Family Practice Vol. 21, No. 4 © Oxford University Press 2004, all rights reserved. Printed in Great Britain
Doi: 10.1093/fampra/cmh408, available online at www.fampra.oupjournals.org

Variation in test ordering behaviour of GPs: professional or context-related factors?

Wim HJM Verstappen[a], Gerben ter Riet[b], Willy I Dubois[a,†], Ron Winkens[a], Richard PTM Grol[a] and Trudy van der Weijden[a]

Verstappen WHJM, ter Riet G, Dubois WI, Winkens R, Grol RPTM and van der Weijden T. Variation in test ordering behaviour of GPs: professional or context-related factors? *Family Practice* 2004; **21:** 387–395.

Objective. The aim of this study was to describe GPs' test ordering behaviour, and to establish professional and context-related determinants of GPs' inclination to order tests.

Methods. A cross-sectional analysis was carried out of 229 GPs in 40 local GP groups from five regions in The Netherlands of the combined number of 19 laboratory and eight imaging tests ordered by GPs, collected from five regional diagnostic centres. In a multivariable multilevel regression analysis, these data were linked with survey data on professional characteristics such as knowledge about and attitude towards test ordering, and with data on context-related factors such as practice type or experience with feedback on test ordering data. The main outcome measure was the percentage point differences associated with professional and context-related factors.

Results. The total median number of tests per GP per year was 998 (interquartile range 663–1500), with significant differences between the regions. The response to the survey was 97%. At the professional level, 'individual involvement in developing guidelines' (yes versus no), and at the context-related level 'group practice' (versus single-handed and two-person practices) and 'more than 1 year of experience working with a problem-oriented laboratory order form' (yes versus no) were associated with 27, 18 and 41% lower numbers of tests ordered, respectively.

Conclusion. In addition to professional determinants, context-related factors appeared to be strongly associated with the numbers of tests ordered. Further studies on GPs' test ordering behaviour should include local and regional factors.

Keywords. Family practice, health care, inter-doctor variation, physician's practice patterns, quality assurance, test ordering.

Introduction

The use of laboratory and imaging tests by GPs is increasing in many countries, and inter-doctor variation has been shown to be large.[1–3] The reasons for the increase in the numbers of tests ordered are still imperfectly understood, and probably complex. Possible explanations include the expansion of modern diagnostic technology, increased fear of litigation and lack of knowledge about appropriate test use.[4–6] Furthermore, monitoring of chronic diseases is increasingly performed by GPs, due to a shift of care from hospital to primary care.[7]

Improving the quality of test ordering requires a thorough understanding of the causal determinants of test ordering behaviour.[8–11] Previous studies into determinants of test ordering have, in general, yielded inconsistent conclusions. Various professional or practice-related factors have been held responsible for the inter-doctor variation (GP's age, years of experience as a GP, GP's attitude towards risk taking, practice size and practice type), but no single determinant has been found to be very influential across all of these studies.[12–18] The present study attempted to investigate the influence of context-related determinants not only at the practice level but also at the level of local GP groups, such as differences between GP groups in patterns of collaboration, and, at the regional level, such

Received 24 February 2003; Revised 27 August 2003; Accepted 10 March 2004.

[†]Deceased. [a]Centre for Quality of Care Research, Department of General Practice, Maastricht University, [b]Department of General Practice, Academic Medical Centre, University of Amsterdam, The Netherlands and Horten Centre, University of Zurich, Switzerland. Correspondence to Wim Verstappen, Department of General Practice, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands; E-mail: wim.verstappen@hag.unimaas.nl

as differences between regions in quality improvement programmes or ways of organizing test requests. We studied the variation in actual test ordering behaviour among a large group of GPs, to assess determinants of inter-doctor variation, at both the professional level and the level of the local and regional context.

Methods

Design and population

We performed a cross-sectional study of the numbers of tests ordered by GPs, and linked these test ordering data with data from a survey among the study population. Test data were retrieved from the files by staff members of five participating diagnostic centres. A diagnostic centre is an institute, usually associated with a hospital, where GPs can order tests without referring the patient to an out-patient clinic. One of the tasks of the medical co-ordinator of such a centre is to provide feedback to the GPs about their test ordering. The five different diagnostic centres included in the study used similar problem-oriented test ordering forms for laboratory tests, with tests categorized into groups based on clinical problems (Fig. 1). The study population consisted of GPs associated with these regional diagnostic centres and whose individual test ordering data could be retrieved. Dutch GPs collaborate with colleagues in so-called local GP groups. They share patient care outside office hours and most groups provide continuing medical education as an important activity. GPs consented to having their individual data on test ordering behaviour used for research purposes.

Variables and instruments

The dependent variable for the multivariable regression analysis was the total number of tests that the GP requested in 1 year (1997). Data of 27 tests (19 laboratory and eight imaging) were retrieved (Table 1). Data on the desktop tests that many GPs regularly perform in their own practice [erythrocyte sedimentation rate (ESR), haemoglobin, glucose and cholesterol] could not be retrieved, and these tests were therefore excluded.

The GPs in the study population were surveyed on the following professional and context-related determinants.

Professional characteristics. These included age, number of years of experience, working full time (5 days) or part time, knowledge of diagnostic accuracy measures, e.g. sensitivity, predictive value, involvement in guideline development and personal opinions on test ordering. The latter variable was measured on a 5-point scale, with options ranging from disagree to agree.

Context-related determinants. At the practice level, we determined practice type, size and location of practice, fraction of privately insured patients (compared with sick fund-insured patients),[16] the fraction of patients older than 65, level of computerization, distance to the laboratory and imaging facility, and use of desktop equipment. Use of desktop equipment was measured on a 4-point scale ranging from never to always. At the local GP group level, we measured quality improvement activities in the GP group setting (yes/no), presence of at least one member who participated (or had participated) in guideline development for the Dutch College of General Practitioners (yes/no) and presence of a joint strategy on medication and test ordering in the local GP group (yes/no). At the regional level, we assessed the experience with feedback from the regional diagnostic centre (yes/no) and whether respondents had at least 1 year's experience with the problem-oriented laboratory form (yes/no).

Analysis

Descriptive analyses were performed on test ordering data relating to the 27 tests selected, both for all 27 and for laboratory and imaging tests separately; differences in test ordering data between regions were tested with the Kruskal–Wallis test. To obtain a normal distribution of the dependent variable, all regression analyses were performed with the log-transformed total number of tests ordered. As a consequence, $e^{\text{regression coefficient}}$ reflects a relative risk, and results are reported as percentage point changes associated with the various independent variables.

As an initial step in the regressions analysis, we first conducted a stepwise backward linear regression analysis for each region separately. This approach shows which variables predict best the number of test orders for each region. In these analyses, all variables initially are entered into the model. The regression algorithm then removes—taking into account the effects of others—those variables that do not have a strong independent association with the number of test orders. Using robust variance estimation, we took into account that, even within the same region, the numbers of test orders GPs requested cannot be assumed to be statistically independent of each other, because the test ordering behaviour of two GPs within the same GP group may be more similar than that of two GPs from different GP groups. In this initial step of the regression analyses, we adjusted for working full time or part time, and the practice size, i.e. these variables were forced into the model and were never omitted. The effect of any other variables should be seen in the context of these two. In accordance with the statistical literature, the *P*-values for entry into or removal from the multivariable model were set at 0.15 and 0.20. In an effort to avoid the selection of too many variables and over-fitting of the data set, only those variables that were selected in each region by this stepwise procedure were eligible for entry into the multilevel multivariable analysis.

St. Jans Hospital Weert		GPs' Order For Laboratory Tests		Route 40
GP Code				
<0> <1> <2> <3> <4> <5> <6> <7> <8> <9>				
Copy to:				
Patient number				
<0> <1> <2> <3> <4> <5> <6> <7> <8> <9>				
<0> <1> <2> <3> <4> <5> <6> <7> <8> <9>				
Clinical data:	☐	Excluding a possible disorder		
	☐	Confirming a possible disorder		
■	☐	Monitoring purposes		
GENERAL COMPLAINTS		**COAGULATION DEFECT**		**LIVER FUNCTION TEST**
☐ ESR ☐ glucose	☐	APTT, PT, trombocytes	☐	ALAT, gammaGT
☐ Hb screening ☐ TSH screening		**CARDIAC COMPLAINTS**	☐	bilirubin (only in case of icterus)
ANAEMIA		**Angina pectoris**		**MONONUCLEOSIS INFECTIOSA**
Diagnosis	☐	Hb (suspicion anaemia)	☐	WBC, leuco diff. count
☐ Hb, MCV	☐	TSHscreening(suspicion thyriod disorder)		**NEONATAL ICTERUS**
further tests depending Hb, MCV		**control hearth failure**	☐	bilirubin
Control	☐	kreatinin, Na, K. Before start		**KIDNEY FUNCTION TEST**
☐ Hb		medication after 2 weeks, than each six months	☐	Serum kreatinin
ACUTE INFECTIOUS DISEASE				
☐ CRP				**PROSTATE FUNCTION**
To diagnose/exclude appendicitis		**CHOLESTEROL**	☐	Prostate specic antigen
☐ CRP, WBC, leuco diff count		**diagnosis**		**PSYCHOGERIATRY**
ARTHRITIS	☐	cholesterol / HDL-cholesterol	☐	**Screening.** Hb, -indices,
Diagnosis / control RA	☐	lipids (empty stomach)		ESR, TSH screening, Na, K
☐ ESR		**control with statin-use**		ALAT, gammaGT, vit B1, B6, B12,
☐ rheum factor	☐	cholesterol, ALAT		foliumacid
Control Sulfasalazine Therapy		**DIABETES MELLITUS**		**THYREOID FUNCTION TEST**
☐ Hb, WBC, leuco diff count,		**Diagnosis**	☐	TSH screening
trombocytes, gammaGT, ALAT, kreatinin,	☐	glucose	☐	Thyroid Antibodies (TPO)
albumin in urine.		**3-mnd.control diabetes type 2**		**control hypothyroid therapy**
Each month	☐	glucose	☐	TSH, Free Thyroid Hormone
diagnosis / control gout		only when suspecting a bad regulation:		when stable 3 monthly till yearly.
☐ uric acid	☐	HbA1c		**control hyperthyroid therapy**
ATOPIC SYNDROME		**Risicostratification/1-year control**	☐	Free Thyroid Hormone every 6 weeks.
☐ **Inhalation-allergic screening** (grass, dog, cat	☐	glucose (empty stomach)		.if stable, 3 monthly till yearly
☐ **Food-allergic screening** (milk, soya, egg, fish	☐	lipids (empty stomach)	☐	TSH 3 monthly till yearly
peanut)		**In morning urine**		**PREGNANCY**
If positive: follow-up test of specific antigen	☐	albumin / kreatinin ratio	☐	pregnanacy test
☐dust / mite		repeat twice if positive		**Diabetes during pregnancy**
☐ cat o milk		**HYPERTENSION**	☐	glucose (empty stomach)
☐ dog o soya	☐	glucose, cholesterol / HDL, kreatinin, K	☐	glucose tolerance test
☐ grass o chicken egg		**When starting ACE-remmer**		**OTHER TESTS**
☐ three o fish	☐	kreatinin, K	☐	
☐ funghi o peanut		**In morning urine**	☐	
☐ herbs	☐	albumin / kreatinin ratio	☐	
☐ horse		repeat twice if positive	☐	

FIGURE 1 *Problem-oriented laboratory order form*

In the final regression model, the data had a clear hierarchical structure, with GP groups operating under single regional diagnostic centres and GPs collaborating within GP groups. Again, one should not assume that test ordering behaviour of two GPs within the same GP group will be more similar than that of two GPs from different GP groups. The same holds for GP groups within a region being perhaps more similar than two GP

TABLE 1 *Tests retrieved from diagnostic centres*

Laboratory tests		Imaging tests
Packed cell volume	Alanine aminotransferase	Chest X-ray
White blood count	Aspartate aminotransferase	Double contrast barium enema
C-reactive protein	γ-Glutamyltransferase	Ultrasound of hepatobiliary tract
Thyroid-stimulating hormone	Alkaline phosphatase	X-ray of cervical spine
Potassium	Lactate dehydrogenase	X-ray of lumbar spine
Creatinine	Amylase	X-ray of hip
Blood urea nitrogen	Bilirubin	X-ray of knee
Sodium	Immunoglobulin E	X-ray of shoulder
Uric acid	Allergic screening test	
Prostate-specific antigen		

groups randomly chosen from different regions. Therefore, the data were modelled in a three-level multilevel analysis model using the Stata command *gllamm* (Generalized Linear Latent and Mixed Models), with GP group and region as the random coefficients. The variables selected by the previously described stepwise procedure for each region separately were eligible for the multilevel model. In addition, all context-related factors measured at local GP group and regional level were entered. Thus, the initial multilevel model contained 11 independent determinants (see Table 3). To adjust for practice size, the natural logarithm of practice size was entered as an offset variable.[19] Briefly, this was done because it was the number of tests ordered that was essential, rather than the order rate, i.e. the number of orders per potential patient who triggered the order by his or her visit to the GP. No tests for interactions were performed to avoid the risk of false-positive associations in subgroups before the theoretical mechanisms underlying test ordering are better understood. The likelihood ratio test was used to decide which levels would be retained. All analyses were carried out using Stata statistical software (Release 7.0. Stata Corporation, College Station, TX).

Results

Individual test ordering data were retrieved for 229 GPs, working in 40 local GP groups in the five selected regions in The Netherlands (Table 2). Figure 2 demonstrates the large variation between regions in the total number of tests ordered ($P < 0.001$). In region III, the median number of tests ordered proved to be more than twice that in region II.

Of the 229 GPs, 221 (97 %) returned the questionnaire. Compared with all Dutch GPs, the study population included more male GPs and more GPs working in urban practice locations. Two-person practices were under-represented, while relatively more GPs practised in group practices (data not shown). Table 3 presents some characteristics of the study population at GP, practice and local GP group levels. Eighteen GPs were actually involved in developing guidelines. A knowledge question, involving the application of Bayes' theorem to a patient case, was correctly answered by 16% of the study population. A total of 111 GPs (55%) answered that they would feel uncomfortable if it appeared that they clearly ordered more tests than their colleagues. In contrast, nine GPs (4.1%) would be uncomfortable if they ordered fewer tests. There was a desire to discuss personal test ordering behaviour in local GP groups, and to receive feedback on test ordering from the diagnostic centre. At the local group and regional levels, 22 local GP groups had experience of discussing their test ordering behaviour in the local GP group, which had led to (group) plans for change. At the regional level, there was only one region (region I) where the diagnostic centre was already providing individualized feedback on test ordering behaviour, while two of the five regions had introduced the problem-oriented form >1 year previously (regions I and II).

Determinants of test ordering variation

Table 3 also shows the professional and context-related variables that were eligible for entry in the multilevel model. The variable location of practice, whose omission had a negligible effect on the coefficients of the remaining variables, was omitted. The random variation due to the local GP group level proved to be small and insignificant after the three GP group level variables had been omitted. Therefore, the local GP group level was omitted, and our final multilevel model contained seven variables. Our final two-level model explained ~30% of the variation in test ordering. Two of the variables of the final multilevel model were at the professional level: working full time or part time; and participation in the production of a guideline. Three variables were at the context-related practice level: type of practice; distance to an imaging facility; and distance to a laboratory

TABLE 2 *Distribution of numbers of tests ordered by 229 GPs in five regions*

		Total	Region I	Region II	Region III	Region IV	Region V
Total numbers of tests ordered	P5	364	261	322	617	349	577
	P25	663	576	499	1085	694	1125
	P50	998	860*	666*	1742*	891*	1273*
	P75	1500	1436	847	2781	1344	1608
	P95	2648	1960	1293	3805	2413	2674
Total numbers of laboratory tests ordered	P5	303	157	250	498	332	448
	P25	565	456	400	942	569	903
	P50	839	691*	568*	1469*	799*	1078*
	P75	1271	1116	730	2498	1194	1398
	P95	2297	1732	1104	3445	2071	2249
Total numbers of imaging tests ordered	P5	40	57	34	74	35	38
	P25	99	110	61	173	96	128
	P50	146	159*	90*	243*	142*	162*
	P75	218	245	132	316	175	221
	P95	370	470	254	379	382	383

* $P < 0.001$ Kruskal–Wallis.
P represents the percentile of the distribution. For example, P25 means that 25% of all values are lower than this value. P50 is identical to the median.

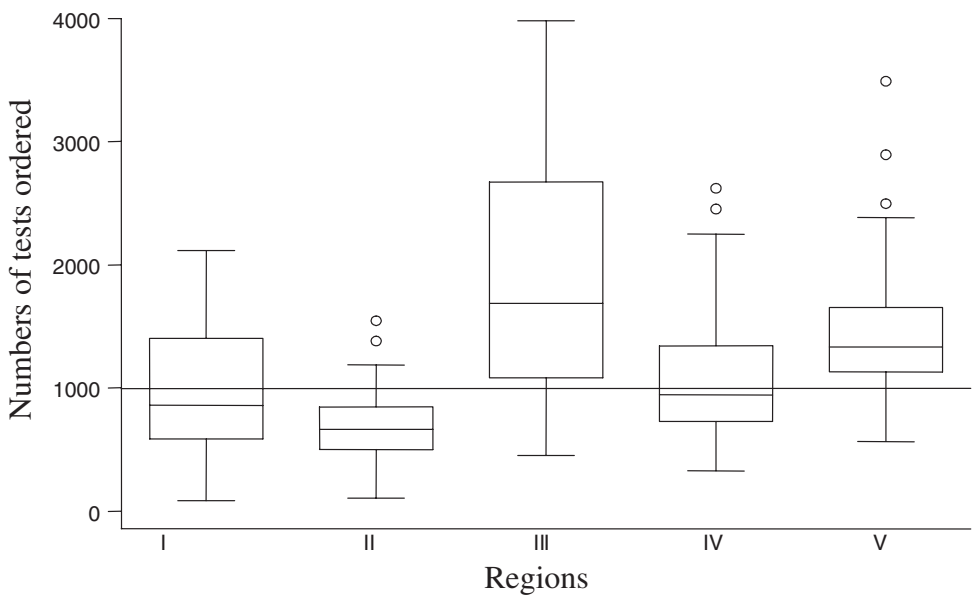

FIGURE 2 *Box plot showing the distribution of the numbers of laboratory and imaging tests ordered by 229 Dutch family physicians in each of five regions in 1997. The horizontal line shows the overall median number of tests (998) ordered. The horizontal lines within the boxes represent the medians for each respective region. The lower and upper ends of the boxes are the lower and upper quartiles. The 'antennae' sticking out from the boxes delineate where 95% of the observations lie. Dots represent the number of tests ordered by physicians who ordered very many tests compared with colleagues within their region. The graph shows the large inter-regional differences with respect to the average number of test orders as well as with respect to the variation in the numbers of tests ordered. For example, 50% of physicians in region II ordered between 499 and 847 tests, whereas these numbers are 1085 and 2781 for the physicians in region III*

facility. Two variables were at the context-related regional level: feedback on test ordering; and experience with the problem-oriented form. Table 4 shows detailed results of the final two-level model. At the professional GP level, having been actively involved in national guideline setting was associated with a 27% lower volume of tests ordered compared with non-active GPs. The practice type contributed significantly to the

TABLE 3 *Individual and context-related determinants of the number of tests ordered by 221 GPs in 1997*

Study population characteristics	Eligible for initial multilevel model	β	SE β	Total
Determinants related to GPs (n = 221)				221
Male				191
Age (SD)				46.1 (6.2)
GP's number of years of experience in years (SD)				15.5 (7.6)
	Work time factor 5 days	Reference		
	>4.5 days	−0.0756	0.1201	169
	4 days	0.2333	0.1158	25
	<4 days	−0.1031	0.0988	26
	Involved in developing guidelines	−0.2300	0.1269	18
GPs answering questions on diagnostic accuracy correctly				16
Don't want to order more tests than colleagues (scale 1–5)[a]				3.2
Desire to discuss test ordering in local groups (scale 1–5)[a]				4.1
Desire to receive feedback on test ordering (scale 1–5)[a]				4.1
Attitude to risk taking (scale 1–5)[a]				2.7
Desire to have direct access to MRI facility (scale 1–5)[a]				2.1
Context-related determinant level				
Practice				
	Practice size (SD)	Offset variable		2545 (525)
% Privately insured (SD)				35.4 (11.2)
% Older than 65 years (SD)				14.4 (6.8)
No. of GPs working in computerized practice				206
No. of GPs using medical module information system				146
	Practice location: urban	Reference		108
	Semi-urban	−0.0195	0.1022	56
	Rural	−0.0804	0.1132	56
	Practice type: single-person	Reference		103
	Two-person	−0.0989	0.0954	40
	Group practice	−0.1641	0.1052	77
% of GPs using desk top testing always for Hb, ESR and glucose				12.8
	Distance to imaging facility in km	0.0004	0.0087	6.2 (5.3)
	Distance to laboratory facility in km	0.0120	0.0130	2.3 (2.5)
Local GP group (n = 40)				
	No. of local GP groups receiving feedback on test ordering	−0.0678	0.2157	22
	No. of local GP groups making group plans for change	−0.0508	0.0994	26
	At least one GP in the GP group is involved in developing guidelines	−0.1220	0.1033	12
Region (n = 5)				
	No. of diagnostic centres providing feedback on test ordering	−0.4776	0.1251	1

The second column shows the 11 determinants eligible for the initial multilevel model analysis, including practice size (offset variable).
β = regression coefficient; MRI = magnetic resonance imaging.
[a] Personal opinions of GPs on test ordering 1 = disagree … 5 = agree.

TABLE 4 *Results of final two-level multivariable analysis*

Level	Determinant		Difference (%)	*P*	95% CI
Professional	Working full time or part time	5	(0) reference		
		4.5 days	−13.5	0.210	−31.0 to 8.5
		4 d	15.7	0.204	−7.6 to 45.0
		1.5–3.5 days	−14.3	0.105	−28.9 to 3.3
	Actively involved in developing guideline(s)	No	(0) reference		
		Yes	−26.9	0.013	−43.0 to −6.4
Context-related practice	Practice type	Single-person	(0) reference		
		Two-person	−5.9	0.516	−21.8 to 13.1
		Group	−18.0	0.022	−30.9 to −2.8
	Distance to imaging facility (per 10 km)		−9.4	0.168	−21.9 to 3.2
	Distance to laboratory facility (per 10 km)		19.1	0.142	−7.4 to 43.5
Context-related regional	Diagnostic centre providing feedback	No	(0) reference		
		Yes	24.1	0.311	−18.2 to 88.3
	Problem-oriented form >1 year	No	(0) reference		
		Yes	−41.0	0.001	−57.2 to −18.7

Differences are percentage point changes compared with a reference category $r^2 = 0.304$.

variation in test ordering: GPs working in group practices ordered ~18% fewer tests than those in single-person or two-person practices. At the context-related regional level, having had at least 1 year's experience with the problem-oriented laboratory form was associated with a 41% lower volume of tests ordered. The intra-class correlation coefficient at region level was 0.304, meaning that the variation between regions was large compared with the variation within regions, which supports the assumption that variability in test ordering is strongly correlated with a region factor.

Discussion

To our knowledge, the present study is the first explicitly to include context-related variables at GP group and regional level. This enabled us to focus on the variation in GPs' test ordering behaviour in relation to both professional and context-related determinants. We found, to our surprise, a large variation in test ordering between the regions, and we determined three variables that were independently and strongly associated with the volume of tests, namely involvement in developing guidelines, working in a group practice, and having had more than 1 year's experience with a problem-oriented form.

At the level of the professional, GPs who were involved in developing national clinical guidelines (in the context of the Dutch College of GPs programme for guideline setting) ordered fewer tests than other GPs. Although this subgroup represents a minor and probably selected proportion of the GPs, discussing guidelines and the underlying medical evidence might be an important part of a strategy to improve test ordering behaviour.[20–23] Secondly, at the context-related practice level, working in a group practice was also associated with a considerably lower number of tests ordered. This finding, which probably results from general discussions of and reflections on practice behaviour in such group practices, is in line with earlier findings related to prescription behaviour.[24–26] Finally, at the regional level, it was particularly the level of experience with a problem-oriented test ordering form that appeared to have a large impact on the numbers of tests ordered. It is not so much the influence of the order form itself that is surprising, but rather the magnitude of this effect.[27,28] The present study was unable to explain all of the inter-regional variation. Of course, disease-related factors are also important in the variation of test ordering. Although there might be slight differences in morbidity between the regions, it is unlikely that differences in case mix play an important role, because a total of ~550 000 patients were involved. Explaining this inter-regional variation will require more research, which should include patient-related, organizational and socio-cultural determinants.

Our study population differed from the total population of GPs in The Netherlands in some features, but we do not think that these differences influenced

our results. Further, in The Netherlands, diagnostic facilities only perform tests when a physician orders them. Sometimes, however, diagnostic centres perform test cascades, depending on the results of the previous test. Further, only data from the diagnostic facility were available, so the tests that were ordered but not performed, e.g. because the patient did not visit the diagnostic centre, were not included. However, both situations probably constitute a small part of the ordered tests.

Based on the present results, it is tempting to recommend the introduction of problem-oriented forms in diagnostic facilities for GPs; however, further study to replicate our findings is necessary. The problem-oriented form was developed as a quality improvement instrument, aimed at efficient and cost-efficient use of tests. Of course, it is also important to study patient-related factors, such as whether patients are actively demanding tests and how to 'sell' such a cost-conscious approach to such demanding patients. These patient factors should be discussed with colleagues, as some of them may have developed effective strategies for dealing with them. Despite the small influence of the local GP group in our study, many GPs mentioned the social influence of colleagues as an important determinant of test ordering. The medical co-ordinators of the diagnostic centres, who provide the feedback on test ordering and may as such be regarded as experts on this topic, could function as opinion leaders in these discussions.[29–31] Based on the strong correlations we found between several factors and test ordering patterns, we conclude that a quality improvement programme, consisting of discussions on guidelines and feedback reports in a local GP group, and collaborating with a diagnostic centre that uses problem-oriented test ordering forms and provides the feedback, appears to be a promising intervention to decrease overuse of GPs' test ordering.

Acknowledgements

The researchers gratefully acknowledge financial support for the present study from the College for Health Insurances.

References

[1] Ayanian JZ, Berwick DM. Do physicians have a bias toward action? A classic study revisited. *Med Decision Making* 1991; **11:** 154–158.

[2] Zaat JO, van Eijk JT. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care* 1992; **30:** 846–854.

[3] 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.

[4] Ferrier BM, Woodward CA, Cohen M, Goldsmith CH. Laboratory tests: which physicians order more? *Can Fam Physician* 1991; **37:** 347–352.

[5] Wong ET. Improving laboratory testing: can we get physicians to focus on outcome? *Clin Chem* 1995;**41:** 1241–1247.

[6] Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. *Science* 2000; **290:** 2261–2262.

[7] Kaag ME, Wijkel 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.

[8] Kristiansen IS, Hjortdahl P. The general practitioner and laboratory utilization: why does it vary? *Fam Pract* 1992; **9:** 22–27.

[9] Little P, Cantrell T, Roberts L, Chapman J, Langridge J, Pickering R. Why do GPs perform investigations: the medical and social agendas in arranging back X-rays. *Fam Pract* 1998; **15:** 264–265.

[10] Miller WL, McDaniel RR, Jr, Crabtree BF, Stange KC. Practice jazz: understanding variation in family practices using complexity science. *J Fam Pract* 2001; **50:** 872–878.

[11] Mabeck CE, Kragstrup J. Is variation a quality in general practice? *Scand J Prim Health Care Suppl* 1993; **1:** 32–35.

[12] Goold SD, Hofer T, Zimmerman M, Hayward RA. Measuring physician attitudes toward cost, uncertainty, malpractice, and utilization review. *J Gen Intern Med* 1994; **9:** 544–549.

[13] Bugter Maessen AM, Winkens RA *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.

[14] Malcolm L, Wright L, Seers M, Davies L, Guthrie J. Laboratory expenditure in Pegasus Medical Group: a comparison of high and low users of laboratory tests with academics. *NZ Med J* 2000; **113:** 79–81.

[15] Cherkin DC, Deyo RA, Wheeler K, Ciol MA. Physician variation in diagnostic testing for low back pain. Who you see is what you get. *Arthritis Rheum* 1994; **37:** 15–22.

[16] van Merode GG, Stroink AE, Maarse JA, Goldschmidt HM. Impact of insurance coverage type on laboratory test ordering behaviour of general practitioners. *World Hosp Health Serv* 2000; **36:** 7–12.

[17] Sherwood P, Lyburn I, Brown S, Ryder S. How are abnormal results for liver function tests dealt with in primary care? Audit of yield and impact. *Br Med J* 2001; **322:** 276–278.

[18] Winkens R, Dinant GJ. Evidence base of clinical diagnosis: rational, cost effective use of investigations in clinical practice. *Br Med J* 2002; **324:** 783.

[19] Armitage PBG, Matthews JNS, eds. *Statistical Methods in Medical Research*, 4th edn. Oxford: Blackwell; 2002.

[20] Mittman B, Tonesk X, Jacobson P. Implementing clinical practice guidelines: social influence strategies and practitioner behavior change. *Qual Rev Bull* 1992: 411–422.

[21] Gorton TA, Cranford CO, Golden WE, Walls RC, Pawelak JE. Primary care physicians' response to dissemination of practice guidelines. *Arch Fam Med* 1995; **4:** 135–142.

[22] Solberg LI, Brekke ML, Fazio CJ *et al*. Lessons from experienced guideline implementers: attend to many factors and use multiple strategies. *Joint Committee J Qual Improv* 2000; **26:** 171–188.

[23] Grimshaw JM, Shirran L, Thomas R *et al*. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care* 2001; **39:** Ii2–45.

[24] Wyatt TD, Reilly PM, Morrow NC, Passmore CM. Short-lived effects of a formulary on anti-infective prescribing—the need for continuing peer review? *Fam Pract* 1992; **9:** 461–465.

[25] von Ferber L, Bausch J, Koster I, Schubert I, Ihle P. Pharmacotherapeutic circles. Results of an 18-month peer-review prescribing-improvement programme for general practitioners. *Pharmacoeconomics* 1999; **16:** 273–283.

[26] Gill PS, Makela M, Vermeulen KM *et al*. Changing doctor prescribing behaviour. *Pharm World Sci* 1999; **21:** 158–167.

[27] Zaat JOM, Van Eijk JT, Bonte HA. Laboratory test form design influences test ordering by general practitioners in the Netherlands. *Med Care* 1992; **30:** 189–198.

[28] van Walraven C, Goel V, Chan B. Effect of population-based interventions on laboratory utilization: a time-series analysis. *J Am Med Assoc* 1998; **280:** 2028–2033.

[29] 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. *J Am Med Assoc* 1991; **265:** 2202–2207.

[30] Soumerai SB, McLaughlin TJ, Gurwitz JH *et al*. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *J Am Med Assoc* 1998; **279:** 1358–1363.

[31] Borbas C, Morris N, McLaughlin B *et al*. The role of clinical opinion leaders in guideline implementation and quality improvement. *Chest* 2000; **118:** 24–32.