, at the post-measurement stage 0.40–0.60. However, this level is acceptable,

R. Grol, Psychologist; H. Mokkink, Sociologist; F. Schellevis, General Practitioner, Department of General Practice, University of Nijmegen, The Netherlands.

© *Journal of the Royal College of General Practitioners*, 1988, 38, 10–13.

if one considers the complexity of the care situation and the nature of the criteria.¹⁰

To ensure that distortion was reduced to a minimum, the consultations of each general practitioner were spread between different assessors, and a check and correction procedure was carried out.

Changes in the medical and non-medical performance. The median percentages and scores of the 43 participants for the consultations assessed before peer review were compared with the median scores after peer review. It proved impossible to form a comparable control group for this analysis. A t-test for paired observations was used to determine the significance of the differences found.

Prescribing behaviour

The prescribing behaviour of some of the participants was compared with that of a control group. Data from a regional sick fund were available for the period 1980–83 for 19 participants. General practitioners with pharmacy facilities in their practice, who worked in a group practice or for whom data were incomplete were excluded. Forty-five comparable non-participants for whom data were available from the same sick fund were selected as controls. For both groups the mean annual number of prescriptions per sick fund patient in the following categories was calculated:¹¹ (1) all medicines; (2) drugs with a clearly defined range of indications, for example, drugs for lowering blood pressure or blood sugar level; (3) drugs with a wide range of indications and/or of disputable application, for example, analgesics, tranquillizers, antibiotics.

Figures for the two-year period of peer review (1982–83) were compared with the two-year period before peer review (1980–81). In 1984 the doctors in the control group started to take part in peer review, so their figures could no longer be used.

Results

Medical performance

The percentage of 'obligatory' activities performed in the consultations was higher after participation in peer review (Table 1). This applied particularly to history taking and providing information. The changes were slight for carrying out a physical examination. Overall, 74% of the participants had

Table 1. Changes in medical performance: percentage of 'obligatory' activities performed by the 43 general practitioners before and after peer review (736 consultations were assessed before peer review and 543 after peer review).

	Percentage of activities performed	
	Before peer review	After peer review
<i>Diagnostic phase</i>		
Total	47	53**
History taking	50	58**
Physical examination	50	52
<i>Strategic phase</i>		
Total	40	47**
Providing information/advice	38	46**
Asking patients to report back	45	49
<i>Entire consultation</i>	45	51**

** $P < 0.01$ (t-test) for after peer review versus before peer review.

changed their performance to bring the entire consultation more in line with the criteria.

The changes in medical performance were also visible at a detailed level. For example, Table 2 shows the 'obligatory' activities for hypertension control. It can be seen that more activities in the history-taking category were performed after peer review than before. The results of blood pressure measurements were usually revealed to the patient but in 30% of consultations a follow-up appointment was not made and advice regarding risk behaviour was seldom given.

Table 2. Medical performance of the 43 general practitioners before and after peer review for hypertension control: percentage of 'obligatory' activities performed (145 consultations were assessed before peer review and 79 after peer review).

	Percentage of activities performed (n)	
	Before peer review	After peer review
<i>History-taking</i>		
Asking about complaints, feelings, etc.	67 (145)	73 (79)
Discussing complaints	71 (58)	89 (27)
Clarifying risk factors	55 (145)	67 (79)
Discussing compliance with medication	29 (115)	51 (67)
Blood pressure high: discussing compliance with medication	28 (43)	70 (23)
<i>Examination</i>		
Measuring blood pressure	97 (145)	100 (79)
<i>Information</i>		
Explanation of findings of examination	88 (145)	95 (79)
Advice regarding behaviour	20 (35)	8 (37)
<i>Medication</i>		
Adjusting medication when diastolic blood pressure ≥ 100 mmHg	29 (28)	29 (14)
Decreasing medication when < 100 mmHg	19 (31)	39 (53)
<i>Follow-up</i>		
Making a follow-up appointment	65 (145)	70 (79)

n = number of consultations at which an activity is obligatory, according to the protocol.

Non-medical performance

The group of 43 general practitioners worked more in agreement with the criteria after peer review (Table 3). This difference was most marked for general orientation, clarifying the questions and expectations of the patient, systematic structure to the consultation, active stimulation of patient involvement and providing an explanation and information to the patient. Nearly all of the participants changed in the direction intended in these categories. Little difference could be seen for listening to the patient and openness, which was already at a high level, and paying attention to psychosocial aspects.

Differences between participants

Some participants changed considerably after peer review, others hardly at all. The greatest changes were to be found among those general practitioners who, before participating in peer review, worked least in agreement with the criteria (low scorers) (Table 4). The 'regression to the mean effect' is slight in this case because the high scorers also changed after peer review.

Table 3. Changes in non-medical performance: mean scores before and after peer review for the 43 general practitioners (612 consultations were assessed before peer review and 543 after peer review).

	Mean scores ^a	
	Before peer review	After peer review
General orientation	2.2	2.9**
Clarifying the questions and expectations of patient	1.6	2.7**
Systematic structure	3.6	4.0**
Listening to patient and openness	3.5	3.6
Active stimulation of patient involvement	1.7	2.8**
Providing explanation and information to patient	2.9	3.3**
Attention to psychosocial aspects	1.6	1.7

** $P < 0.01$ (t-test) for after peer review versus before peer review.

^a Score 1 'not in agreement with the standards', score 5 'in total agreement with the standards'.

Prescription behaviour

There was a significant difference between the study group and the control group in the prescribing of drugs with a wide range of indications (Table 5) — during peer review the participants prescribed, on average, half an item less per patient than in the two years before review, while the control group prescribed half an item more. For a standard practice with 1600 sick fund patients, this difference represents approximately £4000 for the auditing period or £2000 a year.

Discussion

Following the programme of peer review the participants began to work more in agreement with the criteria which they had accepted before the audit. General practitioners who had conformed least with the criteria beforehand showed the most marked change. In addition, the prescribing by participants of drugs such as analgesics, tranquillizers and antibiotics — drugs which should be prescribed in moderation — was shown to decrease and this change was significantly different from the increase shown by the control group. When it is remembered that general practitioners do not change their behaviour easily, and that the effects of auditing and continuing education are generally small^{12,13} it can be seen that the project produced a number of positive results.

However, criticisms can be made about the study. The first concerns the framework of the research. The study suffered several methodological problems such as forming equal control groups, standardizing the experimental stimulus and making accurate and valid measurements.^{14,15} Therefore, care must be taken when making generalizations. In addition, the study was based on a small sample, especially for changes in prescribing behaviour. It is difficult to draw conclusions about the relationship between taking part in peer review and the changes observed although this relationship is plausible. For nearly all the doctors the programme of peer review was the first educational activity they had taken part in for a long time and changes took place in several aspects of general practitioner care, especially those emphasized in the programme.

The results of this study can be compared with those of other studies.^{8,16,17} In a large-scale audit Fleming and Lawrence¹⁶ found even larger changes in the medical performance of general practitioners after peer review, particularly among doctors who had low scores before audit. However, Fleming and Lawrence analysed records whereas our study made a more valid assess-

ment of real performance in the practice.

The second criticism of the study concerns the criteria for adequate performance which were used. Even after taking part in peer review, the general practitioners were still a long way from 'optimal' performance and those doctors who had conformed least to the criteria beforehand changed most. This suggests that there must be a maximum level of performance that can be attained and the criteria may be too stringent. However, there were large differences in the performance of individual general practitioners. Some doctors carried out 70% of the 'obligatory' activities and some only 30%. Subjective observations confirmed that many general practitioners are seriously concerned about providing a qualitatively high level of care, but many lack commitment. At present lack of commitment and differences in working methods are accepted too easily and a change in outlook is called for if real changes are to be achieved in general practice medical care.¹⁸

The final criticism of this study concerns the value of the changes in performance. The changes observed are not great, they may not be sustained and the energy invested may not have been worthwhile. It is sensible to look not only at the changes in performance, but also at the more subjective assessments made by the participants.³ The programme motivated a high percentage of the trainers in the region to examine their own work critically for a long period and most of the participants had a very positive attitude towards the project, even six months after it had been completed. Opposition to peer review largely disap-

Table 4. Changes in percentages/median scores after peer review among general practitioners who had low and high scores before peer review.

	Change in percentage/ median score	
	Low scorers ($n = 20-23$)	High scorers ($n = 20-23$)
<i>Medical performance</i>		
Diagnostic phase	+ 8.4%	+ 2.3%*
Strategic phase	+ 7.7%	+ 5.1%
<i>Non-medical performance</i>		
General orientation	+ 0.8	+ 0.5*
Clarifying the questions and expectations of patient	+ 1.1	+ 1.2
Systematic structure	+ 0.7	+ 0.2**
Listening to patient and openness	+ 0.1	+ 0.1
Active stimulation of patient involvement	+ 1.2	+ 0.9
Providing explanation and information to patient	+ 0.7	+ 0.1**
Attention to psychosocial aspects	+ 0.2	0.0

** $P < 0.01$; * $P < 0.05$ (t-test) for high scorers versus low scorers. n = number of doctors.

Table 5. Changes in the mean annual number of prescriptions per sick fund patient between 1980-81 (before peer review) and 1982-83 (during peer review).

Prescriptions	Study group ($n = 19$)	Control group ($n = 45$)
All drugs	- 0.81	+ 0.18
Drugs with a wide range of indications	- 0.54	+ 0.46*
Drugs with a clearly defined range of indications	+ 0.59	+ 0.52

* $P < 0.05$ (t-test) for study group versus control group. n = number of doctors.

peared, as did feelings of being threatened. The participants reported a greater degree of consciousness about their activities and a more critical approach to their profession. They also had less fear of being criticized and judged.

This study has provided the basis for the long-term development of quality control in general practice. The measured changes in behaviour form only one of the parameters by which the success of the auditing project can be measured.

References

1. Grol R, Mesker P (eds). *Huisarts en onderlinge toetsing*. Utrecht: Wetenschappelijke Uitgeverij Bunge, 1986 (English translation in preparation).
2. Grol R, Eiji J van, Mesker P, Schellevis F. Audit: a project on peer review in general practice. *Fam Pract* 1985; 2: 219-224.
3. Grol R. *Kwaliteitsbewaking in de huisartsgeneeskunde. Effecten van onderlinge toetsing*. Dissertatie Katholieke Universiteit Nijmegen, 1987.
4. Smits A, Mokkink H, Grol R, et al. Spreekuurgedrag van huisartsen. *Huisarts en Wetenschap* 1985; 28: 121-125.
5. Mokkink H, Smits A, Grol R. Prevara: een observatie instrument voor het handelen van de huisarts in het kader van processen van somatische fixatie. *Ned T Psychologie* 1982; 37: 35-50.
6. Royal College of General Practitioners. *What sort of doctor? Report from general practice 23*. London: RCGP, 1985.
7. Fabb W, Marshall J. *The assessment of clinical competence in general family practice*. Lancaster: MTP Press, 1983.
8. Maguire P, Fairbairn S, Fletcher C. Consultation skills of young doctors. *Br Med J* 1986; 292: 1573-1578.
9. Koran L. The reliability of clinical methods, data and judgements. *N Engl J Med* 1975; 293: 642-646, 695-701.
10. Katz F, Snow R. *Assessing health workers performance. A manual for training and supervision. WHO public health papers no. 72*. Geneva: World Health Organization, 1980.
11. Mokkink H. *Ziekenfondscijfers als parameter voor het handelen van huisartsen*. Dissertatie Katholieke Universiteit Nijmegen, 1986.
12. Haynes R, Davis D, McKibbin A, Tugwell P. A critical appraisal of the efficacy of continuing medical education. *JAMA* 1984; 251: 61-64.
13. Sanazaro P. Determining physicians performance. Continuing medical education and other interacting variables. *Evaluation and the Health Professions* 1983; 2: 197-210.
14. Smith N. The feasibility and desirability of experimental methods in evaluation. *Evaluation and Program Planning* 1980; 3: 251-256.
15. Winickof R, Coltin R, Morgan R, et al. Improving physician performance through peer comparison feedback. *Med Care* 1984; 6: 527-534.
16. Fleming D, Lawrence M. Impact of audit on preventive measures. *Br Med J* 1983; 287: 1852-1854.
17. Verby J, Holden P, Davies R. Peer review of consultations in primary care. The use of audiovisual recordings. *Br Med J* 1979; 1: 1686-1688.
18. McIntyre N, Popper K. The critical attitude in medicine: the need for a new ethics. *Br Med J* 1983; 287: 1919-1923.

Address for correspondence

Dr R. Grol, Department of General Practice, University of Nijmegen, Medical Faculty, Postbus 9101, 6500 HB Nijmegen, The Netherlands.



The Royal College of General Practitioners ACCOMMODATION AND CATERING

Members of the College are welcome to stay at 14 and 15 Princes Gate; early booking is recommended. Bed and breakfast may be obtained. Bookings should be sent to Mrs Lindsey Demetriou, the Accommodation Secretary. Public rooms may be hired subject to availability. Please contact Miss Elizabeth Monk, Secretary to the Establishments Officer, at the Royal College of General Practitioners, 14 Princes Gate, Hyde Park, London SW7 1PU. Telephone 01-581 3232.

INFECTIOUS DISEASES UPDATE

Lyme disease

This is a recently recognized tick borne spirochaetal infection caused by *Borrelia burgdorferi*. It appears to be spreading quite rapidly in parts of the USA and has been recognized in other parts of the world including the UK. The ixodid tick feeds mainly on field mice but also on deer which appear to be responsible for spreading the infection over larger areas. Those at most risk include ramblers and farmers who pick up ticks, usually from long grass and bracken. The resulting illness may be a rash (erythema chronicum migrans) which starts at the site of the tick bite and spreads progressively outwards. However, infection can cause joint problems which may be confused, for example, with rheumatoid arthritis which can be acute or chronic. In addition neurological features have been noted and meningitis can occur. It is important to think of Lyme disease as a cause of these symptoms especially in those with occupations putting them at risk. A reliable serological test is available and treatment with penicillin or tetracycline can be helpful especially if given early in the illness.

Cryptosporidiosis

A recent cluster of cryptosporidium reports from south Wales has drawn attention to this infection which has only in recent years been widely recognized as a cause of disease. The cystic forms are only recognized in stool samples if special staining techniques are used, so it may be necessary for general practitioners to request the test specifically. In the UK approximately 4000 cases were recognized during 1986 and there is possibly a seasonal increase in cases in the late spring and the autumn. Outbreaks have occurred particularly in day nurseries. Acute diarrhoeal illnesses can occur but grumbling symptoms similar to those produced by giardial infection are typical. In the immuno-compromised, especially those with the acquired immune deficiency syndrome, infections can be more severe and become generalized. Treatment in the otherwise healthy does not normally involve antimicrobial drugs and symptoms are usually self limiting.

Rubella vaccination

Evidence is accumulating that when rubella vaccine is given inadvertently during pregnancy the congenital rubella syndrome is unlikely to occur. There is an ongoing study in the USA, of over 1000 pregnancies so far, where live attenuated rubella vaccine was given within three months before or after the dates of conception. Cendihill, HPV-77 and the RA 27/3 vaccines were involved. In the occasional infant there was serological or culture evidence of rubella infection but there has been no evidence of the fetal abnormalities that are recognized to result from 'wild' virus infection. These encouraging data are consistent with experience in the UK and Europe. Pregnancy remains a contraindication to rubella vaccination because of the theoretical risks from these live attenuated viruses but the discovery that a woman was pregnant when she received the vaccine does not now make termination of pregnancy on these grounds a logical step.

Yellow fever

Yellow fever is currently being reported from Mali and southern districts of Mauritania. The World Health Organization advises that vaccination is especially important for travellers to this part of West Africa.

Suggestions for topics to include in future updates are welcomed and should be passed to the contributor, Dr E. Walker, Communicable Diseases (Scotland) Unit, Ruchill Hospital, Glasgow G20 9NB (041-946-7120), from whom further information about the current topics can be obtained.