Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians
A Randomized Trial

Wim H. J. M. Verstappen, MD
Trudy van der Weijden, MD, PhD
Jildou Sijbrandij, MSc
Ivo Smeele, MD, PhD
Jan Hermsen, MD
Jeremy Grimshaw, PhD
Richard P. T. M. Grol, PhD

Context Numbers of diagnostic tests ordered by primary care physicians are growing and many of these tests seem to be unnecessary according to established, evidence-based guidelines. An innovative strategy that focused on clinical problems and associated tests was developed.

Objective To determine the effects of a multifaceted strategy aimed at improving the performance of primary care physicians' test ordering.

Design Multicenter, randomized controlled trial with a balanced, incomplete block design and randomization at group level. Thirteen groups of primary care physicians underwent the strategy for 3 clinical problems (arm A; cardiovascular topics, upper and lower abdominal complaints), while 13 other groups underwent the strategy for 3 other clinical problems (arm B; chronic obstructive pulmonary disease and asthma, general complaints, degenerative joint complaints). Each arm acted as a control for the other.

Setting: Primary care physician groups in 5 regions in the Netherlands with diagnostic centers recruited from May to September 1998.

Study Participants: Twenty-six primary care physician groups, including 174 primary care physicians.

Intervention During the 6 months of intervention, physicians discussed 3 consecutive, personal feedback reports in 3 small group meetings, related them to 3 evidence-based clinical guidelines, and made plans for change.

Main Outcome Measure According to existing national, evidence-based guidelines, a decrease in the total numbers of tests ordered per clinical problem, and of some defined inappropriate tests, is considered a quality improvement.

Results For clinical problems allocated to arm A, the mean total number of requested tests per 6 months per physician was reduced from baseline to follow-up by 12% among physicians in the arm A intervention, but was unchanged in the arm B control, with a mean reduction of 67 more tests per physician per 6 months in arm A than in arm B ($P=.01$). For clinical problems allocated to arm B, the mean total number of requested tests per 6 months per physician was reduced from baseline to follow-up by 8% among physicians in the arm B intervention, and by 3% in the arm A control, with a mean reduction of 28 more tests per physician per 6 months in arm B than in arm A ($P=.22$). Physicians in arm A had a significant reduction in mean total number of inappropriate tests ordered for problems allocated to arm A, whereas the reduction in inappropriate test ordered physicians in arm B for problems allocated to arm B was not statistically significant.

Conclusion In this study, a practice-based, multifaceted strategy using guidelines, feedback, and social interaction resulted in modest improvements in test ordering by primary care physicians.

JAMA. 2003;289:2407-2412 www.jama.com

In many countries, the number of diagnostic tests ordered by primary care physicians is growing, while according to established evidence-based guidelines, many of these tests are seen as unnecessary.[1-3] Possible explanations are test ordering routines that are difficult to change, a more defensive attitude among primary care physicians out of fear of medical errors, or a lack of knowledge about the appropriate use of tests.[4-7] Moreover, patients more actively ask for tests and often attach greater value to test results than is justified by the facts.[8,9] Unfortunately, little is yet known about the negative effects of performing such tests, in terms of, for example, unnecessary exposure to radiation or false-positive results, that may induce fear and anxiety in patients or may result in a cascade of unnecessary further testing.

Given these problems it is challenging to learn how to change test ordering performance effectively and bring it into line with existing evidence or guidelines on optimal testing. Many such attempts have been made with mixed results, showing that successful strategies require a well-balanced

Author Affiliations are listed at the end of this article.
Corresponding Author and Reprints: Wim Verstappen, MD, Department of General Practice, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands (e-mail: wim.verstappen@hag.unimass.nl).
Toward Optimal Laboratory Use Section Editor: David H. Mark, MD, MPH, Contributing Editor.

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

Clinical Problems	Tests
Arm A groups	
A1: cardiovascular diseases/ hypertension	Cholesterol, subfractions, potassium, sodium, creatinine, ECG (exercise), BUN*
A2: upper abdominal complaints	SGPT, γ-glutamyltransferase, ultrasound scans of hepatobiliary tract, SGOT,* LDH,* amylase,* bilirubin,* alkaline phosphatase*
A3: lower abdominal complaints	Prostate-specific antigen, CRP, ultrasound of the kidney, IVP, double-contrast barium enema, sigmoidoscopy
Arm B groups	
B1: COPD/asthma	Allergic screening test, chest radiograph, immunoglobulin E*
B2: general malaise/fatigue/ vague complaints	ESR, Hb with or without indices, Ht, TSH, Monospot, leukocyte count*
B3: degenerative joint complaints	ESR, uric acid, rheumatoid factors, radiographs of lumbar spine,* cervical spine,* shoulder,* knee,* hip*

Abbreviations: BUN, blood urea nitrogen; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; ECG, electrocardiogram; ESR, erythrocyte sedimentation rate; Hb, hemoglobin; Ht, hematocrit; IVP, intravenous pyelogram; LDH, lactic dehydrogenase; SGOT, serum glutamic-oxaloacetic transaminase; SGPT, serum glutamatic-pyruvate transaminase; TSH, thyroid-stimulating hormone.
*Tests that are inappropriate according to the national evidence-based guidelines.[1-3]

combination of interventions.[10-12] We have developed a multifaceted strategy combining personal feedback and guideline dissemination with quality meetings in small groups of primary care physicians. Social interactions were used as an important motivator for change, as physicians learned how colleagues were handling test ordering problems and as they obtained information about the consequences of medical decision making in daily practice.[13,14] The aim of this strategy was to achieve sustained improvements in test ordering, for example, working in line with the national, evidence-based guidelines. The present article describes the changes in test ordering performance resulting from this innovative strategy in a large population of primary care physicians.

METHODS

Setting and Population

Our study was conducted in 5 regions in the Netherlands, each of which made use of the services of a diagnostic center. A diagnostic center is an institute, usually associated with a hospital, where primary care physicians can order tests without referring patients to the hospital. Thirty-seven local groups of primary care physicians linked to 1 of these 5 diagnostic centers were eligible for the study. These groups are a common feature of Dutch general practice, involving teams of primary care physicians collaborating in a specific region. These teams share patient care outside office hours and many of them also engage in continuing medical education. From May until September 1998 the coordinators of the 5 diagnostic centers recruited local groups in their regions to participate.

Intervention

The strategy consisted of the following elements: personalized graphical feedback, including a comparison of each physician's own data with those of colleagues; dissemination of national, evidence-based guidelines; and regular meetings on quality improvement in small groups. The strategy focused on specific clinical problems and the diagnostic tests used for these problems (TABLE 1). These tests covered about 90% of all tests a primary care physician can order in a diagnostic center. For the tests used in the trial, national guidelines for optimal test ordering had to be available.

During the first 6 months of 1999, each of the recruited physicians received by mail 3 consecutive feedback reports on 3 different clinical problems, together with concise information on the 3 evidence-based clinical guidelines for these problems, developed by the Dutch College of Primary Care Physicians.

Each postal contact was followed by a 90-minute standardized small group quality improvement meeting about 2 weeks later, supervised by the medical coordinator of the diagnostic center. At the 3 meetings, physicians were asked to discuss and compare their feedback reports with colleagues and to relate them to the national guidelines. They also discussed Bayesian decision rules to help them understand the probability of false-positive results in low-prevalence disorders. Another important topic of debate was how to deal with the frequent requests by patients to have inappropriate tests performed. This discussion of the guidelines was followed by a thorough discussion of the difficulties of achieving changes at the individual primary care physician level, the practice level, or at the patient level. The next step was to try to implement the guidelines in their own practice, and at the end of each session, plans were drawn up for change, both at individual and group level. Subsequent meetings were used to evaluate whether targets had been met.

Design and Measurements

The effect of the intervention was evaluated in a multicenter, randomized controlled trial that was conducted in the first 6 months of 1999 with a balanced, incomplete block design, consisting of 2 arms, with the local group of primary care physicians as the unit of randomization. (FIGURE) One group of local groups (arm A) underwent the strategy with respect to tests associated with the 3 clinical problems allocated to arm A (Table 1), while the other group of local groups (arm B) underwent the strategy with respect to tests associated with the 3 problems allocated to arm B (Table 1). The groups in arm A acted as blind controls for the groups undergoing the arm B intervention, and vice versa. This rigorous design was used to balance the influence of nonspecific effects on the test ordering performance between the 2 arms and to neutralize the Hawthorne effect, that is, the effect that physicians might change their test ordering because they

were aware of taking part in a trial.[15,16] After stratification for region and group size, randomization was performed centrally with Duploran, a random numbers program.

The physicians gave informed consent for the retrieval of anonymous data on the numbers and results of all tests ordered. To avoid seasonal influences, the numbers of tests for effect evaluation were assessed during the last 6 months of 1998 (the baseline period) and the last 6 months of 1999 (the follow-up period).

Intervention Effect Measures

Characteristics of primary care physicians and local groups were collected by means of a written questionnaire. Two effect measures were used to evaluate intervention effects:

1. A decrease in the total numbers of requested tests per 6 months per physician: since most of the recommendations in the national, evidence-based guidelines advise ordering fewer tests, a decrease in the total numbers of tests ordered was regarded as an improvement in patient care. Separate analyses were performed for the 6 different clinical problems.

2. A decrease in the numbers of inappropriate tests as defined in the guidelines (Table 1 and BOX): these tests were regarded as inappropriate for the associated clinical problems for various reasons, for example, because the results of these tests seldomly have an influence on the treatment, because the high likelihood of false-positive results can occur, because better alternatives are available, or because adverse effects to some tests can occur (eg, radiology tests).

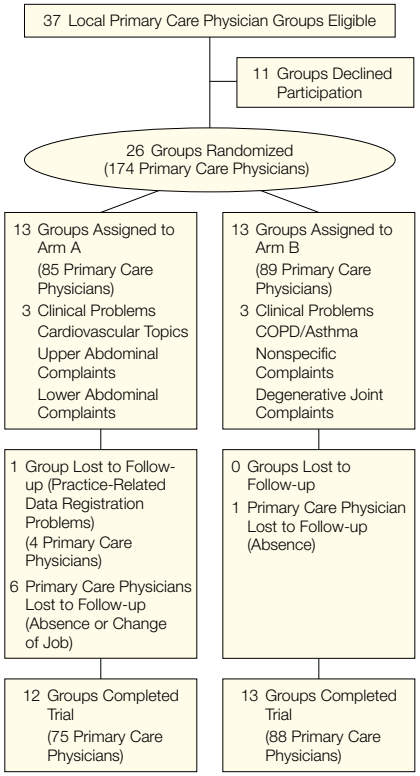
Statistical Analysis

Differences in individual characteristics of the primary care physician were tested for significance with Pearson χ^2 test. In the evaluation of intervention effects, the unit had to be the local group of primary care physicians because this unit was also the unit of randomization. To account for clustering within local groups, a 3-level model was used with the local group as level 3, individual physicians as level 2, and numbers of tests as level 1. The analysis was carried out with 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. All effects were analyzed with analysis of covariance using the numbers of tests during the follow-up period as the dependent variable and the numbers of tests at baseline and the region, which appeared to be an important determinant, as independent variables.

RESULTS

One hundred seventy-four primary care physicians, belonging to 26 local groups, expressed their willingness to participate on first request, so no further recruitment was necessary. After randomization, both arms included 13 local groups (Figure). No differences were found among the characteristics of our individual study primary care physicians (TABLE 2). Likewise, no differences were found in the characteristics of the local primary care physician groups (data not shown). The mean size of the local groups and experience with continuing medical education in small groups of colleagues did not differ between the 2 arms, nor was there any statistically significant difference between the 2 arms in the mean numbers of tests during the baseline period (data not shown). In multilevel analyses, the point estimation and SD were about the same as in the analysis

Figure. Flow of Randomized Trial



Box. Inappropriate Tests as Defined in Evidence-Based Guidelines[1-3]

Upper Abdominal Complaints

1. There is no reason to order liver function tests for vague upper abdominal complaints without jaundice. The risk of false-positive results is too large because of the low prevalence of patients with liver diseases in general practice
2. If screening is necessary, order serum glutamic-pyruvate transaminase and γ-glutamyltransferase in patients without jaundice
3. Order total bilirubin, serum glutamic-pyruvate transaminase, and γ-glutamyltransferase in patients with jaundice

General Malaise, Fatigue, and Vague Complaints

1. Order hemoglobin and erythrocyte sedimentation rate in patients with general fatigue that has persisted for longer than 1 month
2. Do not order leukocyte counts in cases of general fatigue

Degenerative Joint Complaints

Do not order radiographs of the joints since the results of these tests have no influence on the treatment

of covariance at individual physician level and therefore no correction for local groups was needed, even though the intraclass correlation coefficient for block A tests was .12 and that for block B tests was .10.

Decreases in Numbers of Tests

All the changes in the intervention group were in agreement with the national evidence-based guidelines (TABLE 3), that is, the represented reductions in the numbers of tests ordered. The number of tests ordered were always larger in the intervention arm than in the control arm. The primary care physicians in arm A decreased the total mean numbers of tests relating to problems allocated to arm A by 12% between baseline and follow-up, while no change in the numbers of these tests occurred for primary care physicians in arm B (blind control arm). The decrease for physicians in arm A was 67 tests more per physician compared with the decrease for the physicians in arm B ($P=.01$). The physicians in arm B achieved a decrease of 8% in total number of tests ordered for the problems allocated to arm B between baseline and follow-up, while a 3% decrease was achieved in the numbers of these tests by physicians in arm A (blind control arm). These results correspond with an additional decrease in the total numbers of tests for problems allocated to arm B of 28 compared with the physicians of arm A ($P=.22$).

The results per clinical problem also are shown in Table 3. The mean change in numbers of tests ordered for the 3 clinical problems allocated to arm A was statistically significant (cardiovascular, $P=.01$; upper abdominal, $P=.01$; lower abdominal, $P=.02$), while the change in the numbers of tests ordered for the 3 clinical problems allocated to arm B was in agreement with the recommendations in the national guidelines, although each failed to reach statistical significance.

Decreases in Numbers of Inappropriate Tests

The reduction in the total numbers of inappropriate tests is shown in TABLE 4. After the intervention, significantly fewer total inappropriate tests for the problems allocated to arm A were ordered by the primary care physicians in this arm ($P=.01$). The total numbers of inappropriate tests for the problems allocated to arm B ordered by the primary care physicians in arm B also tended to decrease, which was in agreement with the recommendations in the guidelines, but the reduction failed to reach statistical significance ($P=.11$). A significant reduction in the numbers of tests ordered, compared with the control group, was found for 4 of the tests for upper abdominal complaints: amylase, bilirubin, lactic dehydrogenase, and alkaline phosphatase.

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

Individual Variables	Arm A	Arm B
No. of physicians	85	89
Age, mean (SD), y	46.2 (6.6)	45.8 (5.4)
Female, No. (%)	14 (16)	15 (17)
No. of patients per physician, mean (SD)*	2587 (641)	2637 (519)
Patients >65 y, % mean (SD)	15 (6.8)	13 (7.1)
Physicians with a part-time factor, % mean (SD)†	91 (15)	91 (16)
Physicians with a solo practice, No. (%)	43 (51)	48 (54)
Physicians who use computerized registration system, No. (%)	66 (78)	61 (69)

*Total practice population for whom the primary care physician is responsible.
†Part-time factor is the working time. A full-time factor is 100%, each half of the day is 10%, so the part-time factor of 80% is a physician who works 4 days.

Table 3. Effects of the Strategy by Analysis of Covariance Adjusted for Numbers of Diagnostic Tests at Baseline and for the Region on the Mean (SD) Numbers of Tests, per Primary Care Physician per 6 Months

Clinical Problem	Baseline, Mean (SD)	Follow-up, Mean (SD)	% Change	Baseline, Mean (SD)	Follow-up, Mean (SD)	% Change	β (SE)*	95% CI	*P* Value
	Arm A Tests								
	Arm A (Intervention)			Arm B (Control)					
Total tests	478 (309)	422 (234)	−12	507 (293)	503 (281)	0	−67 (19)	−104 to −30	.01
Cardiovascular/hypertension	293 (189)	276 (157)	−6	290 (182)	302 (184)	+4	−35 (13)	−61 to −10	.01
Upper abdominal complaints	165 (125)	128 (82)	−22	192 (128)	174 (114)	−9	−28 (9)	−45 to −10	.01
Lower abdominal complaints	20 (20)	18 (19)	−10	25 (25)	27 (29)	+8	−5 (2)	−9 to −1	.02
	Arm B Tests								
	Arm A (Control)			Arm B (Intervention)					
Total tests	640 (394)	624 (357)	−3	724 (386)	664 (356)	−8	−28 (23)	−74 to 14	.22
COPD/asthma	39 (31)	31 (25)	−20	53 (27)	38 (19)	−28	−1 (2)	−5 to 3	.58
General complaints	548 (340)	544 (310)	0	599 (340)	568 (321)	−5	−19 (21)	−61 to 22	.36
Degenerative joint complaints	54 (38)	49 (36)	−9	72 (43)	58 (37)	−19	−3 (4)	−10 to 4	.34

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; SE, standard error.
*β is the intervention effect (analysis of covariance) 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 baseline and follow-up in mean (SD) numbers of tests in the intervention group minus the total change between baseline and follow-up in mean numbers of tests in the control group, adjusted for baseline and region.

Table 4. Effects of the Strategy by Analysis of Covariance Adjusted for Numbers of Diagnostic Tests at Baseline and for the Region on the Mean (SD) Numbers of Inappropriate Tests, per Primary Care Physician per 6 Months

Inappropriate Tests	Baseline, Mean (SD)	Follow-up, Mean (SD)	Baseline, Mean (SD)	Follow-up, Mean (SD)	β (SE)*	95% CI	*P* Value
			Arm A Tests				
	Arm A (Intervention)		Arm B (Control)				
Total tests	63 (75)	45 (41)	66 (55)	63 (56)	−16 (4.8)	−27 to −7	.01
BUN	8.7 (19)	7.2 (15)	6.3 (7.2)	6.6 (8.3)	−1 (1.3)	−4 to 2	.37
SGOT	7.7 (11)	5.5 (7.7)	8.3 (13)	7.5 (14)	−2 (1.4)	−5 to 1	.13
LDH	13 (27)	8.8 (16)	12 (20)	11 (18)	−3 (1.5)	−6 to −1	.01
Amylase	5.3 (13)	3.6 (6.9)	3.4 (4.9)	4.5 (10)	−2 (1.1)	−4 to −0.1	.04
Alkaline phosphatase	11 (25)	7.0 (11)	9.3 (13)	9.0 (15)	−3 (1.5)	−6 to −0.3	.03
Bilirubin	20 (27)	15 (19)	31 (43)	27 (35)	−6 (2.6)	−11 to −0.3	.04
			Arm B Tests				
	Arm A (Control)		Arm B (Intervention)				
Total tests	134 (81)	126 (74)	163 (89)	138 (74)	−8 (5.0)	−18 to 2	.11
Immunoglobulin E	3.6 (5.3)	2.8 (4.7)	3.0 (5.3)	1.5 (2.7)	−1 (0.42)	−1 to 1	.14
Leukocyte count	95 (63)	92 (57)	110 (69)	96 (58)	−6 (4.0)	−4 to 2	.11
Total imaging tests†	36 (26)	31 (22)	50 (34)	41 (26)	−1 (2.7)	−4 to 6	.70

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; BUN, blood urea nitrogen; LDH, lactic dehydrogenase; SE, standard error; SGOT, serum glutamic-oxaloacetic transaminase.
*See footnote in Table 3 for the intervention effect β.
†Total imaging tests include chest radiography, radiographs of the lumbar spine, cervical spine, shoulder, knee, and hip.

COMMENT

A new strategy to influence test ordering performance was evaluated in a trial with a large group of primary care physicians in 5 diagnostic center regions in the Netherlands. The relatively short intervention period resulted already in a substantial reduction in the total numbers of tests ordered and in the number of inappropriate tests ordered. Although the effects may seem not very large, it is important to realize primary care physicians in the Netherlands already order fewer tests than their colleagues in other countries.[1] This further reduction can be regarded as quality improvement in terms of test ordering because these changes were in agreement with the recommendations in national evidence-based guidelines.

There are some methodological considerations. We have no reason to believe that the large study population differs from the Dutch primary care physician population. Items relevant for the determinants of test ordering performance of primary care physicians were distributed equally over both arms.[17] However, maybe only motivated, well-functioning groups of physicians participated, and it is therefore questionable if the strategy will work for all groups. Secondly, our study only evaluated effects on volume of tests, because patient data were not available from the diagnostic centers. However, available empirical evidence shows that a general reduction in test ordering in primary care does not lead to more referrals or substitution of care.[18] Furthermore, despite that the guidelines state that a reduction in total test ordering equals quality improvement, this does not implicate that each separate test should always decrease. Finally, the duration of the study is too short to determine long-term effects on test ordering.

Our study underlines that multifaceted interventions are superior to single interventions.[19,20] Significant changes in numbers of tests were not found for all clinical problems included, so conclusions about the effectiveness of our strategy are not straightforward. Some clinical problems may require additional strategies, for example, electronic reminders may be necessary to achieve further improvement.[21] Nevertheless, our strategy would seem to be a powerful effective and tailor-made strategy, which fits in well with routine primary care physician practice in many western countries, is linked to the every day general practice routine, and gives primary care physicians the opportunity to discuss their test ordering performance with colleagues on the basis of actual performance data, making discussions less noncommittal. Discussing feedback reports and guidelines provides physicians the opportunity to change their performance by learning from each other and by learning to implement new strategies. Thus, social influence by peer interaction can be an important motivator for change.[14] Our strategy could also be used for in-hospital teams or other groups of collaborating physicians, as well as for other topics, such as prescription or referral behavior.

Author Affiliations: Center for Quality of Care Research, Department of General Practice (Drs Verstappen, Weijden, and Grol), Maastricht University, Maastricht, the Netherlands; Medical Integration Center Kempenland, Maxima Medical Center, Veldhoven, the Netherlands (Dr Verstappen); Department of Methodology and Statistics, Maastricht University, (Ms Sijbrandij); Center for Diagnostics and Consultation, Elkerliek Hospital, Helmond, the Netherlands (Dr Smeele); Medical Diagnostic Center of the Canisius-Wilhelmina Hospital Nijmegen (Dr Hermsen), Ottawa Health Research Institute, Center of Best Practice, Institute of Population Health, University of Ottawa, Canada (Dr Grimshaw).
Author Contributions: *Study concept and design:* Verstappen, van der Weijden, Smeele, Hermsen, Grimshaw, Grol.

Acquisition of data: Verstappen, Smeele, Hermsen.
Analysis and interpretation of data: Verstappen, van der Weijden, Sijbrandij, Hermsen, Grimshaw, Grol.
Drafting of the manuscript: Verstappen, van der Weijden, Hermsen, Grol.
Critical revision of the manuscript for important intellectual content: Verstappen, van der Weijden, Sijbrandij, Hermsen, Grol.
Statistical expertise: Verstappen, van der Weijden, Sijbrandij.
Obtained funding: Verstappen, van der Weijden, Grol.
Administrative, technical, or material support: Verstappen, Hermsen.
Study supervision: Verstappen, van der Weijden, Grol.
Funding/Support: This study was supported by the Dutch Health Care Insurance Council.

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. Ayanian JZ, Berwick DM. Do physicians have a bias toward action? a classic study revisited. *Med Decis Making*. 1991;11:154-158.
3. Kristiansen IS, Hjortdahl P. The general practitioner and laboratory utilization: why does it vary? *Fam Pract*. 1992;9:22-27.
4. Ferrier B, Woodward C, Cohen M, et al. Clinical practical guidelines: new-to-practice family physicians' attitudes. *Can Fam Physician*. 1996;42:463-468.
5. Zaat JO, van Eijk JT. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care*. 1992;30:846-854.
6. Wong ET. Improving laboratory testing: can we get physicians to focus on outcome? *Clin Chem*. 1995; 41:1241-1247.
7. Hoffrage U, Lindsey S, Hertwig R, et al. Medicine: communicating statistical information. *Science*. 2000; 290:2261-2262.
8. McDonald IG, Daly J, Jelinek VM, et al. Opening Pandora's box: the unpredictability of reassurance by a normal test result. *BMJ*. 1996;313:329-332.
9. Little P, Cantrell T, Roberts L, et al. Why do GPs perform investigations? the medical and social agendas in arranging back x-rays. *Fam Pract*. 1998;15: 264-265.
10. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet*. 1993;342:1317-1322.
11. Oxman AD, Thomson MA, Davis DA, et al. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ*. 1995;153:1423-1431.
12. Solomon DH, Hashimoto H, Daltroy L, et al. Techniques to improve physicians' use of diagnostic tests: a new conceptual framework. *JAMA*. 1998;280:2020-2027.
13. Grol R. Peer review in primary care. *Qual Assur Health Care*. 1990;2:119-126.
14. Mittman B, Tonesk X, Jacobson P. Implementing clinical practice guidelines: social influence strategies and practitioner behavior change. *Qual Review Bull*. 1992:413-422.
15. Brady WJ, Hissa DC, McConnell M, et al. Should physicians perform their own quality assurance audits? *J Gen Intern Med*. 1988;3:560-565.
16. Winkens RA, Pop P, Bugter Maessen AM, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet*. 1995;345:498-502.
17. Bugter Maessen AM, Winkens RA, Grol RP, et al. Factors predicting differences among general practitioners in test ordering behaviour and in the response to feedback on test requests. *Fam Pract*. 1996; 13:254-258.
18. Winkens RA, Grol RP, Beusmans GH, et al. Does a reduction in general practitioners' use of diagnostic tests lead to more hospital referrals? *Br J Gen Pract*. 1995;45:289-292.
19. Wensing M, Grol R. Single and combined strategies for implementing changes in primary care: a literature review. *Int J Qual Health Care*. 1994;6:115-132.
20. 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.
21. 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.

I know of but one freedom and that is the freedom of the mind.
—Antoine de Saint-Exupéry (1900-1944)