

Feedback of patients' evaluations of general practice care: a randomised trial

E Vingerhoets, M Wensing, R Grol

Abstract

Objective—To assess the effects of feedback of patients' evaluations of care to general practitioners.

Design—Randomised trial.

Setting—General practice in the Netherlands.

Subjects—55 GPs and samples of 3691 and 3595 adult patients before and after the intervention, respectively.

Interventions—GPs in the intervention group were given an individualised structured feedback report concerning evaluations of care provided by their own patients. Reference figures referring to other GPs were added as well as suggestions for interpretation of this feedback, an evidence-based overview of factors determining patients' evaluations of care, and methods to discuss and plan improvements.

Main outcome measures—Patients' evaluations of nine dimensions of general practice measured with the CEP, a previously validated questionnaire consisting of 64 questions, using a six point answering scale (1=poor, 6=very good).

Results—Mean scores per CEP dimension varied from 3.88 to 4.77. Multilevel regression analysis showed that, after correction for baseline scores, patients' evaluations of continuity and medical care were less positive after the intervention in the intervention group (4.60 v 4.77, $p < 0.05$ and 4.68 v 4.71, $p < 0.05$, respectively). No differences were found in the remaining seven CEP dimensions.

Conclusions—Providing feedback on patients' evaluations of care to GPs did not result in changes in their evaluation of the care received. This conclusion challenges the relevance of feedback on patients' evaluations of care for quality improvement.

(Quality in Health Care 2001;10:224–228)

Keywords: patient feedback; general practice; quality improvement

A positive evaluation by patients of the care received is an important outcome which is related to patients' adherence with treatments, (re)attendance of the care provider,^{1,2} and functional health status.^{3,4} Performing surveys and interviews with patients to elicit their views on healthcare delivery is increasingly common in all sections of health care.⁵ It is used so frequently that its suitability as a tool for quality improvement and increasing patient involvement in health care seems to be almost

self-evident, yet it remains unclear whether feedback of patients' evaluations of care to clinicians is an effective method for improving professional performance and organisation of services.^{2,6} Quality improvement methods should prove their effectiveness in well designed studies before they are implemented in health care.^{7,8} However, few experimental studies on instruments to elicit feedback from patients have been performed. Only a few small uncontrolled before/after comparisons^{9–15} and anecdotal experiences support expectations that feedback on patients' views may be effective. We have therefore performed a randomised trial aimed at determining the effect of feeding back patients' evaluations of care to general practitioners.

Methods

SAMPLE

Sixty general practitioners (GPs) from 43 general practices were recruited out of a systematic sample of 700 GPs in the Netherlands, stratified for urbanisation level to reflect the national situation. Two independent patient samples were taken: the first at inclusion of the GP in the study in months 1–3 and the second 15 months after the study started. Each GP recruited 100 consecutive patients visiting their practice, giving a total sample of at least 3000 patients before and after the intervention (power calculation below) with an expected response rate of 50%. Patients under 18 years of age, those who were not able to understand the Dutch language, and those who were mentally handicapped or terminally ill were excluded.

PROCEDURES

The study, which was performed between March 1997 and March 1999, was designed as a randomised trial with two equally sized

Key messages

- Despite a long history of research on patients' views of health care, evidence on the relevance of surveys among patients for improving the quality of care is limited.
- Feedback of patients' evaluations of general practice care could help general practitioners to identify opportunities for improvement.
- This study showed, however, that patients' evaluations of care did not improve after the feedback.
- Further development is needed before wide scale implementation of the method can be recommended.

Centre for Quality of Care Research (WOK), Universities of Nijmegen and Maastricht, The Netherlands

E Vingerhoets, general practitioner

M Wensing, senior

research fellow

R Grol, professor in general practice

Correspondence to:

Dr M Wensing, Centre for Quality of Care Research (WOK), University Medical Centre Nijmegen, P O Box 9101, 6500 HB Nijmegen, The Netherlands
M.Wensing@hsv.kun.nl

Accepted 30 July 2001

groups: an intervention group and a non-intervention control group. After matching for practice size (single practitioner versus group practice), general practices were randomly allocated to either the intervention or control group using a computer generated list of random numbers. The allocation was performed by a statistician who was not directly involved in the project and was concealed from others. Patients were blinded for the intervention (assuming that the GPs followed the instructions not to inform the patients about the feedback), but blinding of the GPs was not possible.

MEASUREMENTS

The outcome measure was the evaluation by the patients of general practice care measured with the CEP, a previously validated instrument.¹⁶ This structured questionnaire consists of 51 questions covering nine dimensions of care: organisation of appointments (9 questions), availability for emergencies (3), premises (3), continuity (4), cooperation (4), medical care (6), relation and communication (10), information and advice (6), and support (6). Answers are given on a six point scale ranging from 1=poor to 6=very good. The questionnaire also contains questions on patients' sex, age, education, chronic illness (list of 25 diseases), number of visits to the GP in the past 2 months, and overall health status (5 point scale).

Questionnaires were handed out by the GP or his/her assistant immediately after the consultation. They could be completed at home and returned anonymously to the University of Nijmegen in a prepaid envelope. Measurements were performed directly after inclusion of the GP in the study in months 1–3 and repeated at month 15.

A written questionnaire, distributed to the participating GPs before the intervention, was used to obtain the characteristics of the participating GPs and included questions on sex, age, experience as a GP, practice setting, urbanisation level, personal list system, and provision of GP training. In the intervention group we also asked GPs after the intervention to report on changes as a result of the feedback and to categorise those actions according to the dimensions in the patient questionnaire.

INTERVENTION

The intervention comprised an individual written feedback report on patients' evaluations of care using data from the pre-intervention measurement. Each GP received this 15 page report 3–6 months after the start of the study. It contained the figures for each question and aggregated scores for each of the nine dimensions of care. Figures were exclusively related to patients of the particular GP. Reference figures related to patients from all GPs were added, as well as an abstract of a systematic review of studies on determinants of patients' evaluations of care (E Vingerhoets, M Wensing, P Van Montfort, *et al*, unpublished report, 1997) and a short manual with suggestions on how to interpret and deal with the

information. This manual described a number of ways to use the results including discussions with colleagues and assistants, detailed follow up surveys among patients, and establishment of a patient panel. All written text in the report was standardised, not tailored to the individual GP. The control group received this intervention only after the post-intervention measurements had been taken.

ANALYSIS OF DATA

For each dimension on the CEP questionnaire an aggregated score was calculated which was the mean score for this dimension. An aggregated score was calculated for a patient if less than half of the questions had missing values. Missing values were substituted by the mean of the valid items in the dimension. In remaining cases the aggregated score was not calculated for that particular patient. In order to detect a difference between the intervention and control groups in the patient evaluation score before and after the intervention of 0.3 points on the 6 point scale, samples of 3000 patients with a minimum response rate of 50% per GP both before and after the intervention were required (power=0.80, alpha=0.05, standard deviation=1.3, intraclass rho=0.08).

An analysis was made of the GPs who dropped out of the study after the pre-intervention measurement by comparing them with the remaining GPs with respect to GP characteristics, characteristics of the patient samples, and patients' evaluations of care. The intervention and control groups were compared with respect to GP characteristics, patient characteristics, and patients' evaluations of care before the intervention and repeated for patient characteristics and patients' evaluations of care after the intervention. A χ^2 test or a *t* test was used to test the differences statistically and *p* values of 0.05 or less were considered to be significant.

A series of multilevel regression analyses were performed that took into account the fact that patient data were clustered within GPs (patients with the same GP have something in common). This was done by including a GP factor in all models. For each of the nine dimensions of patients' evaluations of care the following regression models were tested (table 1). The first model (A) used the difference between pre- and post-intervention patients' evaluations of care as a dependent variable and group allocation (treatment versus control) as a potential predictor. The remaining models used the post-intervention patients' evaluation of the dimensions of the CEP as dependent variables. Model B examined the proportion of variation systematically related to differences between GPs. Model C compared the experimental groups (intervention versus control) to determine the straightforward intervention effect. The same model was also used to test pre-intervention differences between groups. Model D added the mean pre-intervention score of the GP to the relevant dimension to determine a "conditional intervention effect". It was not possible to include pre-intervention

Table 1 Overview of the multilevel regression models

Model	Aim	Dependent factor	Potential predictors
A	To determine the difference between experimental groups regarding the within group change in patients' evaluations of care over time, taking into account clustering of data within GPs	Change of patient evaluations per patient (post-intervention score minus baseline score)	GP factor; group allocation (intervention <i>v</i> control group)
B	To determine the proportion of variation associated with systematic differences between GPs regarding post-intervention patients' evaluations of care	Post-intervention patients' evaluations per patient	GP factor
C	To determine the difference between experimental groups regarding post-intervention patients' evaluations of care, taking clustering of data within GPs into account ("straightforward intervention effect")	Post-intervention patients' evaluations per patient	GP factor; group allocation (intervention <i>v</i> control group)
D	To determine the difference between experimental groups regarding post-intervention patients' evaluations of care, taking into account the clustering of data within GPs as well as the pre-intervention scores ("conditional intervention effect")	Post-intervention patients' evaluations per patient	GP factor; group allocation (intervention <i>v</i> control group); pre-intervention score of patient evaluation (mean score per GP)
E	To determine the difference between groups defined by GP reported actions on the relevant dimensions regarding post-intervention patients' evaluations of care, taking into account the clustering of data within GPs as well as the pre-intervention scores	Post-intervention patients' evaluations per patient	GP factor; group allocation defined by GP reported action on the relevant dimension (yes <i>v</i> no); pre-intervention score of patient evaluation

scores of individual patients because the patients before and after the intervention were different. Finally, model E used the GP reported action on a specific dimension (yes versus no) rather than group allocation as a predictor (model E).

Results

During the study five GPs dropped out (one in the intervention group and four in the control group) for reasons of burn out, break up of practice group, or long term illness. These five GPs did not differ from the remaining 55 GPs with respect to the GP characteristics measured, except that none of the five had a personal list system ($p=0.011$) and none provided training for GPs ($p=0.035$). The 333 patients of these five GPs did not differ from the 3691 patients of the remaining 55 GPs on most pre-intervention measures. However, the patients from the five GPs were older (53.1 *v* 49.7 years, $p=0.0002$) and had made fewer visits to the GP (2.2 *v* 2.4, $p=0.05$). In addition, they had less positive evaluations of the following dimensions of care: organisation of services (4.36 *v* 4.46, $p=0.05$), premises (3.71 *v* 3.89, $p=0.0188$), continuity (4.53 *v* 4.67, $p=0.0301$), relation and communication (4.38 *v* 4.63, $p=0.0001$) and information giving (4.46 *v* 4.66, $p=0.0012$). The remainder of this paper focuses on the 55 GPs who provided measurements before and after the intervention.

Table 2 describes the characteristics of the 55 GPs. They were predominantly male with varied experience in primary care. Training

Table 2 Characteristics of general practitioners ($n=55$)

Sex	
Male	46 (84%)
Female	9 (16%)
Mean (SD) age (years)	44.3 (5.3)
Mean (SD) experience (years)	14.5 (6.6)
Practice type	
Office based (1–2 GPs)	40 (73%)
Health centre (>2 GPs)	15 (27%)
Urbanisation	
Rural	16 (29%)
Urban	27 (49%)
Mixed	12 (22%)
Personal list system	32 (59%)
Provider of GP training	26 (49%)

The questions on personal list system and provision of GPs training have one and two missing values, respectively.

practices seem to be overrepresented compared with the national average. There were no differences in characteristics between the 29 GPs in the intervention group and the 26 in the control group. Table 3 shows the characteristics of the two patient samples, pre- and post-intervention. The two groups did not differ although the pre-intervention group had a slightly higher frequency of attendance ($p=0.0001$). A large number (more than two thirds) in both groups reported a chronic illness.

Patients' evaluations of care are shown in table 4. Pre-intervention scores for both the intervention and control groups are comparable for most dimensions, except for the CEP dimension "availability for emergencies" where patients in the intervention group had more positive evaluations of care ($p<0.05$). Model A, which focused on differences between pre- and post-intervention measurements, found only one significant effect of group allocation: patients' evaluations of "medical care" improved in the control group by 0.09 but in the intervention group by only 0.01 ($p=0.0305$).

The regression models which used post-intervention scores as dependent variables found that the proportion of explained variance in the patient evaluation scores determined solely by the GP (model B) was highest for CEP dimensions "premises" and "organisation

Table 3 Characteristics of patients pre- and post-intervention

	Pre-intervention ($n=3691$)	Post-intervention ($n=3595$)
Sex		
Male	1137 (31%)	1172 (33%)
Female	2542 (69%)	2390 (67%)
Mean (SD) age (years)	49.7 (15.8)	50.1 (16.5)
Education		
None/primary	404 (11%)	462 (13%)
Low	867 (24%)	922 (26%)
Medium	1330 (36%)	1202 (34%)
High	681 (19%)	672 (19%)
Unclassified	372 (10%)	305 (9%)
Chronic illness	2597 (70%)	2571 (72%)
Mean (SD) visits in past 2 months	2.4 (1.8)	2.2 (1.8)
Overall health status		
Excellent	230 (6%)	218 (6%)
Very good	509 (14%)	479 (14%)
Good	1962 (55%)	2006 (57%)
Reasonable	818 (23%)	729 (21%)
Bad	66 (2%)	68 (2%)

The questions have varying numbers of missing values.

Table 4 Patients' evaluations of care before and after feedback

CEP dimension	Pre-intervention		Post-intervention	
	Feedback group (n=1917)	Control group (n=1774)	Feedback group (n=1766)	Control group (n=1826)
Organisation of appointments	4.42	4.49	4.43	4.55
Availability for emergencies	4.26*	4.09*	4.18	4.17
Premises	3.88	3.90	3.88	4.03
Continuity	4.62	4.72	4.60*	4.77*
Cooperation	4.35	4.27	4.43	4.38
Medical care	4.68	4.62	4.68	4.71
Relation and communication	4.66	4.61	4.67	4.68
Information and advice	4.67	4.65	4.68	4.71
Support	4.30	4.25	4.31	4.33

Mean score per dimension (scale 1–6). The scores have varying numbers of missing values. * $p < 0.05$ for difference between groups at pre-intervention or post-intervention in model C of the multilevel analysis (the “straightforward intervention effect”).

of appointments” (17.5% and 10.7%, respectively). The explained variances were lower for the dimensions “continuity” (7.2%) and “cooperation” (6.3%). Variances for the other dimensions were low (2.5–4.2%). Inclusion of group allocation (intervention *v* control) as a potential predictor (model C) revealed that patients' evaluations of “continuity” were less positive in the intervention group than in the control group ($p = 0.0236$). This effect was slightly modified in the model that included pre-intervention scores as predictors (model D) but was in the same direction ($p = 0.0479$). Furthermore, after correcting for pre-intervention scores, patients' evaluations of “medical care” were less positive in the intervention group than in the control group ($p = 0.0462$).

The number of patients in whom improvements were expected as GPs reported changes in their professional performance varied between 100 for “cooperation” and 1393 for “organisation of services” out of the 3995 post-intervention patients. However, model E, which explored whether patients' evaluations changed when GPs reported actions, did not show any significant effect.

Discussion

This study challenges the assumption that the feedback of patients' evaluations of care in general practice is an effective method for quality improvement. The study showed that the feedback hardly influenced the patients' evaluations, which is consistent with other studies on feedback of patient based outcomes.¹⁸ This finding is similar to that of other studies on the effects of feedback on clinical topics such as test ordering or preventive care with respect to change in clinical performance.^{19–20} No effects on patients' evaluations of care were found even for aspects of care where the GP reported having taken action to improve the situation.

Nevertheless, many GPs in the intervention group reported changes in their professional performance and the organisation of care. There was clearly a gap between GPs' actions to improve professional performance and observable changes in patients' evaluations of care. Changing the evaluation of care by patients may have been difficult as the pre-intervention evaluations were already very positive and the follow up period may have

been too short for such improvements to have an impact on the scores. The intervention may not have been strong enough since the feedback had the character of a “screening test” on a wide range of aspects of care. The intervention and the working conditions of the GPs were not standardised, and the uncontrolled variables may have influenced the findings. Other types of questionnaires such as those with more scope to provide detailed comments may be more effective. Acting upon this feedback was voluntary. Although a more “directive” intervention might be more effective, it would probably have been less acceptable and therefore not used at all. It is also possible that the measurement instrument (CEP) was not sufficiently responsive to the changes.

External factors might also explain the fact that few changes were found. Although patients are free to choose and change their GP, there is a shortage of GPs so that the actual choice is limited. Patients therefore may be less inclined to criticise their GP, and GPs may be less inclined to respond to patient criticism. Many GPs in the Netherlands have a heavy workload due to many factors including the continuing transfer of patients from secondary to primary care and the GPs' hesitancy to delegate tasks to practice nurses and other care providers. This heavy workload is a barrier to the implementation of changes in physician performance. In future studies we will examine the process of change in the behaviour of GPs in more detail in order to understand better the barriers which prevent GPs learning from their patients' views.

The findings of this study can probably be generalised to other settings in which primary care physicians work in office based practices and possibly also to many other healthcare settings. A crucial feature of our study population may be the fact that we included experienced physicians who have developed their routines over many years. It is probably easier to change the performance of trainees or residents. It is probably also significant that Dutch GPs do not have to compete with other care providers, which is not the case in some other countries. In the Netherlands GPs are “gate keepers” to secondary care and receive a fixed reimbursement based on the number of patients registered. Furthermore, there is a growing shortage of GPs.

Feedback on patients' evaluations of care could be an effective method of improving the quality of care. We believe that it is more effective if it is embedded in an educational programme or quality improvement activity related to a specific clinical topic or patient group. Combinations of interventions are more effective than single interventions—for instance, feedback combined with small group education can change the performance of physicians.²¹ Alternatively, it can be seen as a method of identifying possible topics for quality improvement activities which need to be explored in more detail before specific interventions to improve the situation can be

applied. With this approach it may not be necessary to achieve changes in patients' evaluations of care as a result of the feedback itself. The literature on implementation of innovations also suggests that barriers for change should be identified and interventions to induce changes in behaviour should address these barriers adequately.²¹

Feedback on patients' evaluations of care is, of course, only one possible approach to patient involvement in health care. Other approaches include, for instance, questionnaires to assess patients' needs before they consult a clinician, complaint procedures, and focus group interviews with patients to elicit their priorities on health care. The literature on patient involvement in health care tends to focus on conceptual and measurement issues rather than on achieving actual improvements in clinical practice. Thousands of descriptive patient satisfaction surveys have been published. However, it is important to evaluate methods of involving patients with respect to their effectiveness and feasibility before they are used on a wider scale as a tool for quality improvement.²² This study aimed to deliver this type of evidence and its conclusion is that patients' views alone do not necessarily lead to an improvement in the quality of care.

EV and MW designed the study and wrote the paper; EV developed the intervention and recruited the family physicians; RG supervised the project and participated in discussions on the design of and reports on the project; Ms P van Montfort and Ms N Verheijen coordinated the data collection and provided practical assistance in all phases of the study; Ir R P Akkermans performed the statistical analyses.
Funding: Dutch Organisation for Scientific Research (NWO).
Competing interests: None

- 1 Lewis JR. Patient views on quality care in primary care: literature review. *Soc Sci Med* 1994;39:655-70.
- 2 Scott A, Smith R. Keeping the customer satisfied: issues in the interpretation and use of patient satisfaction surveys. *Int J Qual Health Care* 1995;6:353-9.
- 3 Hall JA, Milburn MA, Epstein AM. A causal model of health status and satisfaction with medical care. *Med Care* 1993;31:84-94.

- 4 Wensing M, Grol R, Van Weel C, et al. Does the health status of chronically ill patients predict their judgements of the quality of primary care care? *Quality of Life Res* 1997;6:293-9.
- 5 Cleary PD. The increasing importance of patient surveys. Now that sound methods exist, patient surveys can facilitate improvement. *BMJ* 1999;319:720-1.
- 6 Tasa K, Baker R, Murray M. Using patient feedback for quality improvement. *Qual Manag Health Care* 1996;4:55-67.
- 7 Fitzpatrick R. Surveys of patient satisfaction: I. Important general considerations. *BMJ* 1991;302:887-9.
- 8 Hearnshaw H, Baker R, Cooper A, et al. The costs and benefits of asking patients for their opinions about primary care. *Fam Pract* 1996;13:52-8.
- 9 Adamson TE, Tschann JM, Gullion DS. Patient feedback as a tool to influence physician counseling. *Patient Educ Counsel* 1988;11:109-17.
- 10 Brody DS. Feedback from patients as a means of teaching non-technical aspects of medical care. *J Med Educ* 1980;55:34-41.
- 11 Cope DW, Linn LS, Leake BD, et al. Modification of residents' behavior by preceptor feedback of patient satisfaction. *J Gen Intern Med* 1986;1:394-8.
- 12 Falvo D. Patient perception as a tool for evaluation and feedback in family practice resident training. *J Fam Pract* 1980;10:471-4.
- 13 Greco MJ, Brownlea AA, McGovern JM. Utilising patient feedback in the RACGP training program. An exploratory study. *Aust Fam Phys* 1995;24:1077-81.
- 14 Niles N, Tarbox G, Schultz W, et al. Using qualitative and quantitative patient satisfaction data to improve the quality of cardiac care. *J Qual Improv* 1996;22:323-35.
- 15 Stillman PL, Burpeau-DiGregorio MY, Nicholson GI, et al. Six years of experience using patient instructors to teach interviewing skills. *J Med Educ* 1983;58:941-6.
- 16 Wensing M, Grol R, Van Weel C, et al. Quality assessment using patients' evaluations of care. *Eur J Gen Pract* 1998;4:155-8.
- 17 Bero L, Freemantle N, Grilli R, et al. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote implementation of research findings by health care professionals. *BMJ* 1998;316:465-8.
- 18 Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Evaluation Clin Pract* 1999;5:401-16.
- 19 Thomson MA, Oxman AD, Davis DA, et al. Audit and feedback to improve health professional practice and health care outcomes. In: *Cochrane Collaboration*. Cochrane Library. Issue 3. Oxford: Update Software, 1998.
- 20 Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *BMJ* 1991;303:398-402.
- 21 Wensing M, Van der Weijden T, Grol R. Implementing guidelines and innovations in primary care: which interventions are effective? *Br J Gen Pract* 1998;48:991-7.
- 22 Wensing M. Evidence-based patient empowerment. *Quality in Health Care* 2000;9:200-1.