

Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only

Wim H. J. M. Verstappen, MD,
PhD^{1,2}

Trudy van der Weijden, MD, PhD¹

Willy I. Dubois, MSc[†]

Ivo Smeele, MD, PhD²

Jan Hermesen, MD³

Frans E. S. Tan, PhD⁴

Richard P. T. M. Grol, PhD⁴

¹Centre for Quality of Care Research (WOK), Care and Public Health Research Institute (CAPHRI), and Department of Primary Care, Maastricht University, The Netherlands

²Centre for Diagnostics and Consultation, St Jans Hospital, Weert, The Netherlands

³Centre for Diagnostics and Consultation, Elkerliek Hospital, Helmond, The Netherlands

⁴Medical Diagnostic Centre, Canisius-Wilhelmina Hospital, Nijmegen, The Netherlands

⁵Department of Methodology and Statistics, Maastricht University, The Netherlands

[†]Deceased.



Conflicts of interest: none reported

CORRESPONDING AUTHOR

Wim Verstappen, MD, PhD
Department of Primary Care
Maastricht University
PO Box 616
6200 MD Maastricht
The Netherlands
wim.verstappen@hag.unimaas.nl

ABSTRACT

PURPOSE We wanted to evaluate the added value of small peer-group quality improvement meetings compared with simple feedback as a strategy to improve test-ordering behavior. Numbers of tests ordered by primary care physicians are increasing, and many of these tests seem to be unnecessary according to established, evidence-based guidelines.

METHODS We enrolled 194 primary care physicians from 27 local primary care practice groups in 5 health care regions (5 diagnostic centers). The study was a cluster randomized trial with randomization at the local physician group level. We evaluated an innovative, multifaceted strategy, combining written comparative feedback, group education on national guidelines, and social influence by peers in quality improvement sessions in small groups. The strategy was aimed at 3 specific clinical topics: cardiovascular issues, upper abdominal complaints, and lower abdominal complaints.

The mean number of tests per physician per 6 months at baseline and the physicians' region were used as independent variables, and the mean number of tests per physician per 6 months was the dependent variable.

RESULTS The new strategy was executed in 13 primary care groups, whereas 14 groups received feedback only. For all 3 clinical topics, the decrease in mean total number of tests ordered by physicians in the intervention arm was far more substantial (on average 51 fewer tests per physician per half-year) than the decrease in mean number of tests ordered by physicians in the feedback arm ($P = .005$). Five tests considered to be inappropriate for the clinical problem of upper abdominal complaints decreased in the intervention arm, with physicians in the feedback arm ordering 13 more tests per 6 months ($P = .002$). Interdoctor variation in test ordering decreased more in the intervention arm.

CONCLUSION Compared with only disseminating comparative feedback reports to primary care physicians, the new strategy of involving peer interaction and social influence improved the physicians' test-ordering behavior. To be effective, feedback needs to be integrated in an interactive, educational environment.

Ann Fam Med 2004;2:569-575. DOI: 10.1370/afm.244.

INTRODUCTION

Numbers of tests ordered by primary care physicians are rising in many countries, and interdoctor variation has been found to be large, whereas according to established guidelines, many of these tests can be regarded as unnecessary.¹⁻³ It is as yet unclear, however, what would be the best method to influence physicians' test-ordering behavior. Several studies evaluating different types of interventions to change this

behavior have so far reported heterogeneous results. Among these widely investigated strategies, one that has met with mixed results is feedback.⁴⁻⁷ Many authorities in Western countries, such as health insurers, regularly disseminate reports about test ordering, prescription, or referral rates to physicians or practices, often without substantial impact.^{8,9}

The literature shows that multifaceted strategies are generally superior to single methods when it comes to influencing behavior.¹⁰⁻¹² Success rates of specific strategies seem to be strongly influenced by the extent to which they fit in with the local and organizational context and the physicians' day-to-day work routine.^{13,14} Favorable experience has been gained in particular with small-group education and interactive quality improvement sessions for primary care physicians.^{15,16} We therefore decided to develop a multifaceted strategy, combining transparent, individual graphic feedback on test-ordering routines, education on clinical guidelines for test ordering, and small-group quality improvement meetings among primary care physicians. At these meetings, test-ordering behavior and changes in routines were discussed, using social influence and peer influence as important motivators for change. Social influence from respected colleagues or opinion leaders seems to have a greater effect on practice routines than do traditional medical education activities focusing on changing professional cognitions or attitudes.¹⁷⁻²¹ We therefore expected our strategy to be useful, because it is closely linked to the everyday setting for many physicians, who tend to work more or less in isolation and have limited contact with peers about topics like test-ordering behavior.

We hypothesized that greater insights into and discussion of the physicians' own performance in a safe group of respected colleagues would be a powerful instrument to improve the quality of test ordering. Because classic feedback is increasingly used as a routine quality improvement strategy, and because this simple and cheap strategy might suffice, we further hypothesized that this innovative, multifaceted strategy would have an added value relative to standardized feedback only. In one multicenter randomized trial with a block design, this strategy was indeed found to have a favorable effect on physicians' test-ordering behavior.²² A cost analysis of the new strategy and a process evaluation showed that it was a cost-efficient and feasible tool for improvement of physicians' test-ordering behavior.^{23,24} This article determines the effects of this innovative, multifaceted strategy, compared with a classic feedback strategy, to assess the added value of the small-group quality improvement meetings.

METHODS

Overall Design and Population

The complete trial consisted of 3 arms. The comparison between 2 arms to assess the clinical relevance of our strategy was described in the *Journal of the American Medical Association*.²² The study reported here compared one of the arms receiving the entire strategy (which was also described in the *JAMA* article) with the third arm that received a feedback intervention strategy. This multicenter randomized-controlled trial was conducted during the first 6 months of 1999 in 5 health care regions, each with a diagnostic center. A diagnostic center is an institute, usually associated with a hospital, where primary care physicians can order laboratory, imaging, and function tests. All 5 diagnostic centers used nationally developed indication-oriented forms for laboratory orders. Because they made use of one of these 5 diagnostic centers, 37 local primary care practice groups, with a total of 294 primary care physicians, were eligible for participation.

Local teams of primary care physicians collaborating in a specific region are a common feature of Dutch primary care. Every physician working in a solo, 2-person, group, or health center practice in the Netherlands is a member of a local primary care practice group. Continuous medical education (for example, quality improvement meetings about prescribing) is an important activity in most groups. One task of a diagnostic center medical coordinator is to give feedback to these physicians on their test-ordering behavior, and the medical coordinators are considered to be opinion leaders concerning test ordering. From May 1998 until September 1998 the coordinators of the 5 diagnostic centers recruited local primary care groups in their regions to participate in the trial.

Intervention

The new strategy consisted of the following elements: (1) personalized graphic feedback, including a comparison of each physician's own data with those of colleagues; (2) dissemination of and education on national, evidence-based guidelines; and (3) continuous quality improvement meetings in small groups. The improvement strategy concentrated on 3 specific clinical topics (cardiovascular conditions, upper abdominal complaints, and lower abdominal complaints) and the tests used for these clinical problems, because it was believed that the physicians would prefer to discuss specific clinical topics rather than specific tests (Table 1). During the first half of 1999 each physician received 3 different feedback reports on these 3 clinical problems by mail, together with concise information on the evidence-based clinical guidelines for these specific clinical topics developed by the Dutch College of Primary Care Physicians (an

Table 1. Clinical Problems and Associated Tests Used in the Trial

Clinical Problems	Tests
Cardiovascular conditions	Cholesterol, subfractions, potassium, sodium, creatinine, BUN, ECG (exercise)
Lower abdominal complaints	Prostate-specific antigen, C-reactive protein, ultrasound scan of the kidney, intravenous pyelogram, double-contrast barium enema, sigmoidoscopy
Upper abdominal complaints	ALT, AST,* LDH*, amylase,* γ -glutamyltransferase, bilirubin,* alkaline phosphatase,* ultrasound scan of hepatobiliary tract

BUN = blood urea nitrogen; ECG = electrocardiogram; ALT = alanine aminotransferase; AST = aspartate aminotransferase; LDH = lactic dehydrogenase.

*Tests that are inappropriate according to national evidence-based guidelines on upper abdominal complaints (see Supplemental Appendix).

example of a feedback report is available online only as a supplemental figure at: <http://www.annfammed.org/cgi/content/full/2/6/569/DC1>).

Each postal contact was followed by a 90-minute standardized small-group quality improvement meeting about 2 weeks later at which one of the clinical problems was discussed based on the feedback reports and the guidelines (Figure 1). In these meetings social influence, which was an important vehicle to reach improvement on test ordering, consisted of the following major components. The first was mutual personal feedback by peers, who worked in pairs at the start of the meeting. The second component was interactive group education in which national guidelines were related to the individual physician's actual test-ordering behavior and an effort to reach a group consensus on the optimal test-ordering behavior. The third was the development of individual and group plans for change to stimulate the physicians to put their plans into daily practice. As a critical follow-up, achieving the goals of these plans was discussed at the next meeting. The medical coordinators disseminated the feedback reports and organized and supervised the quality improvement meetings. That the medical coordinators were respected regional opinion leaders concerning test-ordering behavior was an additional important component in the social influence strategy.

The first was mutual personal feedback by peers, who worked in pairs at the start of the meeting. The second component was interactive group education in which national guidelines were related to the individual physician's actual test-ordering behavior and an effort to reach a group consensus on the optimal test-ordering behavior. The third was the development of individual and group plans for change to stimulate the physicians to put their plans into daily practice. As a critical follow-up, achieving the goals of these plans was discussed at the next meeting. The medical coordinators disseminated the feedback reports and organized and supervised the quality improvement meetings. That the medical coordinators were respected regional opinion leaders concerning test-ordering behavior was an additional important component in the social influence strategy.

The first was mutual personal feedback by peers, who worked in pairs at the start of the meeting. The second component was interactive group education in which national guidelines were related to the individual physician's actual test-ordering behavior and an effort to reach a group consensus on the optimal test-ordering behavior. The third was the development of individual and group plans for change to stimulate the physicians to put their plans into daily practice. As a critical follow-up, achieving the goals of these plans was discussed at the next meeting. The medical coordinators disseminated the feedback reports and organized and supervised the quality improvement meetings. That the medical coordinators were respected regional opinion leaders concerning test-ordering behavior was an additional important component in the social influence strategy.

Design and Measurements

The physicians gave informed consent for the retrieval of anonymous data on the numbers

of tests ordered. To avoid seasonal influences, the numbers of tests for effect evaluation were assessed for the last 6 months of 1998 (the baseline period) and the last 6 months of 1999 (the follow-up period). The strategies were evaluated in a multicenter randomized controlled trial, consisting of 2 arms, with the local primary care practice group as the unit of randomization (Figure 2). After stratification for region and group size, randomization was performed centrally with Duploran, a random numbers program. (Department of Epidemiology, Maastricht University, F. Kessels, methodologist). For the intervention arm the local primary care practice groups received the entire intervention, whereas for the feedback arm the local practice groups received only the feedback reports of their test-ordering behavior for the same clinical problems.

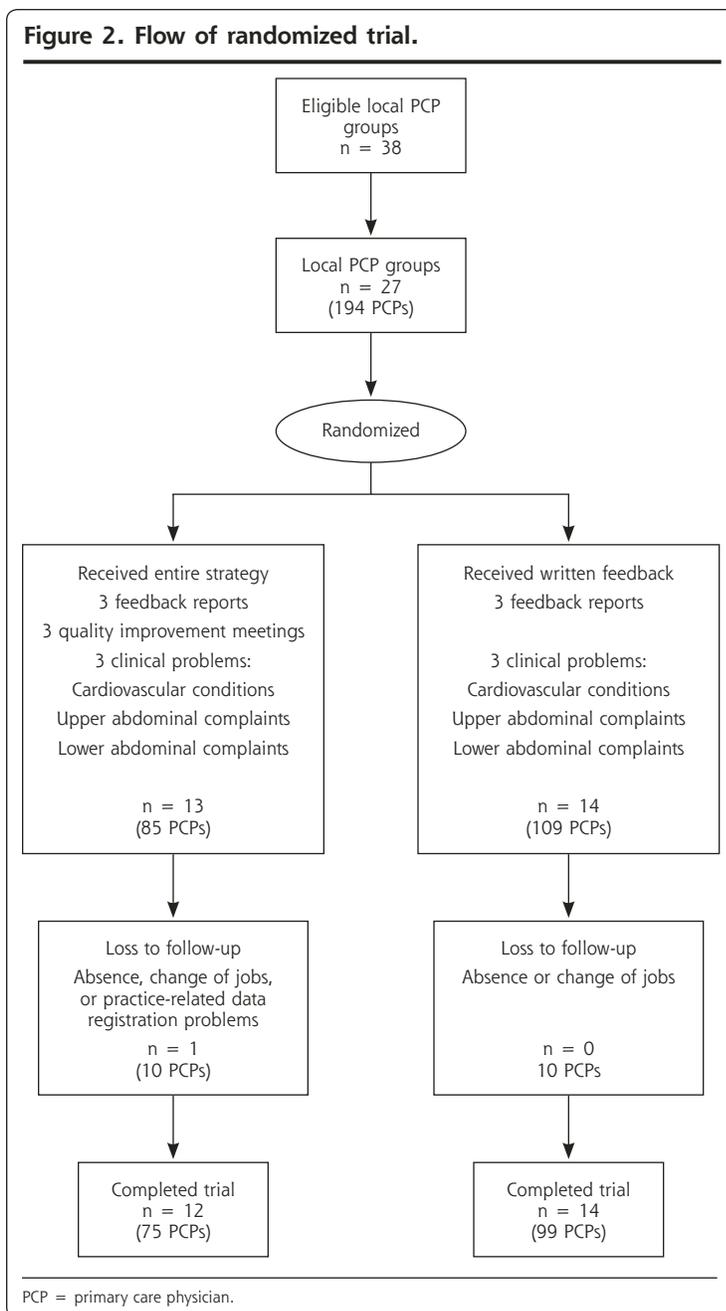
Effect Measures and Measuring Instruments

Characteristics of primary care physicians and local practice groups were collected by a written questionnaire. To evaluate intervention effects, the following effect measures were defined:

1. The total number of requested tests per 6 months per physician for the 3 clinical problems in total and per clinical problem (a decrease in the numbers of tests was considered as better patient care, consistent with national, evidence-based guidelines for test ordering for the included clinical problems)
2. Reduced interdoctor variation in the numbers of test ordered (considered to represent an improvement in performance)
3. The effects on total numbers and on defined inappropriate upper abdominal tests for one specific problem—upper abdominal complaints

Figure 1. Structure of the 90-minute small-group quality improvement meeting.

	5 min	Explaining the method, looking back on the last meeting
	5 min	Critical look at own numbers
	5 min	Pairwise talk about interindividual differences, and explain them to each other
	25 min	Plenary discussion: relate the test-ordering performance to the guidelines
	10 min	Pairwise talk about difficulties in changing test-ordering behavior
	25 min	Plenary discussion on difficulties, resistance to change, learn from each other
	15 min	Making individual and group work agreements, evaluation

Figure 2. Flow of randomized trial.

Statistical Analysis

Differences on individual physician characteristics were tested for significance with the Pearson's chi-square test. To evaluate intervention effects, the unit should be the local primary care practice group, because that group was the unit of randomization. A 3-level model was used with the practice group as level 3, physicians as level 2, and numbers of tests as level 1. This model was analyzed using SAS PROC MIXED Release 8.2 (SAS Institute, Cary, NC). Power calculations based on the baseline data showed that each arm needed

approximately 85 physicians to detect a 10% difference in mean total numbers of tests with 80% power, and a risk of type 1 error of .05. The region appeared to be an important determinant of the between-group variance and was used as independent variable, together with the baseline numbers of tests. All effects were analyzed with analyses of covariance. This regression equation gives the intervention effect β from which the follow-up numbers of tests are the dependent variable and the baseline numbers and the region are the independent variables. β reflects the total change between mean numbers of tests in the intervention arm minus the total change between baseline and follow-up in mean numbers of tests in the feedback arm, adjusted for baseline and region.

Interdoctor variation was calculated by the coefficient of variance, the standard deviation (SD) divided by the mean.

RESULTS

Twenty-seven local primary care practice groups, including 194 physicians, expressed their willingness to participate, so no further recruitment actions were needed. After randomization, the intervention arm included 13 local practice groups, whereas the feedback arm included 14 practice groups (Figure 2). Each physician received feedback on the 3 clinical problems.

Table 2 describes the characteristics of the study population. Mean group size in the intervention arm was 6.9 (SD 2.1) compared with 7.8 (SD 4.2) in the feedback arm. There was a large, but statistically insignificant difference in mean total numbers of tests per physician per 6 months between the 2 arms at baseline; the intervention arm had 478 (SD 309), the feedback arm had 541 (SD 337). An intention-to-treat analysis was not possible for 10 physicians in each arm, including one entire local practice group in the intervention arm. Data for the follow-up measurements for these physicians were lacking because of absence, change of jobs, or practice-related data registration problems. Multilevel analyses showed that the point estimation and standard deviation were the same at the group level as in the analysis of covariance at the individual physician level; therefore, no correction for local practice groups was needed.

Table 2. Study Population Characteristics at Individual Primary Care Physician Level

Characteristic	Intervention Arm	Feedback Arm
Number of physicians	85	109
Age, mean (SD), y	46.2 (6.6)	46.2 (6.6)
Female, No. (%)	14 (16)	11 (10)
Patients per physician, mean No. (SD)*	2,587 (641)	2,444 (416)
Patients > 65 y, mean % (SD)	15 (6.8)	15 (6.5)
Working time factor, % (SD)†	91 (15)	92 (12)
Physicians with a solo practice, No. (%)	43 (51)	44 (40)
Physicians who use computerized registration system, No. (%)	66 (78)	75 (69)

* Total practice population for whom the primary care physician is responsible.

† Working time factor, full time = 100% = 5 days; each half-day = 10%, so a physician with a part-time factor of 70% works 7 half-day periods.

Table 3 shows results of these analyses at the individual physician level for all tests and each clinical problem. The total number of tests ordered decreased in both arms. For the intervention group physicians, the decrease was 51 tests more per physician per half-year than for the feedback physicians ($P = .005$). The differences in changes were significant, except for cardiovascular conditions, which decreased with marginal significance. The data in the Supplemental Appendix

(available online only at: <http://www.annfammed.org/cgi/content/full/2/6/569/DC1>) describe the

intervention and its effects in more detail for the upper abdominal complaints. The differences for the defined inappropriate tests were also significant,

meaning that the intervention physicians ordered 13 fewer inappropriate tests than the feedback physicians per physician per half-year. ($P = .002$). Table 3 also shows that the coefficient of variance decreased more in the intervention arm, meaning that the variation in test ordering between physicians in the intervention group decreased more than they did in the feedback group. Figure 3 depicts graphically the results for all tests at aggregated local practice group level and shows that effects in the intervention arm were more straightforward.

DISCUSSION

A new interactive quality improvement strategy was evaluated and compared with classic feedback alone among 27 local primary care practice groups, including 194 physicians in 5 regions. The first success was the easy recruitment, with practice groups eager to participate in the trial. A considerable improvement in test-ordering behavior was found after 1 year of intervention. In the intervention arm, there was a statistically significant and clinically relevant decrease in numbers of tests ordered in keeping with the national evidence-based guidelines. The numbers of tests ordered for 2 clinical problems fell significantly, and a statistically significant reduction in the numbers of inappropriate tests for upper abdominal complaints was observed. During the intervention period, the guidelines on cholesterol testing were updated nationally, which might have been one reason why the decrease in numbers of cardiovascular tests was only marginally significant. Interdoctor varia-

Table 3. Effects of Strategy on the Mean (SD) Number of Tests and the Coefficient of Variance, per Primary Care Physician and per 6 Months

Study Subjects	Intervention Arm				Feedback Arm				β^\dagger	SE β	95% CI	P
	Baseline Mean (SD)	CV*	Follow-up Mean (SD)	CV*	Baseline Mean (SD)	CV*	Follow-up Mean (SD)	CV*				
Total number of tests	478 (309)	0.65	422 (235)	0.56	541 (337)	0.62	535 (309)	0.58	-51	17.94	-87 to -16	.005
Cardiovascular conditions	293 (189)	0.65	276 (157)	0.57	322 (214)	0.66	333 (205)	0.62	-25	13.08	-51 to 1	.056
Lower abdominal complaints	20 (20)	1.00	18 (19)	1.06	30 (40)	1.43	30 (27)	0.90	-6	2.18	-10 to -2	.008
Upper abdominal complaints	165 (125)	0.76	128 (82)	0.64	188 (143)	0.76	171(117)	0.68	-24	7.98	-40 to -8	.003
Inappropriate upper abdominal tests	55 (60)	1.09	39 (32)	0.82	60 (63)	1.05	56 (54)	0.96	-13	4.1	-22 to -5.2	.002

Note: analysis of covariance adjusted for baseline number of tests and the regions.

CV = coefficient of variance; CI = confidence interval.

* CV = SD / mean.

† β = intervention effect = the total change between baseline and follow-up of mean numbers of tests in Intervention arm less the total change of numbers between baseline and follow-up of mean numbers of tests in the feedback arm.

tion in numbers of tests ordered decreased in both arms, but more so in the intervention arm. The small-group quality improvement meetings successfully discussed the transparent feedback reports on test ordering and the national guidelines.²⁴ The personal interaction and mutual influence between colleagues implicitly resulted in an individual or group contract.^{21,24,25} The role of the medical coordinators as opinion leaders also seems a crucial element of the strategy.^{20,26} Questions can therefore be raised about the impact of written feedback reports in general if they are not integrated in a wider system of quality improvement. This lack of impact could have been the reason why Eccles and colleagues did not find any effect in their trial of feedback on test ordering.⁹

Some methodological comments can be made on our study. It is possible that only motivated, well-functioning groups of physicians participated, and it is therefore questionable whether the strategy would work for all groups. Despite the large numbers of participating physicians, a difference between the 2 study arms was found in baseline performance. This difference is probably due to chance, however, as the number of randomization objects was small ($n = 27$). Despite the smaller mean number of tests at baseline, the intervention arm succeeded in substantially decreasing the numbers of tests ordered. We did not include a nonintervention control arm; we did not consider doing so a relevant contrasting strategy because feedback is now a regularly used strategy in primary care in the Netherlands. Unfortunately, we could not use clinical data to evaluate

the effect, but because the evidence-based guidelines recommend a reduction in the total numbers of tests, the decrease we found can be interpreted as a quality improvement. Moreover, there is empirical evidence that a general reduction in test use in primary care does not lead to more referrals or substitution of care.^{27,28}

We expect that these limitations have had only minor impact on the results, and our results may yield 2 important conclusions. The first is that this new strategy can be a powerful innovative instrument to change physician test-ordering behavior. The strategy gives physicians an opportunity to discuss their test-ordering performance with colleagues on the basis of actual performance data, making the participants feel more committed to the agreements. Our strategy also seems worthwhile because small-group quality improvement meetings can help build a local practice group focusing on quality improvement. Many test-ordering problems that physicians encounter in everyday practice, such as demands for tests by patients and changing guidelines, can be discussed and may be solved in an open and respectful discussion among colleagues. The second conclusion is that merely sending feedback reports to physicians, without other activities, such as peer discussion or other strategies that fit in with everyday practice, does not have much impact on test-ordering behavior. More effort is needed, and feedback reports must fit in with a more ambitious continuous quality improvement program. Further, although our method was applied for test-ordering behavior, it also seems applicable to quality improvement for other issues, such as prescribing and referral behavior, and for other teams of collaborating physicians.

Nationwide implementation of this innovative strategy would be a logical next step and is now being prepared in the Netherlands.

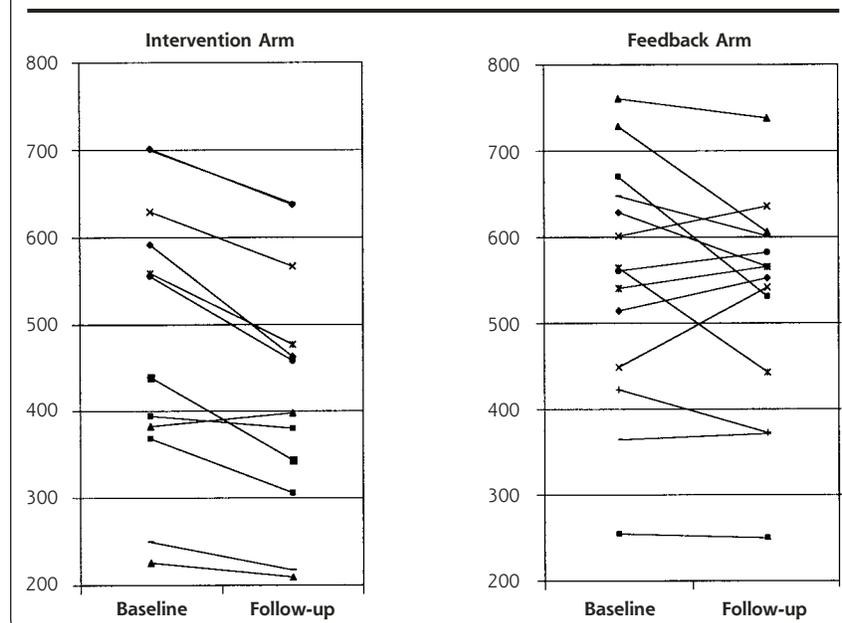
To read or post commentaries in response to this article, see it online at <http://www.annfam.org/cgi/content/full/2/6/569>.

Key words: Quality assurance, health care; test-ordering behavior; feedback; small-group quality improvement; quality of health care

Submitted January 20, 2004; submitted, revised, May 11, 2004; accepted May 19, 2004.

Funding support: The authors gratefully acknowledge the financial contribution to the study provided by the Dutch College for Health Insurances.

Figure 3. Baseline and follow-up measurements in mean total numbers of tests per 6 months at aggregated local practice group level for the 13 intervention and the 14 feedback local practice groups.



References

1. Leurquin P, Van Casteren V, De Maeseneer J. Use of blood tests in general practice: a collaborative study in eight European countries. Eurosentinel Study Group. *Br J Gen Pract.* 1995;45:21-25.
2. Kassiser JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *New Engl J Med.* 1989;320:1489-1491.
3. Ayanian JZ, Berwick DM. Do physicians have a bias toward action? A classic study revisited. *Med Decis Making.* 1991;11:154-158.
4. Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *BMJ.* 1991;303:398-402.
5. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA.* 1995;274:700-705.
6. Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* 1997; CD000259.
7. Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Audit and feedback versus alternative strategies: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* 2000; CD000260.
8. O'Connell DL, Henry D, Tomlins R. Randomised controlled trial of effect of feedback on general practitioners' prescribing in Australia. *BMJ.* 1999;318:507-511.
9. Eccles M, Steen N, Grimshaw J, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet.* 2001;357:1406-1409.
10. Wensing M, Grol R. Single and combined strategies for implementing changes in primary care: a literature review. *Int J Health Care.* 1994;6:115-132.
11. Wensing M, Van der Weijden T, Grol R. Implementing guidelines and innovations in general practice: which interventions are effective? *Br J Gen Pract.* 1998;48:991-997.
12. Grimshaw JM, Shirran L, Thomas R, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care.* 2001;39(8 Suppl 2):112-45.
13. Winkens RAG, Pop P, Bugter-Maessen AMA, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet.* 1995;498-502.
14. Tierney WM. Feedback of performance and diagnostic testing: lessons from Maastricht [editorial; comment]. *Med Decis Making.* 1996;16:418-419.
15. Grol R. Peer review in primary care. *Quality Assur Health Care.* 1990;2:219-226.
16. Grol R, Jones R. Lessons from 20 years of implementation research. *Fam Pract.* 2000;17:S32-S35.
17. Brady WJ, Hissa DC, McConnell M, Wones RG. Should physicians perform their own quality assurance audits? *J Gen Intern Med.* 1988;3:560-565.
18. Mittman B, Tonesk X, Jacobson P. Implementing clinical practice guidelines: social influence strategies and practitioner behavior change. *Qual Rev Bull.* 1992;18:413-422.
19. Lomas J, Enkin M, Anderson GM, Hannah WJ, Vayda E, Singer J. Opinion leaders vs audit and feedback to implement practice guidelines. Delivery after previous cesarean section. *JAMA.* 1991;265:2202-2207.
20. Thomson O'Brien MA, Oxman AD, Haynes RB, Davis DA, Freemantle N, Harvey EL. Local opinion leaders: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* 2000(2): CD000125.
21. Tausch BD, Harter MC. Perceived effectiveness of diagnostic and therapeutic guidelines in primary care quality circles. *Int J Qual Health Care.* 2001;13:239-246.
22. Verstappen WH, Van Der Weijden T, Sijbrandij J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians: a randomized trial. *JAMA.* 2003;289:2407-2412.
23. Verstappen WH, van Merode F, Grimshaw J, Dubois WI, Grol RP, van der Weijden T. Comparing cost effects of two quality strategies to improve test ordering in primary care: a randomized trial. *Int J Qual Health Care.* 2004;16:391-398.
24. Verstappen W, van der Weijden T, Dubois W, et al. Lessons learnt from applying an innovative small group quality improvement strategy to improve test ordering in general practice. *Qual Primary Care.* 2004;12:79-85.
25. Spooner A, Chapple A, Roland M. What makes British general practitioners take part in a quality improvement scheme? *J Health Serv Res Policy.* 2001;6:145-150.
26. Borbas C, Morris N, McLaughlin B, et al. The role of clinical opinion leaders in guideline implementation and quality improvement. *Chest.* 2000;118(2 Suppl):24s-32s.
27. Winkens RA, Grol RP, Beusmans GH, Kester AD, Knottnerus JA, Pop P. Does a reduction in general practitioners' use of diagnostic tests lead to more hospital referrals? *Brit J Gen Pract.* 1995;45:289-292.
28. Kaag ME, Wijkkel D, de Jong D. Primary health care replacing hospital care--the effect on quality of care. *Int J Qual Health Care.* 1996;8:367-373.