

Interpretation of diagnostic test results:

Pretest expectations, test interpretation and management decisions in routine general practice

The studies presented in this dissertation were conducted under the auspices of the school for Public Health and Primary Care (CAPHRI) at Maastricht University, The Netherlands. CAPHRI is part of the Netherlands School of Primary Care Research (CaRe), which has been acknowledged since 1995 by The Royal Netherlands Academy of Art and Sciences (KNAW).

The printing of this thesis was generously sponsored by SBOH, employer of GP trainees.



© 2010 Paul Houben, Maastricht

Cover: Inge van Vuuren: 'twee sporen beleid'

Layout: Tiny Wouters

Production: Datawyse | Universitaire Pers Maastricht

ISBN: 978 90 5278 963 7

Interpretation of diagnostic test results:

Pretest expectations, test interpretation and management decisions in routine general practice

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. mr. G.P.M.F. Mols,
volgens besluit van het College van Decanen in het openbaar te verdedigen
op vrijdag 17 september 2010 om 14.00 uur

door

Paul Houben



Promotores:

Prof. dr. T. van der Weijden
Prof. mr. dr. R.P.T.M. Grol

Copromotor:

Dr. R. Winkens

Beoordelingscommissie:

Prof. dr. G.J. Dinant, voorzitter
Prof. dr. P.J.E. Bindels, Erasmus Medisch Centrum Rotterdam
Prof. dr. M.P. van Dieijen-Visser
Prof. dr. H.E. van der Horst, VU Medisch Centrum Amsterdam
Prof. dr. P. Pop

Contents

Chapter 1	Introduction	7
Chapter 2	Towards better use of test results: current problems and a research agenda	19
Chapter 3	GPS' considerations when interpreting laboratory test results: a qualitative study <i>Huisarts Wet. 2005;48:326-32</i>	31
Chapter 4	Reasons for ordering laboratory tests and relationship with frequency of abnormal results <i>Scand J Prim Health. 2010;28:18-23</i>	47
Chapter 5	Pretest expectations strongly influence interpretation of abnormal laboratory results and further management <i>BMC Fam Pract. 2010; 11:13</i>	59
Chapter 6	Reasons for ordering spinal X-ray investigations: how they influence general practitioners' management <i>Can Fam Physician. 2006;52:1266-7</i>	71
Chapter 7	Cascade effects of laboratory testing are found to be rare in low disease probability situations: prospective cohort study <i>J Clin Epidemiol. 2010;63:452-8</i>	81
Chapter 8	Discussion	95
	Summary	109
	Samenvatting	117
	Appendix	125
	Publications	131
	Dankwoord	135
	Curriculum Vitae	141

Chapter 1

Introduction

Introduction

In routine practice, general practitioners (GP) use laboratory and imaging tests on a daily basis¹⁻³. All the results of these tests need to be interpreted and applied for diagnosis and further management which is a complex process. This process is hardly studied in routine practice and many questions need to be answered. Deeper insight helps to understand the interpretive process and to explore the room for improvement.

This introduction first discusses knowledge about test ordering and interventions to improve test ordering (Figure 1.1, phase 1). This knowledge led to the realization that a better understanding of the use of test results is needed. Therefore, the use of test results and in particular the problems in the interpretation of results are then discussed (Figure 1.1, phase 2). The end of this introduction presents this thesis' research questions.

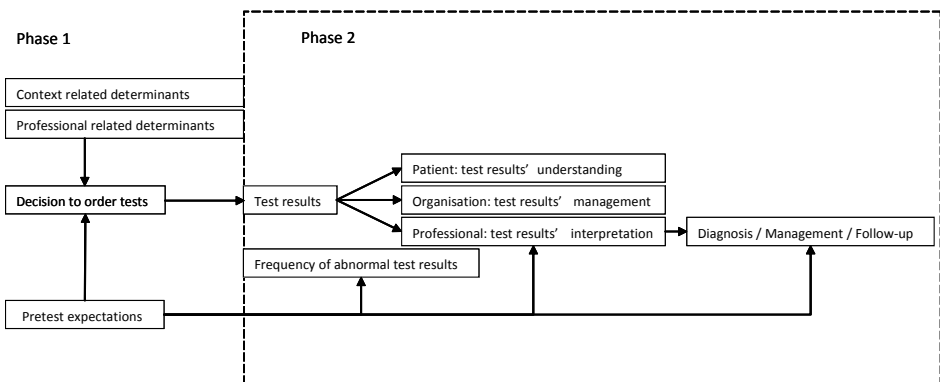


Figure 1.1 Test ordering and the use of test results.

Test ordering behaviour

GPs order laboratory and imaging tests in about 5-10% of the consultations^{1,3}. However, the number of tests that are ordered may differ significantly between GPs. This difference is influenced by both professional and context related determinants (Figure 1.1). Examples of professional related determinants are GPs tolerance towards diagnostic uncertainty, overrating the diagnostic value of tests, defensive considerations and GPs' use of diagnostic tests for reassuring worried patients⁴⁻¹¹. Examples of context related determinants are the socioeconomic status of the patients in the practice, the use of problem oriented forms for ordering tests, working in a group practice and the number of patient contacts per week^{3,12-16}.

Both the volume of tests ordered and variation between GPs in number of tests ordered suggest that tests may be ordered that are not necessary from a strict medical viewpoint. Indeed, many test orders seem unnecessary according to established, evidence based guidelines. It has been tried to reduce such unnecessary tests by interventions targeting the ordering of tests¹⁷⁻¹⁹. These interventions help to reduce the number of tests ordered and they are usually even more successful if multiple behaviourally factors are targeted^{20,21}. Examples are individual feedback on tests ordered, discussing ordering of tests in a local GP group and a problem oriented form for ordering tests²²⁻²⁶. However, even combined interventions may only moderately reduce unnecessary tests ordering. In addition, unnecessary test ordering is only one side of the medal, as it may also be that necessary tests are unjustly not ordered.

Besides the findings of limited effects of test ordering reducing strategies, the aim to reduce unnecessary testing has also been criticized for several reasons. For example, recommendations in guidelines may be incomplete or are sometimes even conflicting²⁷. Also, some tests are frequently used, such as the ESR, although there may be little evidence about the tests' usefulness for diagnosing or monitoring of disease in general practice²⁸⁻³¹. Furthermore, there may be a considerable interindividual variation in the assessment if ordered tests are medically necessary. In a study the interrater agreement about appropriateness of diagnostic test orders was assessed. The interrater kappa values were low and ranged from 0.33 to 0.42³². Finally, in many studies about ordering tests a definition of medically unnecessary is not explicitly formulated and definitions also often differ between studies³³. It may thus be difficult and ambiguous to agree upon what a medically unnecessary test order is.

Reducing unnecessary testing is therefore difficult and some unnecessary test ordering may persist. As a consequence the results of these tests still need to be interpreted and used (Figure 1.1, phase 2). However, this has hardly been studied before in day-to-day care. Better understanding may help to explore the room for improvement in this phase of diagnostic test use, and may improve the understanding of the consequences of ordering medically unnecessary tests. In this second phase of diagnostic test use problems may arise on organisational, professional and patient level. Potential problems on these different levels are further explored in chapter 2 of this thesis. The further chapters in this thesis focus in particular on the interpretation of test results at the professional level.

Interpretation of test results

The incidence of serious pathology for patients presenting in general practice is generally low. Although patients may present with a whole range of different complaints, many complaints are innocent and self limiting within a few weeks^{34,35}. Also, much pathology presents in an early stage of disease when symptoms are often

still vague and specific signs cannot yet be found. A definite diagnosis or clear pathophysiological or anatomic substrate may lack³⁶⁻³⁹. It is indeed a specific characteristic of general practice that the pretest probability for finding a disease is low in a significant part of the situations for which diagnostic tests are ordered.

Nevertheless, many test results in general practice are abnormal. However, not every abnormal test result is clinically significant⁴⁰. How often an abnormal test result without clinical significance is found, is not known. But the proportion of abnormal tests that have no clinical significance may increase if the pretest probability is low⁴¹. An explanation is the methods that are used to define the reference values for laboratory test results. For most tests the reference values are defined as two standards deviations of the mean value (95% interval). Thus, by definition, 5% of the normal population will fall outside the reference range. The probability for finding an abnormal test increases if more tests are ordered. This is illustrated in a study where healthy individuals were screened with 25 laboratory tests⁴⁰. Of the individuals, 89% had at least one abnormal test result. Test results may also be misleading because most tests have no 100% sensitivity and specificity which means that a test result may be abnormal without the patient having the disease, or a test results may be normal while the patient has the disease. Finally, there is usually biological and analytical variation in an individual between repeated measurements and there may also be a small risk of an analytical error⁴²⁻⁴⁵.

Thus, it is difficult for a GP to correctly interpret abnormal test results. However, it is striking how little evidence there is about interpretation of test results in daily practice. It is supposed, and supported by case reports, that doctors may unjustly react on abnormal results. Case reports describe how an initial abnormality may induce a process, also called cascade process, where ever more investigations are ordered to explain an initial abnormality⁴⁵⁻⁴⁷. The possible occurrences of such cascades are used as an argument to reduce the number of unnecessary orders for diagnostic investigations²⁴. However, there is no empirical evidence about the occurrence of such cascades in general practice, and there is little evidence about the circumstances which may cause erroneous interpretations of laboratory results. Anyway, overreacting is not the only problem as abnormal results sometimes may also be unjustly not further investigated⁴⁸. Therefore, to better understand problems that may occur, it is necessary to increase the understanding of GPs' interpretation of test results.

The interpretation of results may be discussed from the background of medical decision making theory^{49,50}. It describes how Bayes' theorem (see box 1.1) can be applied to calculate the posttest probability for disease from the pretest probability, the test results and the tests' sensitivity and specificity⁵¹. Alternatively, the likelihood ratio may be used⁵². However, these calculations are difficult and may be too complex for GPs' daily practice⁵³⁻⁶⁰. Several aspects of daily practice make application of these calculations even more difficult: GPs usually consider several diagnoses at one time, several tests are ordered, tests are mutually dependent, and sensitivity and specificity

are often not specifically known for the situations where the tests are used⁶¹. Indeed, there were just few doctors in daily practice (3% of the doctors in a telephone survey) that used formal Bayesian calculations. An important reason mentioned was impracticality of the Bayesian method. Of the doctors that said that they used sensitivity and specificity indexes, a major part (95%) did not do so in the recommended formal manner⁶².

The Bayesian model may therefore have shortcomings in describing how GPs interpret results in daily practice. Maybe that instead of using a formal appraisal of test results, GPs are more likely to interpret test results in the light of general clinical knowledge and pretest expectations such as the presumed diagnosis, the probability for disease and maybe also the reasons for test ordering. However, there is a lack of empirical evidence and there is insufficient knowledge about the interpretation of test results, the role of pretest expectations and further diagnostic and management decisions. Deeper insight helps to better understand the interpretive process and helps to explore the room for improvement. Therefore the following research questions for this thesis were formulated:

1. Which problems in the use of test results are recognized on organizational, professional and patient level? Which research questions need to be addressed in research about the use of test results?
2. Which considerations influence GPs' interpretation of laboratory tests and how do these considerations interact?
3. What is the frequency of mildly and markedly abnormal results in routine practice, and how do they relate to the reasons for ordering the laboratory tests?
4. What is the influence of physicians' pretest expectations, in terms of estimated pretest probability and reasons for ordering tests, on the interpretation of the test results, posttest probability estimates and further management?
5. Are GPs less likely to pursue active management (referral to specialists, prescribing medication, physical therapy, or additional investigations) if tests are ordered to reassure patients rather than to exclude or confirm disease?
6. How often do cascade processes, in situations of low pretest disease probability, occur? Which pretest determinants (reason for ordering tests, diagnosis) and posttest determinants (interpretation of tests, estimates of disease probability, diagnosis) increase the risk for such cascade processes?.

Box 1.1

BAYES' THEOREM

Bayes' theorem is a central principle of medical practice because it helps to interpret a test result for an individual patient.

Hypothetical example for the Prostate Specific Antigen test (PSA test):

1% of men have prostate cancer (and therefore 99% do not).

80% of PSA tests detect prostate cancer when it is there (and therefore 20% of PSA tests miss it: a false negative result).

90.4% of PSA tests correctly detect the absence of prostate cancer when it's **not** there (and therefore 9.6% incorrectly detect cancer: a false positive result).

Put in a table, the probabilities look like this:

	Cancer (1%)	No Cancer (99%)
Test positive	80%	9.6%
Test negative	20%	90.4%

If 10.000 patients are tested, the table looks like this:

	Cancer (1%)	No Cancer (99%)
Test positive	80	950
Test negative	20	8950
Patients	100	9.900

The chance that a patient has cancer when there is a positive PSA test is $80/(950+80)=7,8\%$.

The chance that a patient has cancer when there is a negative PSA test is $20/(8950+20)=0,2\%$

Outline of the thesis

Chapter 2 in this thesis summarizes existing knowledge about the problems that are encountered in the use of test results on organizational, professional and patient level. It analyzes gaps in knowledge and it aims to give direction to future research by providing a research agenda.

Chapter 3 presents the results of a qualitative interview study. It investigates GPs' considerations when they interpret laboratory results. The GPs show by reasoning how they interpret abnormal results, the considerations they find important, and how they place the test results in the patient's clinical picture.

The data reported on in chapters 4, 5 and 7 in this thesis were collected in the IDO-study (Interpretatie Diagnostische Onderzoeken). The IDO-study was a prospective study conducted in 2004-2005 amongst 87 GPs in 7 regions in the South of the Netherlands: Heerlen, Sittard, Weert, Geldrop, Eindhoven, Helmond and 's-Hertogenbosch. The participating GPs recorded data on 25 adult patients both when they ordered laboratory tests and when they received the test results. Data for the 6 months follow-up were extracted from the electronic medical records.

Chapter 4 investigates the relation between GPs' pretest expectations and the laboratory results. It shows how often abnormal results are found if tests are ordered for reasons such as patient reassurance. It also formulates 'action limits' for test results to distinguish mildly and markedly abnormalities.

Chapter 5 subsequently explores the relation between GPs' pretest expectations, in terms of their pretest estimate of disease probability and their reasons for ordering tests, and their interpretation of laboratory results, the diagnosis, and the management. It shows how strongly GPs' pretest expectations influence the interpretation of laboratory test results.

Chapter 6 also focuses on the relation between GPs' pretest expectations and their management. But in this chapter the relation is examined in a different set of data compared to chapter 5. In this chapter is investigated if the reasons for ordering spinal x-ray films influence management decisions.

Chapter 7 focuses on situations with a low pretest probability for disease and the occurrence of cascade processes if abnormal test results were found. The number of further diagnostic investigations and referrals to specialists during 6 months follow-up are registered. Also, the cascades that occur are qualitatively described.

Chapter 8 is the general discussion of this thesis and discusses the results of the studies in this thesis and their methodology. It critically reviews the results against existing literature and it places the results in the broader perspective of theories about interpretation of results. It discusses the implications for daily practice and further research.

References

1. Kluijnt I, Zaat J, Van der Velden J, Van Eijk J, Schellevis G. Voor een prikje? Het gebruik van klinisch-chemische, hematologische en serologische bepalingen door huisartsen. *Huisarts Wet* 1991;34:67-71.
2. Speets AM, Kalmijn S, Hoes AW, van der Graaf Y, Smeets HM, Mali WP. Frequency of chest radiography and abdominal ultrasound in the Netherlands: 1999-2003. *Eur J Epidemiol*. 2005;20:1031-6.
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;45:21-5.
4. Van Boven K, Dijksterhuis P, Lamberts H. Defensive testing in Dutch family practice. Is the grass greener on the other side of the ocean? *J Fam Pract* 1997;44:468-72.
5. Hartley RM, Epstein AM, Harris CM, McNeil BJ. Differences in ambulatory test ordering in England and America. Role of doctors' beliefs and attitudes. *Am J Med* 1987;82:513-7.
6. Guthrie B. Why do general practitioners take blood? A cross-sectional study of the use of blood tests in UK general practice. *Eur J Gen Pract* 2001;7:138-42, 60.
7. van der Weijden T, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002;52:974-80.
8. Holtgrave DR, Lawler F, Spann SJ. Physicians' risk attitudes, laboratory usage, and referral decisions: the case of an academic family practice center. *Med Decis Making* 1991;11:125-30.
9. Shye D, Freeborn DK, Romeo J, Eraker S. Understanding physicians' imaging test use in low back pain care: the role of focus groups [see comments]. *Int J Qual Health Care* 1998;10:83-91.
10. Dijksterhuis PH, Van Boven C, Lamberts H. Waarom aanvullend onderzoek? De functie van het aanvullend onderzoek voor de huisarts. *Huisarts Wet* 1994;37:467-72.
11. Zaat JO, van Eijk JT. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care* 1992;30:846-54.
12. Verstappen WH, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract* 2004;21:387-95.
13. Verstappen W, ter Riet G, van der Weijden T, Hermsen J, Grol R. Variation in requests for imaging investigations by general practitioners: a multilevel analysis. *J Health Serv Res Policy* 2005;10:25-30.
14. Hartley RM, Charlton JR, Harris CM, Jarman B. Influence of patient characteristics on test ordering in general practice. *Br Med J Clin Res Ed* 1984;289:735-8.
15. Epstein AM, Begg CB, McNeil BJ. The effects of group size on test ordering for hypertensive patients. *N Engl J Med* 1983;309:464-8.
16. Kristiansen IS, Hjortdahl P. The general practitioner and laboratory utilization: why does it vary? *Fam Pract* 1992;9:22-7.
17. Oakeshott P, Kerry SM, Williams JE. Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br J Gen Pract* 1994;44:197-200.
18. Smith BJ, McNeely MD. The influence of an expert system for test ordering and interpretation on laboratory investigations. *Clin Chem* 1999;45:1168-75.
19. Tierney WM, McDonald CJ, Hui SL, Martin DK. Computer predictions of abnormal test results. Effects on outpatient testing. *JAMA* 1988;259:1194-8.
20. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to improve physicians' use of diagnostic tests: a new conceptual framework. *JAMA* 1998;280:2020-7.
21. van Walraven C, Goel V, Chan B. Effect of population-based interventions on laboratory utilization: a time-series analysis. *JAMA* 1998;280:2028-33.
22. Winkens RA, Pop P, Bugter Maessen AM, Grol RP, Kester AD, Beusmans GH, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet* 1995;345:498-502.
23. Winkens RA, Pop P, Grol RP, Bugter Maessen AM, Kester AD, Beusmans GH, et al. Effects of routine individual feedback over nine years on general practitioners' requests for tests. *BMJ* 1996;312:490.

24. Verstappen WH, van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, Grimshaw J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians. A randomized trial. *JAMA* 2003;289:2407-12.
25. Zaat JO, van Eijk JT, Bonte HA. Laboratory test form design influences test ordering by general practitioners in The Netherlands. *Med Care* 1992;30:189-98.
26. Durand Zaleski I, Rymer JC, Roudot Thoraval F, Revuz J, Rosa J. Reducing unnecessary laboratory use with new test request form: example of tumour markers. *Lancet* 1993;342:150-3.
27. van Wijk MA, Bohnen AM, van der Lei J. Analysis of the practice guidelines of the Dutch College of General Practitioners with respect to the use of blood tests. *J Am Med Inform Assoc* 1999;6:322-31.
28. Brigden ML. Clinical utility of the erythrocyte sedimentation rate. *Am Fam Physician* 1999;60:1443-50.
29. Gronlie M, Hjortdahl P. The erythrocyte sedimentation rate; its use and usefulness in primary health care. *Scand J Prim Health Care* 1991;9:97-102.
30. Dinant GJ, Knottnerus JA, Van Wersch JW. Discriminating ability of the erythrocyte sedimentation rate: a prospective study in general practice. *Br J Gen Pract* 1991;41:365-70.
31. Dinant GJ, Knottnerus JA, van Wersch JW. Diagnostic impact of the erythrocyte sedimentation rate in general practice: a before-after analysis. *Fam Pract* 1992;9:28-31.
32. Bindels R, Hasman A, van Wersch JW, Pop P, Winkens RA. The reliability of assessing the appropriateness of requested diagnostic tests. *Med Decis Making* 2003;23:31-7.
33. van Walraven C, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits [see comments. *JAMA* 1998;280:550-8.
34. Kroenke K, Jackson JL. Outcome in general medical patients presenting with common symptoms: a prospective study with a 2-week and a 3-month follow-up. *Fam Pract* 1998;15:398-403.
35. Kenter EG, Okkes IM, Oskam SK, Lamberts H. Tiredness in Dutch family practice. Data on patients complaining of and/or diagnosed with "tiredness". *Fam Pract* 2003;20:434-40.
36. Muris JW, Starmans R, Fijten GH, Knottnerus JA. One-year prognosis of abdominal complaints in general practice: a prospective study of patients in whom no organic cause is found. *Br J Gen Pract* 1996;46:715-9.
37. Burton C. Beyond somatisation: a review of the understanding and treatment of medically unexplained physical symptoms (MUPS). *Br J Gen Pract* 2003;53:231-9.
38. van der Weijden T, van Velsen M, Dinant GJ, van Hasselt CM, Grol R. Unexplained complaints in general practice: prevalence, patients' expectations, and professionals' test-ordering behavior. *Med Decis Making* 2003;23:226-31.
39. Peveler R, Kilkenny L, Kinmonth AL. Medically unexplained physical symptoms in primary care: a comparison of self-report screening questionnaires and clinical opinion. *J Psychosom Res* 1997;42:245-52.
40. Mold JW, Aspy CB, Lawler FH. Outcomes of an insurance company-sponsored multichannel chemistry screening initiative. *J Fam Pract* 1998;47:110-7.
41. Cebul RD, Beck JR. Biochemical profiles. Applications in ambulatory screening and preadmission testing of adults. *Ann Intern Med* 1987;106:403-13.
42. Franzini C. Relevance of analytical and biological variations to quality and interpretation of test results: examples of application to haematology. *Ann Ist Super Sanita* 1995;31:9-13.
43. Irijala KM, Gronroos PE. Preanalytical and analytical factors affecting laboratory results. *Ann Med* 1998;30:267-72.
44. Plebani M, Carraro P. Mistakes in a stat laboratory: types and frequency. *Clin Chem* 1997;43:1348-51.
45. Brigden ML, Heathcote JC. Problems in interpreting laboratory tests. What do unexpected results mean? *Postgrad Med* 2000;107:145-6, 51-2, 55-8 passim.
46. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician* 1996;25:971-3.
47. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health* 2002;23:23-44.
48. 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. *BMJ* 2001;322:276-8.
49. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ* 2002;324:729-32.
50. Kassirer JP. Diagnostic reasoning. *Ann Intern Med* 1989;110:893-900.

51. Knottnerus JA. Diagnostic prediction rules: principles, requirements and pitfalls. *Prim Care* 1995; 22:341-63.
52. Sonis J. How to use and interpret interval likelihood ratios. *Fam Med* 1999;31:432-7.
53. Richardson WS. Five uneasy pieces about pre-test probability. *J Gen Intern Med* 2002;17:882-3.
54. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making* 2001;21:37-44.
55. Puhan MA, Steurer J, Bachmann LM, ter Riet G. A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Ann Intern Med* 2005;143:184-9.
56. Lyman GH, Balducci L. The effect of changing disease risk on clinical reasoning. *J Gen Intern Med*. 1994;9:488-95.
57. Dolan JG, Bordley DR, Mushlin AI. An evaluation of clinicians' subjective prior probability estimates. *Med Decis Making* 1986;6:216-23.
58. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med* 1998;73:538-40.
59. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002;324:824-6.
60. Heller RF, Sandars JE, Patterson L, McElduff P. GPs' and physicians' interpretation of risks, benefits and diagnostic test results. *Fam Pract* 2004;21:155-9.
61. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 1997;8:12-7.
62. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104:374-80.

Chapter 2

Towards better use of test results: current problems and a research agenda

Summary points

- Optimal use of test results should be an important skill for doctors in routine care.
- Practices should implement systems for management of test results including (a) tracking results, (b) informing patients, (c) documenting that the patient has been informed, and (d) ensuring follow-up.
- Benchmarks for good management of test results are needed and practices should be regularly evaluated on the quality of their test results management systems.
- Teaching doctors Bayesian reasoning might not be a good method to improve the interpretation of test results in routine care.
- Research is needed to examine how doctors interpret test results in routine care, to identify potential pitfalls and problems and to explore the room for improvement.
- Patients need clear information to understand their results, but the literature provides little evidence about the best way to communicate results.
- There is conflicting evidence on the value of testing to reassure patients.

Introduction

Optimal utilisation of health care resources such as diagnostic and screening tests is an important skill for all doctors. Crucial tasks in this respect include the management, interpretation and communication of test results, which are common, routine tasks for most physicians both in general practice and in hospital settings. However, these tasks encounter specific problems in day-to-day care. For example, whereas test results management involves providing patients with timely information about abnormal results and ensuring that adequate follow-up is initiated, this is often not achieved^{1,2}. Because dealing with test results is a routine task, even minor improvements may have major effects on health care. Therefore, the present analysis explores ways to improve test results management, interpretation and communication. We searched for papers in Medline using the key words ‘test(s)’, ‘result(s)’, ‘interpretation’, ‘management’ and ‘communication’. For each relevant paper found, we screened the reference lists for further relevant studies and we also forward-tracked relevant studies with the Science Citation Index (snowballing). Because relevant literature about this topic is relatively scarce, we also propose an agenda for further research (Table 2.1).

Table 2.1 Research questions for better use of test results in routine care.

Problem	Research questions
Test results management	<ul style="list-style-type: none"> - What are appropriate methods to measure the quality of results management systems? - What errors occur due to malfunctioning systems, how often do these errors occur and what is the nature of these errors? - What are effective interventions to improve test result management and do these prevent medical errors?
Test results interpretation	<ul style="list-style-type: none"> - What determinants and/or principles underlie the interpretation of results in routine care? To what extent does Bayesian reasoning play a role? - What are the causes of mistakes in interpretation and in which circumstances do they occur? What are the problems and pitfalls in the interpretation of results? - What are effective interventions to improve the interpretation of results?
Test results communication	<ul style="list-style-type: none"> - What elements are necessary for a good communication about test results? - What is the effect of improving communication on patient health outcomes? - What are the effects of testing for the purpose of patient reassurance, and how appropriate is it?

Test results management

Test results management involves the timely and correct processing of results and taking action if necessary. There are several reasons to advocate good test results management. Firstly, failing management may lead to medical errors, like important diagnoses being missed or delays in treatment^{3,4}. For instance, Dovey found that 82 of 284 errors in the process of health care delivery in American family practice were due to errors in the testing process⁵. Also, a majority of patients prefer to be informed about both normal and abnormal results, and swift contact about results is an important factor in satisfaction with care⁶⁻⁸. Moreover, doctors may be held legally accountable for providing timely and appropriate information about the results of tests ordered^{9,10}. A majority of doctors seem to agree with this view and many regard themselves and the medical system as being partly or fully responsible for appropriate management of test results¹¹. Remarkably, however, many doctors feel dissatisfied with their current management of test results¹².

Several researchers distinguish four elementary steps in the results management process¹¹⁻¹⁴. These seem reasonable and describe the process from the moment that the doctor receives the results until the appropriate actions have been undertaken. These four steps are: (a) tracking whether results of tests ordered have been received, (b) informing patients about the test results, (c) documenting that the patient has been informed and (d) ensuring that actions are undertaken if necessary and that follow-up is carried out¹³. These four steps can provide a basis for the development of a test results management system. However, there has only been little research about the management of test results, and some of these studies are in fact relatively old. But they all show that many practices do not achieve optimal management of test results. For example, two studies among doctors in family practice and internal medicine found that a relevant proportion of them (between 17% and 32%, depending on the type of test) used no or inadequate methods to track whether the results of tests they had ordered were received at the office^{11,14}. Several studies in various settings, including an emergency department and general practice, have shown that many patients (18%-36%) may not be correctly informed of abnormal test results^{11,15-17}. These studies also showed that few doctors (33%-39%), whether in hospital or family practice, used a method to ensure the necessary follow-up^{11,14}.

Several test results management systems have been developed, ranging from telephone message lines, tear-off sections of result forms and direct reports from the laboratory, to sophisticated computerised result management systems¹⁸⁻²². However, most interventions relate to just one type of test (e.g. pap smear test results or mammography), or to just one step in the management process. The implementation of a test results management system based on the above four basic steps might help to improve test results management in clinical practice. Ideally, a system should

encompass all four steps and should be applicable to all the results that doctors deal with. An example of such a system might be the Results Manager, a computer application developed by Poon et al. for outpatient departments²⁰. One of its functions is to help doctors notify patients of their results by producing results letters to patients, and it also provides reminders for follow-up in case of abnormal results.

Furthermore, there is the question if the physician or patient should be responsible for notification of results. Is it the physicians' responsibility to notify every patient of their results, even if these are normal? And to what extent is the physician responsible that the patient goes for follow-up tests after 3 months? Professional organisations or health institutions can help to formulate benchmarks with which practices should comply. The development of benchmarks for test results management should involve a comparative assessment of costs, such as expenditures and time investment, and benefits, such as fewer medical errors. After benchmarking and the implementation of such a system, a practice's performance should be regularly evaluated. This may help to consolidate the implementation of the system and provide an incentive for further improvement.

Doctors' interpretation of test results

Problems may arise not only in correctly managing the test results, but also in the interpretation of these results, as the complex nature of clinical practice means that many factors play a role in the interpretation. For example, various test results usually need to be interpreted in combination, various diagnostic hypotheses may be considered at the same time, co-morbidity can influence results, and reference values may differ according to patient characteristics like body weight, ethnicity, sex and age. Not surprisingly, results will not always be interpreted correctly. For example, several authors have described the possible effect of false-positive results, which may cause a cascade of further medical investigations to explain the initial abnormal result²³⁻²⁵. On the other hand, insufficient attention may be paid to abnormal results, for example because they deviate only slightly from reference values. Doctors may assume that such results are not significant, and the abnormalities may be not appropriately investigated^{26,27}.

A method which might help doctors to interpret results is to link the diagnostic hypothesis to the test results and subsequently evaluate the probability of the diagnosis. This approach is also referred to as Bayesian reasoning. Key concepts in this respect are pretest probability, sensitivity/specificity and posttest probability, but physicians frequently make mistakes in these calculations²⁸. They wrongly estimate pretest and posttest probabilities and tend to overestimate the diagnostic value of

tests^{29,30}. For example, Steurer found that only 60% of general practitioners knew the correct definition of sensitivity and positive predictive value, and only 22% correctly predicted the posttest probability³¹. If different hypotheses are being considered and multiple laboratory tests are involved, calculations may become too complicated. Indeed, Reid found that fewer than 10% of doctors use such calculations in routine practice³². That physicians do not use formal calculations may not be surprising, as extensive psychological research on medical decision making has shown that they often do not explicitly test hypotheses when establishing diagnoses. Instead, they may depend on a pattern of recognition or some form of automatic diagnostic retrieval that leads to a diagnosis^{33,34}. Such mechanisms may also be important in the interpretation of test results.

Apparently, then, doctors are not good at using calculations, and such calculations may become too complicated in many clinical situations. It may therefore not be very useful to try and improve doctors' interpretation of test results by teaching them the skills needed for such calculations. More attention to the way doctors interpret test results in routine practice may provide new viewpoints for improving this interpretation.

There has been little research in this area, but the literature mentions several principles that seem to play a role in the routine interpretation of results. For example, several authors have shown that general practitioners use tests especially to increase their certainty about the diagnosis and to reduce the number of diagnostic options³⁵⁻³⁷. Another finding is that doctors seem to work with variable thresholds when interpreting results, which may differ from laboratory reference values. For example, Norwegian general practitioners differed in their estimation of ESR values at which action was necessary³⁸. Doctors' thresholds for tests may also vary with the patients' age: haemoglobin values at which anaemia work-up is initiated are lower for elderly patients, and doctors who see more elderly patients also seem to use lower thresholds^{39,40}. In a vignette-based study, Thue found that doctors may also underestimate the influence of analytical and biological variation and the importance of previous test values⁴¹. Overall, however, there is limited information available about the way physicians interpret test results in routine practice. More research can help improve our understanding of the causes of mistakes in interpretation and can help find effective methods to improve interpretation.

Communication of test results

Studies on screening tests, such as antenatal tests, pap smears and mammography, have made it clear that patients have difficulties understanding the meaning and accuracy of results. This is illustrated by a study about screening for Down syndrome

in pregnancy; almost all patients (91%) drew incorrect conclusions regarding the risk of the syndrome from a positive result⁴². Difficulties with understanding the meaning and accuracy of test results contribute to increased health worries. This was shown, for instance, by Barton, who found increased health worries after false-positive results in mammography screening and an increase in health care utilisation^{43,44}. Although false-positive results may increase health worries, many patients also acknowledge that false-positives may occur in screening tests like mammography. In a cross-sectional survey in the United States, patients were highly tolerant of false-positives in mammography screening and 63% even accepted 500 false-positives per life saved^{45,46}. By contrast, a qualitative interview study about diagnostic blood tests by van Bokhoven found that patients often did not realise that a result could be false-positive or false-negative, that patients assumed false-positives to be impossible or rare, and that they thought that any abnormal result required further follow-up⁴⁷. Patients tend to interpret boundary values for tests, such as reference values for cholesterol level, as a fact against which they interpret their results⁴⁸. Furthermore, Hamm showed that patients make inaccurate estimates of pretest and posttest probabilities of common diseases and they make inaccurate estimates of sensitivity and specificity of common diagnostic tests. Lipkus showed that even highly educated patients have problems understanding probabilistic information^{49,50}. In conclusion, many patients do not understand the meaning or accuracy of test results, and good patient-doctor communication is essential for a better understanding of results. Unfortunately, the literature provides little evidence about the best way of discussing test results with patients.

If results of diagnostic tests are normal, doctors often hope that this will reassure their patient. Many doctors see normal results as a powerful tool for reassurance, and an important proportion of all requests for diagnostic investigations are also partly done for this purpose⁵¹. Research has indeed shown that explaining and discussing normal results reassures patients and that this may decrease healthcare costs. For example, Lucock found that both patients with high and low anxiety felt reassured after hearing that gastroscopy had shown no serious illness which was maintained after 1 year follow-up⁵². In a randomised controlled trial, Howard et al. investigated if patients with chronic daily headache were reassured by neuroimaging. They found that patients who had undergone MRI scans were (temporarily) less worried after 3 months, and that healthcare costs for these patients were significantly lower after one year than for those who had had no scan, as the former used fewer medical resources⁵³. However, the findings of recent studies on this issue have been inconsistent, and some studies found that patients may not feel reassured after a normal result⁵⁴. In one study, 28 patients were referred for echocardiography because of a heart murmur; 20 of these patients had normal test results but despite this, 11 were left with residual anxiety⁵⁵. A randomised controlled trial found that patients undergoing radiography of the lumbar spine for low back pain reported more back

pain after the radiography and had more consultations⁵⁶. Donkin investigated which patients are not reassured by normal results. She found that reassurance of patients undergoing exercise stress testing was significantly related to prior anxiety and illness perceptions, such as concerns about illness or symptoms experienced. This suggests that there may be a subgroup of patients for whom anxiety about symptoms or signs is inspired by social and psychological morbidity, resulting in residual anxiety despite normal results⁵⁷⁻⁶⁰.

Research agenda

Both in the management, interpretation and communication of test results, there is clearly some room for improvement. Many research questions are still left unanswered, and more attention to these questions is needed for a better use of test results.

Regarding the management of test results, researchers should develop methods to assess the quality of results management systems that are currently used. A basis for this is provided by the four steps for results management mentioned in the literature. Comparison of characteristics and achievements of management systems may then help professional organisations to develop benchmarks. In addition, errors due to malfunctioning systems must be investigated; how often do errors occur and what is the nature of these errors? Finally, further research is needed to investigate if test results management systems prevent errors and how they can be implemented in routine practice.

Regarding the interpretation of test results, research should initially aim to investigate the principles and determinants that underlie results interpretation. Important issues include the extent to which test characteristics like sensitivity and specificity and diagnostic probability estimates are included in doctors' routine test interpretation, and the way that doctors interpret test results in routine practice. Secondly, research should investigate the causes of mistakes and the circumstances in which they are made. For instance, why and when can abnormal results trigger a cascade of investigations? Or why and when are abnormal results ignored? Problems and pitfalls in the interpretation of results need a detailed description. Eventually, interventions should be developed to improve the interpretation of results in routine care, and further research will then be needed to evaluate the effectiveness of these interventions.

Regarding the communication of test results, it is necessary to investigate the elements in doctor-patient communication that help to increase patients' understanding of test results. Such elements may for example concern the ways to

accurately explain false positive and false negative results or the ways to accurately explain that disease probabilities influence test interpretation. Also, further research is needed to better understand the effects of improving communication on patient health outcomes such as patient reassurance, health anxiety and illness perceptions. Lastly, future studies should evaluate the effect and appropriateness of testing for reassurance. These findings may help doctors to understand when testing for reassurance could be useful or should be avoided.

References

1. Gandhi TK. Fumbled handoffs: one dropped ball after another. *Ann Intern Med* 2005;142:352-8.
2. Schoen C, Osborn R, Huynh PT, Doty M, Zapert K, Peugh J, Davis K. Taking the pulse of health care systems: experiences of patients with health problems in six countries. *Health Aff (Millwood)* 2005;Suppl(Web Exclusives):W5-509-25.
3. Wahls TL, Cram PM. The frequency of missed test results and associated treatment delays in a highly computerized health system. *BMC Fam Pract* 2007;8:32.
4. Garvey CJ, Connolly S. Radiology reporting--where does the radiologist's duty end? *Lancet*. 2006;367:443-5.
5. Dovey SM, Meyers DS, Phillips RL Jr, Green LA, Fryer GE, Galliher JM, Kappus J, Grob P. A preliminary taxonomy of medical errors in family practice. *Qual Saf Health Care* 2002;11:233-8.
6. Dolan NC, Feinglass J, Priyanath A, Haviley C, Sorensen AV, Venta LA. Measuring satisfaction with mammography results reporting. *J Gen Intern Med* 2001;16:157-62.
7. Meza JP, Webster DS. Patient preferences for laboratory test results notification. *Am J Manag Care* 2000;6:1297-300.
8. Baldwin DM, Quintela J, Duclos C, Staton EW, Pace WD. Patient preferences for notification of normal laboratory test results: a report from the ASIPS Collaborative. *BMC Fam Pract* 2005;6:11.
9. Birenbaum R. Who is responsible if a patient is not told of negative lab results? *CMAJ* 1989;141:970-2.
10. Berlin L. Communicating radiology results. *Lancet*. 2006;367:373-5.
11. Boohaker EA, Ward RE, Uman JE, McCarthy BD. Patient notification and follow-up of abnormal test results. A physician survey. *Arch Intern Med* 1996;156:327-31.
12. Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care. *Arch Intern Med* 2004;164:2223-8.
13. Hickner JM, Fernald DH, Harris DM, Poon EG, Elder NC, Mold JW. Issues and initiatives in the testing process in primary care physician offices. *Jt Comm J Qual Patient Saf* 2005;31:81-9.
14. Mold JW, Cacy DS, Dalbir DK. Management of laboratory test results in family practice. An OKPRN study. Oklahoma Physicians Resource/Research Network. *J Fam Pract* 2000;49:709-15.
15. Schofield MJ, Sanson Fisher R, Halpin S, Redman S. Notification and follow-up of Pap test results: current practice and women's preferences. *Prev Med* 1994;23:276-83.
16. Kelly MH, Barber JH. Use of laboratory services and communication of results to patients in an urban practice: an audit. *J R Coll Gen Pract* 1988;38:64-6.
17. Keren R, Muret-Wagstaff S, Goldmann DA, Mandl KD. Notifying emergency department patients of negative test results: pitfalls of passive communication. *Pediatr Emerg Care* 2003;19:226-30.
18. Mitchell H, Medley G. Notification of Pap smear results. A Victorian survey. *Aust Fam Physician* 1998;27(Suppl 1):S7-10.
19. Ridgeway NA, Ginn DR, Harvill LM, Hubbs DT, Massengill RM. An efficient technique for communicating reports of laboratory and radiographic studies to patients in a primary care practice. *Am J Med* 2000;108:575-7.
20. Poon EG, Wang SJ, Gandhi TK, Bates DW, Kuperman GJ. Design and implementation of a comprehensive outpatient Results Manager. *J Biomed Inform* 2003;36:80-91.
21. Stephenson P. Audit of a system for dealing with a practice's laboratory test results. *Br J Gen Pract* 1993;43:383-5.
22. Priyanath A, Feinglass J, Dolan NC, Haviley C, Venta LA. Patient satisfaction with the communication of mammographic results before and after the Mammography Quality Standards Reauthorization Act of 1998. *AJR Am J Roentgenol* 2002;178:451-6.
23. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health* 2002;23:23-44.
24. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician* 1996;25:971-3.
25. Mold JW, Stein HF. The cascade effect in the clinical care of patients. *N Engl J Med* 1986;314:512-4.
26. 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. *BMJ* 2001;322:276-8.
27. Henny J. Interpretation of laboratory results: the reference intervals, a necessary evil? *Clin Chem Lab Med* 2007;45:939-41.

28. Puhan MA, Steurer J, Bachmann LM, ter Riet G. A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Ann Intern Med* 2005;143:184-9.
29. Lyman GH, Balducci L. The effect of changing disease risk on clinical reasoning. *J Gen Intern Med* 1994;9:488-95.
30. Noguchi Y, Matsui K, Imura H, Kiyota M, Fukui T. Quantitative evaluation of the diagnostic thinking process in medical students. *J Gen Intern Med* 2002;17:839-44.
31. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002;324:824-6.
32. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104:374-80.
33. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ* 2002;324:729-32.
34. Kassirer JP. Diagnostic reasoning. *Ann Intern Med* 1989;110:893-900.
35. Zaat JO, Schellevis FG, van Eijk JT, van der Velden K. Do out-of-office laboratory tests affect diagnoses in general practice? *Scand J Prim Health Care* 1995;13:46-51.
36. Dinant GJ, Knottnerus JA, van Wersch JW. Diagnostic impact of the erythrocyte sedimentation rate in general practice: a before-after analysis. *Fam Pract* 1992;9:28-31.
37. Gronlie M, Hjortdahl P. The erythrocyte sedimentation rate; its use and usefulness in primary health care. *Scand J Prim Health Care* 1991;9:97-102.
38. Thue G, Sandberg S, Fugelli P. The erythrocyte sedimentation rate in general practice: clinical assessment based on case histories. *Scand J Clin Lab Invest* 1994;54:291-300.
39. Daly MP, Sobal J. Anemia in the elderly. A survey of physicians' approaches to diagnosis and workup. *J Fam Pract* 1989;28:524-8.
40. Nazareth I, King M. Decision making by general practitioners in diagnosis and management of lower urinary tract symptoms in women. *BMJ* 1993;306:1103-6.
41. Thue G, Sandberg S, Fugelli P. Clinical assessment of haemoglobin values by general practitioners related to analytical and biological variation. *Scand J Clin Lab Invest* 1991;51:453-9.
42. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *BMJ* 2006;333:284.
43. Barton MB, Moore S, Polk S, Shtatland E, Elmore JG, Fletcher SW. Increased patient concern after false-positive mammograms: clinician documentation and subsequent ambulatory visits. *J Gen Intern Med* 2001;16:150-6.
44. Marteau TM, Cook R, Kidd J, Michie S, Johnston M, Slack J, Shaw RW. The psychological effects of false-positive results in prenatal screening for fetal abnormality: a prospective study. *Prenat Diagn* 1992;12:205-14.
45. Schwartz LM, Woloshin S, Sox HC, Fischhoff B, Welch HG. US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *BMJ* 2000;320:1635-40.
46. Gram IT, Lund E, Slenker SE. Quality of life following a false positive mammogram. *Br J Cancer* 1990;62:1018-22.
47. van Bokhoven MA, Pleunis-van Empel MC, Koch H, Grol RP, Dinant GJ, van der Weijden T. Why do patients want to have their blood tested? A qualitative study of patient expectations in general practice. *BMC Fam Pract* 2006;7:75.
48. Adelsward V, Sachs L. The meaning of 6.8: numeracy and normality in health information talks. *Soc Sci Med* 1996;43:1179-87.
49. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making* 2001;21:37-44.
50. Hamm RM, Smith SL. The accuracy of patients' judgments of disease probability and test sensitivity and specificity. *J Fam Pract* 1998;47:44-52.
51. van der Weijden T, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002;52: 974-80.
52. Lucock MP, Morley S, White C, Peake MD. Responses of consecutive patients to reassurance after gastroscopy: results of self administered questionnaire survey. *BMJ* 1997;315:572-5.

53. Howard L, Wessely S, Leese M, Page L, McCrone P, Husain K, Tong J, Dowson A. Are investigations anxiolytic or anxiogenic? A randomised controlled trial of neuroimaging to provide reassurance in chronic daily headache. *J Neurol Neurosurg Psychiatry* 2005;76:1558-64.
54. van Kerkhoven LA, van Rossum LG, van Oijen MG, Tan AC, Laheij RJ, Jansen JB. Upper gastrointestinal endoscopy does not reassure patients with functional dyspepsia. *Endoscopy* 2006;38:879-85.
55. McDonald IG, Daly J, Jelinek VM, Panetta F, Gutman JM. Opening Pandora's box: the unpredictability of reassurance by a normal test result. *BMJ* 1996;313:329-32.
56. Kendrick D, Fielding K, Bentley E, Kerslake R, Miller P, Pringle M. Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial. *BMJ* 2001;322:400-5.
57. Donkin L, Ellis CJ, Powell R, Broadbent E, Gamble G, Petrie KJ. Illness perceptions predict reassurance following a negative exercise stress testing result. *Psychol Health* 2006;21:421-30.
58. Howard LM, Wessely S. Reappraising reassurance--the role of investigations. *J Psychosom Res* 1996;41:307-11.
59. Channer KS, James MA, Papouchado M, Rees JR. Failure of a negative exercise test to reassure patients with chest pain. *Q J Med* 1987;63:315-22.
60. Meechan GT, Collins JP, Moss-Morris RE, Petrie KJ. Who is not reassured following benign diagnosis of breast symptoms? *Psychooncology* 2005;14:239-46.

Chapter 3

GPS' considerations when interpreting laboratory test results: a qualitative study

PHH Houben, T van der Weijden, MA van Bokhoven, AEJ Droog, RAG Winkens,
RPTM Grol

Huisarts Wet 2005;48:326-32

Abstract

Objective

To examine considerations involved in general practitioners' (GPs') interpretations of laboratory test results.

Methods

Qualitative analysis of semi-structured interviews with 21 GPs about their considerations when interpreting the 10 most recent laboratory test results they had received.

Results

GPs' considerations when interpreting laboratory test results were highly varied. Considerations related to the working hypothesis, for instance to the estimated pretest probability of a disorder, or to the laboratory tests themselves, for instance to reference values or the level of abnormality of a result. There were often interactions between considerations from these two categories, especially in situations where a GP's working hypothesis was 'no disease' but the test results were abnormal. In such situations, the official reference values were not regarded as absolute and abnormal test results were often interpreted as normal.

Conclusion

The process of interpreting laboratory test results by GPs involves a variety of considerations, relating to aspects of the results and working hypothesis. As a result, abnormal test results do not always influence the diagnosis or management.

Introduction

General practitioners (GPs) order laboratory tests virtually on a daily basis¹. Previous research has shown that their motives to do so are highly varied; they do not use such tests purely for medical reasons, but often partly to reassure their patients²⁻⁴. When the test results come in, they need to be interpreted. How GPs do this and how they then use the results in their diagnostic workup and management remains largely unknown.

It is important to know GPs' considerations when interpreting test results, as such interpretations may lead to incorrect decisions. The literature provides examples of abnormal test results causing unjustified cascades of supplementary diagnostics and therapy⁵. Conversely, however, abnormal test results may be erroneously regarded as irrelevant, while normal results may be used to reassure patients without justification⁶. It is therefore useful to examine the way GPs interpret test results, to improve our understanding of the situations in which certain errors arise. Such an understanding may also help us design interventions to improve the use of laboratory tests in general practice⁷.

Laboratory test results can be interpreted on the basis of the pretest probability of disorders and of two characteristics of the tests: sensitivity and specificity. This does, however, involve some fairly complex calculations, which GPs may not be familiar with^{8,9}. Since these calculations are part of the curriculum of medical schools and GP training programmes^{10,11}, we would nevertheless expect concepts like pretest probability and test characteristics to play a certain role in GPs' everyday interpretations of test results. In addition, the reason for ordering a test may also affect the interpretation of its results. What we need to know is what role is played by these considerations, and what other considerations GPs use.

Since a study of the available literature did not yield any information on other considerations, we set up a qualitative study based on interviews with GPs. Our aim was to examine the considerations that play a role when GPs interpret laboratory test results.

Methods

We recruited a sample of GPs, and tried to ensure it would be representative in terms of general characteristics like the number of years of experience and one-doctor versus group practice. We approached 35 GPs in the region around the Dutch towns of Heerlen and Kerkrade, 21 of whom agreed to take part. Before the interviews, we asked the respondents to collect the 10 most recent laboratory test results, together

with all relevant patient data, and have them available during the interview. We explained to the GPs that we did not intend to evaluate the quality of their work. The actual interviews were held by one of the authors (AD), a medical student, who had studied literature on qualitative research and had been trained in interview techniques.

Results of screening or monitoring tests like cholesterol and HbA1c were not discussed. The interviews were semi-structured and included questions about the reasons for ordering a test, the preliminary diagnosis at the moment when tests were ordered, the interpretation of the results, the influence of reference values, sensitivity and specificity, the eventual diagnosis, the way the results were communicated to the patient and any sense of insecurity on the part of the GP. The interviewer encouraged respondents to discuss the aspects that they felt were important by means of an interview technique involving reflection, critical questions, open-ended questions and an open, inviting attitude on her part. The GPs had been asked to have all relevant information about the patients at hand during the interview. They received a financial remuneration for their participation.

Data analysis

All interviews were audiotaped and transcribed verbatim, after which the texts were checked against the tapes. The data were then processed by special software for qualitative data analysis (Atlas.ti). Two of the researchers (PH and TvdW) independently coded the interviews. Any coding discrepancies were resolved by consensus. The coding process, which was facilitated by the Atlas.ti program, resulted in categories, concepts and hypotheses. In a later stage of the analysis, we specifically searched for information that might confirm or refute the ideas we had developed. If the analysis yielded new information, the categories, concepts and hypotheses were adapted, altered or refined as needed. A third researcher (RW), who is a GP, read a random selection of one third of the interviews and assisted in the discussions. Data saturation was reached after 14–16 interviews. The remaining 5–7 interviews were nevertheless also analysed, to ensure that no new information would be found.

Results

Interviews were held with 21 GPs: 7 GPs were younger than 50 years; 9 GPs had fewer than 2500 registered patients; 10 GPs were, or had been, involved in training other GPs; 12 GPs worked in a one-person practice. One GP was female. The interviews lasted for an average of 55 minutes. The GPs ordered diagnostic tests from various regional laboratories.

Laboratory test results of 191 patients were discussed in the interviews. Their average age was 55 years (SD 19.9; range 7–95); 38% were male. More than half of the laboratory tests ordered involved assessments of haemoglobin, glucose, leucocytes, haematocrit, ESR and creatinine. More than 20% of the test results for glucose, cholesterol, gamma-GT, urea, creatinine, erythrocytes and TSH were abnormal (Table 3.1). The most common reason for ordering tests was to rule out diseases (Table 3.2).

Table 3.1 Numbers of tests and abnormal results.

Test	Number of tests (n=152) [†]	Abnormal results	
		Number below reference value *	Number above reference value *
Haemoglobin	95	8	0
Glucose	95	1	29
Leucocytes	88	1	8
Haematocrit	85	5	0
ESR	84	- [‡]	13
Creatinine	79	7	13
Cholesterol	73	-	48
MCV/MCH/MCHC	73	n.a. [¶]	n.a.
Thrombocytes	69	0	5
Erythrocytes	69	13	1
Triglycerides, HDL-cholesterol, LDL-cholesterol, Cholesterol/HDL ratio	64	n.a.	n.a.
ALT	60	-	10
TSH	44	1	9
AF	39	-	7
Gamma-GT	36	-	11
AST	36	-	4
Potassium	33	1	1
HbA1c	31	-	3
Sodium	28	1	3
Free T4	22	2	0
Differential leucocyte count	18	n.a.	n.a.
Urea	14	0	4
Iron	11	1	0
CRP	10	-	1

* The reference values used in this table were derived from the Dutch guideline *Diagnostisch Kompas* 1999/2000. Individual laboratories may use different reference values. [†] Four GPs copied the test results onto patients' 'green cards' or in a laboratory notebook. Their laboratory test results were excluded from the above table. This left 152 test orders. The table only includes those tests done at least 10 times. [‡] No lower limit. [¶] Not applicable.

Table 3.2 The five main reasons for ordering tests, in percentages (n=191).

Reason for ordering tests	percentage
Ruling out a particular diagnosis	21.7
Monitoring	20.1
At patient's request	15.3
Confirming a particular diagnosis	11.6
Reassuring the patient	11.6
Other reasons	19.7

We divided the considerations into two categories, one including considerations relating to the working hypothesis and the other including considerations relating to the laboratory test (Table 3.3). No further categorisation proved possible. The various considerations are discussed one by one below, as are the interactions between the two categories. They are illustrated with some quotations from the interviews. Abnormal results (in the laboratory's opinion) have been marked with an asterisk.

Table 3.3 Considerations that may influence the interpretation of test results.

Working hypothesis	Laboratory test result
A Disorder and diagnosis: <i>The significance of a test result is clearer if a clear working hypothesis has been formulated.</i>	A Test characteristics: <i>Some test results provide only a vague indication of possible disease. Concepts like the sensitivity and specificity of tests are occasionally taken into account.</i>
B Pretest probability: <i>Inferences drawn from tests results depend on how likely GPs think a disease is.</i>	B Reference values: <i>GPs sometimes regard a laboratory's reference values as too narrow, especially in relation to patients' age or sex.</i>
C Harmless comorbidity: <i>GPs may attribute abnormal test results to comorbidity.</i>	C Results are interpreted in sets: <i>Some test results are regarded as more important than others. GPs look at all results in a set of related tests.</i>
	D Continuous scale: <i>GPs evaluate results as part of a continuum ranging from normal to seriously abnormal.</i>
	E Previous results: <i>GPs relate test results to those of previous tests.</i>

Considerations relating to the working hypothesis

Considerations relating to the working hypothesis did not always play a role in the diagnostic decision process. The situations in which they did play a role are discussed below under the heading 'Interactions between the considerations'.

Disorder and diagnosis

GPs sometimes wanted to use laboratory tests to confirm or rule out a specific working hypothesis, such as rheumatoid arthritis or allergic rhinitis. In such cases, normal or abnormal test results had clear consequences.

Patient is a 22-year-old woman. Tests: gx3 grass mix3 20.50 U/ml (class 4); D1 housedust mites 21.60 U/ml (class 4)*.*

"A 22-year-old woman with respiratory tract complaints. Used to take medication for hay fever; never had allergy tests done; recently increased complaints of coughing and runny nose. So we decided to have this tested to

confirm the hay fever. So she's had allergy tests done, which showed that she was, let's see, positive for grasses and housedust mites. Both class 4 allergies. [...] So I'm having her come to my surgery to talk about sanitation and that sort of thing."

In many cases, however, the GPs did not formulate a working hypothesis, but searched the laboratory results for clues to the presence of somatic disorders. In such cases, they were frequently in doubt about the significance of abnormal test results.

Patient is a 32-year-old man. Tests: serum iron 5.6; glucose 6.0*; ESR 1; Hb 10.3; haematocrit 0.50; leucocytes 12.3*; differential leucocyte count: basophils 0%; eosinophils 2%; neutrophils 82%*; lymphocytes 11%*; monocytes 6%. "He tells me that he regularly feels ill, coughing, phlegm, constant colds, feeling tired, currently not really feeling fit. Well, he felt it needed to be looked into. Like so often. The tests didn't yield much useful information."*

Pretest probability

GPs tried to estimate the probability of a patient having a particular disorder. The more likely a GP thought a disorder was, the more likely he was to concur with the laboratory's opinion that a specific test result was abnormal, and to take further action. The GPs usually mentioned their reasons for ordering a test in conjunction with their estimation of the likelihood of a disorder. When they ordered tests mostly to reassure a patient, they estimated that a disorder was unlikely.

Below are three quotations about cases where the GP regarded the presence of a disease to be unlikely, doubtful and likely, respectively.

Patient is a 57-year-old woman. Tests: glucose 5.4; ESR 2; Hb 8.9; haematocrit 0.43; MCV 91; MCH 1.88; MCHC 20.7; erythrocytes 4.74; leucocytes 13.2; thrombocytes 331; differential leucocyte count: basophils 0.0; eosinophils 0.0; neutrophils 8.9*; lymphocytes 3.1; monocytes 1.0; TSH 3.63. "She's under a lot of stress, and I think that's where the complaints stem from. [...] The tests didn't show any abnormal results. [...] I know there's an asterisk against the leucocytes, but I decided not to follow up on that."*

Patient is a 42-year-old man. Tests: glucose 6.9; ESR 37*; Hb 9.5; haematocrit 0.45; MCV 92; MCH 1.96; MCHC 21.3; erythrocytes 4.84; leucocytes 8.9; HbA1c 6.1*. "Lower abdomen complaints. But they were not such that I got the idea that he should go to A&E immediately. So I was in two minds, in doubt what to do. [...] The ESR was not really characteristic of appendicitis, but in my opinion the leucocyte count was. I told him: 'Let's wait and see.'"*

Patient is a 75-year-old woman. Tests: creatinine 73; ALAT 10; glucose 5.9; ESR 30*; Hb 7.8; haematocrit 0.37; MCV 91; MCH 1.91; MCHC 20.9; erythrocytes 4.09; leucocytes 6.6; thrombocytes 241; TSH 0.78.*

“This patient had some vague complaints, like dizziness, palpitations. And what’s more important: she didn’t really look healthy to me. I felt she looked unwell. [...] I thought her ESR was rather high, at 30. [...] I think I’ll have to ask her some more detailed questions and then I’ll decide whether I want to have other things checked myself or whether, for instance, I’ll refer her to an internist...”

Harmless comorbidity

A patient’s comorbidity sometimes played a role in the interpretation of test results, particularly when the GP did not suspect a disorder but there were test results exceeding the laboratory’s range of reference values. Such abnormal results could then be explained by referring to harmless comorbidity. For instance, elevated leucocyte counts were frequently attributed to presumed mild and self-limiting viral infections. Another example is that the GPs might attribute abnormal values in liver function tests to excessive alcohol consumption.

Patient is a 46-year-old man. Tests: sodium 141; potassium 4.2; creatinine 82; alkaline phosphatase 86; gamma-GT 58; AST 24; ALT 39*; cholesterol 6.2*; glucose 5.8*; ESR 2; Hb 9.7; haematocrit 0.46; MCV 88; MCH 1.83; MCHC 21.0; erythrocytes 5.29; leucocytes 11.0*; thrombocytes 300; HbA1c 5.6*; TSH 1.06.*

“This is a patient in a risk category, with a history of alcohol abuse and smoking. [...] This time he has an abnormal gamma-GT. To me, this suggests he is still drinking.”

Considerations relating to the laboratory tests

This category includes those considerations used by GPs that relate to the nature of the laboratory tests. Again, these considerations did not always play a role.

Test characteristics

The GPs distinguished various types of test. Some, like glucose and TSH, were specifically used to confirm or rule out a particular disease, while others, like ESR and leucocyte count, yielded more general information about possible somatic problems. This second category of tests only provided the GPs with some indication of the possible presence of a disease. Results of such non-specific tests which exceeded the range of reference values used by the laboratory did not always prompt the GPs to change their diagnosis or management. In addition, the sensitivity and specificity of a test sometimes played a limited role. Though the GPs did not always use these specific terms, they might mention related concepts like false-positive results and laboratory

errors. Such concepts were mainly used to explain abnormal results and interpret them as 'probably normal'.

Patient is a 20-year-old woman. Tests: BSE 4; Hb 8.7; haematocrit 0.42; MCV 92; MCH 1.9; MCHC 20.8; erythrocytes 4.58; leucocytes 9.4; thrombocytes 276; differential leucocyte count: lymphocytes 25%; monocytes 9%; basophils 0%; eosinophils 2%; neutrophils 64%.*

"To reassure the patient. She had recurrent throat complaints. This [pointing at the monocyte count] signifies nothing to me; it is 9, up to 8 is normal, so that's irrelevant. There are several test values that are irrelevant. This doesn't mean much to me. Slightly elevated. Monocytes, unless they're extremely increased ... that's hardly relevant. What I was mostly interested in was any abnormal cells to do with a blood disease. Like leukaemia, or a clearly abnormal differential leucocyte count, the lymphocytes and granulocytes and their ratio."

Reference values

In the GPs' view, some of the reference values used by the laboratories had too narrow a range for a primary care population. Some GPs felt that the reference values are based on research among hospital patients and are therefore unsuitable for general practice. In addition, a patient's age and sex play a role in deciding whether a test result is normal or abnormal, and the GPs felt that these aspects were not always taken into account in the laboratories' reference values. They often mentioned that the reference values for elderly people were too strict. Nor did the reference values always differentiate between men and women. Hence, the GPs often used their own reference values.

Patient is a 69-year-old man. Tests: urea 4.9; creatinine 85; total bilirubin 10.8; AF 83; ALT 12; glucose 5.1; ESR 19; Hb 9.6; leucocytes 7.3; thrombocytes 265; PT INR 1.1*; HbA1c 5.5*.*

GP: "My working hypothesis is that this man is suffering from depression but is afraid he has a somatic disease. This ESR, that's another one of those differences between the clinic and what we regard as normal. We think an ESR of 19 is normal for a 69-year-old man, while the clinic doesn't."

I: "Up to what ESR value do you consider to be normal?"

HA: "In a 69-year-old man? His age divided by two. So that's 35. For women it's the age plus 10 divided by two. That's the rule among GPs. Somehow there's a discrepancy between what we use and what they use at the hospital."

Results are interpreted in sets

Laboratory tests often involve sets of related tests for a particular organ system or a pathophysiological process. Examples of such sets include ESR, leucocyte count, and differential leucocyte count, or gamma-GT, ALT, and AST. The GPs sometimes interpreted individual results and sometimes sets of results. An individual abnormal result was sometimes regarded as irrelevant if the other results in the set were normal. In addition, the GPs distinguished a certain hierarchy within sets of tests. If they order Hb, laboratories often automatically also provide results on MCV, MCHC and erythrocytes. When the results of one of these three extra tests exceeded the range of reference values, the GPs sometimes paid little attention to these results. The test that was higher up in the hierarchy was considered to be more decisive for the diagnosis and management.

Patient is a 47-year-old woman. Tests: creatinine 85; ALT 61; ESR 11; Hb 8.1; haematocrit 0.4; erythrocytes 4.56; MCV 87; MCH 17.7; MCHC 20.3; leucocytes 6.9; differential leucocyte count: neutrophils 39.1%*; lymphocytes 56.1%*; monocytes 3.4%; eosinophils 1.4%; basophils 0.0%; EB-mono-slide negative; EBV-VCA IgM negative.*

"A 47-year-old woman with general malaise, the kind of complaints that go with viral infections. Slightly elevated ALT. She has lymphocytosis, at least in the differential leucocyte count. Total leucocyte count is normal, 6.9. She has a relatively large number of lymphocytes and rather few neutrophils. So there's a shift in the blood count. But the mononucleosis tests are negative. So it's not glandular fever. I think this woman is going through a persistent viral infection that's taking a bit longer to clear up than expected."

Continuous scale

Although the laboratories marked abnormal results with an asterisk, GPs did not interpret results in such a dichotomous way. Instead, they saw them as part of a continuum ranging from normal to severely abnormal, with a whole range of options in between. Whether a GP classified a particular result as normal, dubious or abnormal depended on the patient and on the consultation. It was not only individual results which were interpreted on such a continuous scale, but also sets of results, such as those of kidney or liver function tests. In addition, other considerations played an important role in the eventual assessment whether a result or set of results was considered normal or abnormal.

Patient is a 55-year-old man. Tests: BSE 34; Hb 9.3; haematocrit 0.44; erythrocytes 4.5; MCV 99; MCH 20.7; MCHC 20.9*

"[This patient] has been coughing up greenish yellow sputum for the last four to five years, and has been wondering whether it might be tuberculosis. His

ESR was slightly elevated. An ESR of 22 to me means slightly elevated but within the normal range, so I take no further action. But 34 is a bit above normal. I'm not going to intervene immediately, but I will monitor it."

Previous results

Whenever possible, the GPs related results to previous results for the same patient. They checked whether abnormalities increased or decreased and over what period of time.

Patient is a 60-year-old man. Tests: glucose 6.7; creatinine 87; gamma-GT 12, BSE 22*; Hb 9.5; haematocrit 0.47; erythrocytes 4.81; MCV 98; MCH 19.7; MCHC 20.1; leucocytes 8.8.*

"[These tests were done] to reassure the patient and to reassure the doctor, and there are also tactical considerations involved when you're dealing with [these] vague complaints. If you look at the previous ESR values, 23, 31, 22, it's clear that this man could have a slightly higher than normal ESR. Yeah, I take these things into consideration. Well, the other values are OK. HB, haematocrit, all perfectly fine. I also compare with other results. And I check against past results."

Interactions between considerations

GPs interpret laboratory test results on the basis of the working hypothesis which led them to order the tests. This means that there are two options: the results may be in agreement with the working hypothesis or they may conflict with it.

The results are in agreement with the working hypothesis

In those cases where the GPs expected no abnormalities when they ordered the laboratory diagnostics, the tests did indeed usually yield no abnormal results. There were also cases in which the GPs expected a disease to be present and in which the laboratory tests confirmed this suspicion. What these two situations have in common is that the test results were in agreement with the GPs' assumptions about the presence of disease. In other words, the test results confirmed the GPs' diagnosis and there was no reason for further interpretation. The GPs then immediately moved on to management, and the above considerations played a minor role.

Patient is a 31-year-old woman. Tests: uric acid 0.18; gamma-GT 13; AST 20; ALT 19; glucose 5.2; BSE 8; Hb 8.3; haematocrit 0.4; MCV 86; MCH 1.78; MCHC 20.8; erythrocytes 4.67; leucocytes 6.2; thrombocytes 195; TSH 1.47.

"A 31-year-old woman. She's tired, drained, absolutely no energy, she's afraid there's a somatic cause, has gained weight. It's the thyroid again, right, all thyroids here. There's also a swollen metatarsal-I joint in the left foot. I had

uric acid checked to rule out gout. Well, I thought I might as well put that in too, but that was not abnormal. It just confirms that there is no somatic problem, no thyroid problem, glucose is fine, cholesterol is fine, no gout. That's something I can tell people: you don't have to worry about that."

The results do not agree with the working hypothesis

There were many cases in which the GPs did not expect a disease to be present when they ordered the laboratory diagnostics, but one or more results nevertheless turned out to exceed the reference values. By contrast, there were very few cases where the GPs suspected a disease to be present but all test results proved normal. In both situations, the GPs' assumptions about the presence of disease were not in agreement with the test results. It is in these situations that all of the considerations outlined above played an important role in interpreting the results: the GPs tried to make the results fit their working hypothesis. If GPs had assumed that a patient had no disease, they would have been likely to interpret any results exceeding the laboratory's reference values as normal. In some of these cases, the GPs decided to have the test repeated at a later time.

Patient is a 33-year-old woman. Tests: creatinine 56; total bilirubin 30.7; AF 55; ALT 14; CK 54; glucose 4.9; ESR 2; Hb 8.7; haematocrit 0.42; MCV 94; MCH 1.92; MCHC 20.5; erythrocytes 4.53; leucocytes 6.9; thrombocytes 243; differential leucocyte count: basophils 0%; eosinophils 0%; neutrophils 76%*; lymphocytes 17%; monocytes 6%; TSH 2.17; FT4 13.3.*

GP: "A 33-year-old woman. Fibromyalgia, chronic fatigue syndrome [...] It's got worse. [...] There are two abnormal values, but I'm not going to follow up on those. Total bilirubin is of no significance at all, might just as well not have ordered that."

I: "Why did you order it?"

GP: "Just routine. It's more a matter of ruling out than detection. I've interpreted the abnormal values as follows: those neutrophils are probably a lab error. And the bilirubin has no clinical consequences in my view. And the rest is normal; no abnormalities."

I: "And has this influenced your diagnosis?"

GP: "No, that remains the same [...] I think I'm going to tell her that there are no further physical problems, and I hope that will reassure her and just reduce the problem to the chronic fatigue. So we might be able to assist her with that."

Discussion

Main findings

Our results show that interpreting the results of laboratory tests is a complex process. The GPs used a wide variety of considerations, particularly in problem situations, for instance if the tests yielded unexpected abnormal results. These considerations were used to put isolated abnormal values into perspective. As a result, such abnormal results often did not make the GPs more likely to consider a particular disorder, nor did they change their diagnosis. In situations with a low pretest probability, the GPs would explain abnormal results by referring to reference values being too strict or to the fact that other results were normal.

Strengths and weaknesses of this study

The explorative nature of this qualitative study allowed us to identify a wide variety of considerations used by the GPs. In our view, the GPs felt free to talk about laboratory diagnostics to a medical student who, although a future colleague, would not be expected to criticise them¹². Obviously, we had to instruct this student carefully about qualitative research methods, and she had to be trained in interview techniques. We instructed her to inquire further into the underlying reasons for the answers the GPs gave. This led to a wealth of information in the taped interviews. The student had indeed inquired about the reasons underlying the GPs' interpretations in nearly every case. It is our impression that the GPs did try to formulate serious answers to these questions, since the tapes show that each of the GPs spoke to the interviewer about the laboratory results for nearly an hour.

We tried to compose our sample in such a way that it represented a variety of general characteristics like gender, number of years of experience and practice situation (single-doctor or group practice). Since this was a qualitative study, it was not essential to have a representative sample, but it was important to have all relevant categories represented¹³. We managed to achieve this, except as regards the male-female ratio, since only one female GP participated. We were therefore unable to assess whether female GPs use different considerations than their male colleagues, or whether they might use the considerations we identified in a different way than the men.

Comparison with previous research

GPs are known to frequently order laboratory tests just to reassure patients, or because patients ask for them³. In the present study, this happened in 11.6 and 15.3% of the consultations, respectively, percentages which are in agreement with those reported by previous studies^{14,15}. The types of test ordered by the GPs in our study

also correspond to those reported by previous studies of the test-ordering behaviour of Dutch GPs, as does the frequency of abnormal results^{1,16}.

Significance of the findings for practice and future research

Our findings show that GPs use a variety of considerations in interpreting test results, and that these interpretations are affected by the estimated pretest probability and the degree to which results deviate from reference values. As a result, the GPs often interpret results exceeding the laboratory's reference values as normal if the estimated pretest probability of a disorder is low. Quantitative research could be used to assess the magnitude of this effect, and to evaluate, in view of the eventual diagnosis, whether the GPs' interpretations of test results prove justified with hindsight.

Our findings are contradicted by a few case reports showing that unjustified tests yielding unexpected abnormal values may lead to a cascade of further tests and therapeutic interventions^{5,17}. These reports did not mention how frequently such situations arise. Although the present study did not yield any specific information about the frequency with which such cascades of tests and therapeutic interventions occur, it does seem to indicate that such overactive follow-up policies may not be all that common. In any case, we think it would be useful to design a quantitative study to assess the diagnostic and therapeutic consequences of abnormal laboratory test results in situations with a low pretest probability.

References

1. Kluijnt I, Zaat JOM, Van der Velden J, Van Eijk JThM, Schellevis FG. Voor een prikje? Het gebruik van klinisch-chemische, hematologische en serologische bepalingen door huisartsen. *Huisarts Wet* 1991;34:67-71.
2. Van der Weijden T. Het aanvragen van laboratoriumtests bij patiënten met onbegrepen klachten. Op het kruispunt van evidence-based en practice-based geneeskunde. *Huisarts Wet* 2004;47:404-7.
3. Van der Weijden T, Van Bokhoven MA, Dinant GJ, Van Hasselt CM, Grol RPTM. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002;52:974-80.
4. Van der Weijden T, Van Velsen M, Dinant GJ, Van Hasselt C, Grol R. Unexplained complaints in general practice. Prevalence, patients' expectations, and professionals' test ordering behavior. *Med Decis Making* 2003;23:226-31.
5. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician* 1996;25:971-3.
6. 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. *BMJ* 2001;322:276-8.
7. Foy R, Warner P. About time: diagnostic guidelines that help clinicians. *Qual Saf Health Care* 2003;12:205-9.
8. Steurer J, Fisher JE, Bachmann LM, Koller M, Ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002;324:824-6.
9. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104:374-80.
10. Van der Weijden T, Van den Akker M. Dwalingen in de methodologie. XXXII. Foute testuitlagen. *Ned Tijdschr Geneesk* 2001;145:906-8.
11. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med* 1998;73:538-40.
12. Chew-Graham CA, May CR, Perry MS. Qualitative research and the problem of judgement: lessons from interviewing fellow professionals. *Fam Pract* 2002;19:285-9.
13. Malterud K. Qualitative research: standards, challenges, and guidelines. *Lancet* 2001;358:483-8.
14. Van Boven K, Dijksterhuis P, Lamberts H. Defensive testing in Dutch family practice. Is the grass greener on the other side of the ocean? *J Fam Pract* 1997;44:468-72.
15. Dijksterhuis PH, Van Boven C, Lamerts H. Waarom aanvullend onderzoek? De functie van het aanvullend onderzoek voor de huisarts. *Huisarts Wet* 1994;37:467-72.
16. Dijksterhuis PH, Van Boven C. De schatbare waarde van aanvullend onderzoek in de huisartspraktijk. Lelystad: Meditekst, 1993.
17. Brigden ML, Heathcote JC. Problems in interpreting laboratory tests. What do unexpected results mean? *Postgrad Med* 2000;107:145-6,151-2,155-8.

Chapter 4

Reasons for ordering laboratory tests and relationship with frequency of abnormal results

PHH Houben, RAG Winkens, T van der Weijden, RCRM Vossen, AJM Naus, RPTM Grol

Scand J Prim Health 2010;28:18–23

Abstract

Objective

Laboratory tests are ordered on a daily basis, even though disease probability is often very low. Abnormal results, especially mildly abnormal results, can be difficult to interpret in these circumstances. Further insights into the occurrence of abnormalities can help improve rational test ordering and test interpretation. The objective was therefore to examine the frequency of mildly and markedly abnormal results and their relation with physicians' reasons for ordering tests.

Design

Prospective study.

Participants

A total of 87 primary care physicians in the Netherlands collected data on 1775 patients.

Measurements

The physicians recorded the reason for ordering the tests, the most probable diagnosis and the pretest probability. The laboratories' reference values and specified 'action limits' were used to assess the number of abnormal results and markedly abnormal results, respectively.

Results

Laboratory results were received for 1621 patients and 15,603 tests were reported (mean 9.6). The proportion of abnormal test results increased with increasing pretest probability (from 13.9% to 34.7%) and was 13.4% for tests ordered to reassure the patient and 13.3% for psychosocial diagnoses. The proportion of patients with at least one abnormal test result was high: 53.1% for tests ordered to reassure and 57.7% in patients with low pretest probability. Corresponding values for a marked abnormality were 11.1% and 12.4%, respectively.

Conclusion

Abnormal laboratory test results were frequent, even when pretest probability was low. Physicians should therefore carefully consider when tests are necessary. Future research could explore physicians' interpretation of test results and its impact on diagnosis and management.

Introduction

In primary care, correct interpretation of abnormal test results may be difficult. For instance, abnormal results can sometimes lead to an unjustified cascade of further investigations^{1,2}, while on the other hand, they may erroneously be failed to be further investigated³.

These difficulties with the interpretation of abnormal results may have several causes. For example, the pretest probability of disease is usually low in primary care. Also, physicians may suspect that complaints and symptoms are related to psychosocial problems^{4,5}. Many tests are indeed ordered for reasons such as excluding pathology or reassuring the patient⁶⁻⁸. Furthermore, many abnormal lab results represent only minor deviations, and physicians may doubt their clinical relevance⁹. It is in particular the combination of minor abnormalities and a low pretest probability which may be difficult to interpret, as it is often easier to decide on the basis of more markedly abnormal results.

While several authors reported high percentages of abnormal laboratory results^{10,11}, the relation with the physicians' reasons for ordering tests and the frequency of mildly or markedly abnormal results are unknown. More information on this relation could be important for education as well as to improve rational test ordering. It may also help exploring areas of potential difficulty in the interpretation of results.

The goal of this study was therefore to examine the frequency of mildly and markedly abnormal results in routine practice, and to investigate how they relate to the reasons for ordering laboratory tests.

Methods

Design and setting

We conducted a prospective study among 87 primary care physicians and their patients in 7 rural, suburban and urban areas in the south of the Netherlands. Physicians' age, sex, and working time were representative to those in the Netherlands according to data from the Netherlands Institute for Health Services Research (NIVEL). To prevent selection bias, each participating physician was personally instructed to include and to record data on the first 25 consecutive adult patients for whom they had decided to order laboratory tests. Also, to prevent physicians from changing their ordering patterns, it was stressed that we did not intend to measure performance or to give feedback on test ordering. Physicians working part-time included a smaller number of patients, proportional to the number of hours a week they were working. Patients were asked to give informed consent.

Measurements

The physicians recorded data when they ordered the laboratory tests, using forms that were specifically designed for the study and took about two minutes to complete. The forms had been pilot-tested and evaluated as regards validity, reliability and user convenience in an iterative process among a sample of ten primary care physicians and a questionnaire expert. The laboratories provided the researchers with a copy of the laboratory results.

Outcome variables

Test results

We used the reference values from the laboratories, adjusted for age and sex where appropriate, to determine the number of abnormal results. We asked an expert group to help us discriminate between minor and marked abnormalities. This expert group comprised two clinical chemists, one primary care physician with expertise in laboratory testing and two primary care physicians from the research team. The experts went through a consensus procedure to specify action limits for tests, indicating values for which they considered urgent medical intervention to be necessary. The action limits were established in an iterative process and were based as much as possible on evidence-based guidelines and the literature. After 3 discussion rounds, consensus about the action limits for 14 tests was reached (Table 4.1).

Independent variables

Reason for ordering tests

We distinguished 9 reasons for test ordering, which were chosen on the basis of a qualitative interview study among primary care physicians¹². Physicians recorded the most important reason for ordering the investigations by ticking one of nine boxes. We summarized these into 5 categories: (1) to exclude disease and reduce the physician's uncertainty, (2) to confirm diagnosis and to determine treatment, (3) to reassure patients and at patient's request, (4) monitoring of disease (screening for hypertension/cholesterol/diabetes and check-up for known disorder), and (5) other reasons.

Estimated pretest disease probability

The form physicians had to complete at the time of test ordering asked: 'Do you suspect a disease?' The physicians answered on a 5-point Likert scale: 'definitely not', 'probably not', 'maybe', 'probably yes' and 'definitely yes'.

Presumed diagnosis

Physicians recorded their most likely diagnosis at the time of test ordering. We coded these according to the International Classification of Primary Care (ICPC), defining two

groups: (1) the most likely diagnosis refers to non-somatic disease, i.e. ICPC chapters P (psychological) and Z (social) and ICPC code A97 ('no disease'); and (2) all other diagnoses.

Analysis

We used SPSS 15 to analyze the data. We counted the number of abnormal test results according to the reference values and the number of markedly abnormal test results based on the action limits (test level). We also counted the number of patients with at least one abnormal or one markedly abnormal test result (patient level). We used the chi-square test to test for significant differences in the numbers of patients with abnormal results within the categories of age, sex, reason for ordering tests, presumed diagnosis and estimated pretest probability.

Results

A total of 87 primary care physicians participated in the study, and they included 1775 patients. We received no laboratory results for 154 patients (8.7%), the most important reasons being failure to visit the laboratory (31.8%) and failure to give informed consent (13.6%), while the reason could not be retrieved in 42.9% of the cases. The laboratories reported 15,603 tests for the remaining 1621 patients. The mean of reported tests was 9.9 for tests ordered to exclude disease, 8.9 for tests ordered to reassure, and 7.9 for tests ordered for monitoring. There were 2875 (18.4%) abnormal test results according to the reference values. The 1621 patients included 440 (27.1%) with one abnormal test result, 310 (19.1%) with two abnormal test results and 437 (27.0%) with three or more abnormal test results.

The tests for which action limits had been specified accounted for 47.4% (n=7400) of all reported tests (n=15,603). There were 1484 (1484/1621=91.5%) patients for whom at least one of these tests was ordered. Table 4.1 shows how often each test was reported and the percentage of abnormal and markedly abnormal test results.

At test level, the frequency of markedly abnormal test results was clearly smaller than the frequency of abnormal test results (Table 4.2). The frequency of abnormal test results increased with increasing age (13.7-25.8%) and increasing pretest probability (13.9-34.7%), and was high for tests ordered to confirm disease (27.2%) and for screening (27.0%).

At patient level, the frequency of patients with one or more abnormal test results was considerable (Table 4.3). Even if the physician had ordered to reassure, 53.1% of the patients had abnormal results and 11.1% had a markedly abnormal result. Patients' age and sex, the reason for ordering tests and the estimated pretest probability were significantly related to the percentage of patients with abnormal results.

Table 4.1 Reference values, action limits and abnormal results.

Test	N tests	% abnormal test results according to reference values (95% CI)	Action limits	% markedly abnormal test results according to action limits (95% CI)
ALAT (U/l)	440	10% (8-13%)	<70	3% (1-4%)
Cholesterol (mmol/l)	584	38% (34-41%)	<8.0	3% (1-4%)
Creatinine Clearance (ml/min/1.73 m ²)*	780	61% (58-65%)	<60	21% (18-24%)
CRP (mg/l)	175	38% (31-45%)	<50	5% (1-8%)
ESR (mm/hr)	877	22% (19-25%)	<40	5% (3-6%)
Ferritin (µg/l)	57	47% (34-60%)	15-1000	35% (23-47%)
Gamma GT (U/l)	239	24% (19-30%)	<90	6% (3-9%)
Glucose (mmol/l)	949	8% (6-9%)	3.0-11.0	2% (1-3%)
Hemoglobin (mmol/l)	878	13% (11-16%)	♀ 6.5-10.5, ♂ 7.5-11.5	2% (1-3%)
Leukocytes (*10e9/l)	574	13% (10-25%)	2.5-15.0	2% (1-3%)
MCV (fl)	716	6% (5-8%)	75-105	1% (0-2%)
Potassium (mmol/l)	334	16% (12-20%)	3.0-5.5	1% (0-3%)
Sodium (mmol/l)	137	7% (2-11%)	130-150	0%
TSH (mU/l)	660	11% (9-13%)	0.1-8.0	5% (3-7%)

* MDRD formula, adjusted to the laboratory method for creatinine measurement (jaffe/enzymatic).

Table 4.2 The distribution of abnormal test results at test level.*

	N tests	% abnormal test results according to reference values	% markedly abnormal tests according to action limits
Patients (n=1484)	7400	20.7%	4.9%
Sex			
men	2857	21.4%	4.3%
women	4543	20.3%	5.3%
Age, years			
<40	1747	13.7%	2.5%
40-60	2773	19.8%	4.0%
> 60	2880	25.8%	7.4%
Reason for ordering tests			
confirm / determine treatment	725	27.2%	9.1%
exclude / physician's uncertainty	3654	18.6%	4.2%
reassure / patient's request	1139	13.4%	2.0%
monitoring (screening / check-up)	1761	27.0%	6.6%
other reasons	33	15.2%	3.0%
Estimated pretest probability [†]			
definitely no disease	541	13.9%	2.2%
probably no disease	1990	14.2%	2.6%
possibly disease	1651	19.9%	4.5%
probably disease	1127	24.3%	7.0%
definitely disease	196	34.7%	14.3%
Presumed diagnosis [†]			
A97, P, Z	1083	13.3%	2.7%
other ICD9 codes	4350	20.1%	4.9%

* Limited to the 14 tests of Table 4.1. [†] The physicians did not estimate a pretest probability nor mention a presumed diagnosis if they ordered the tests for screening or check-up.

Table 4.3 The distribution of abnormal result at patient level.[†]

Reasons for ordering tests	N patients	% patients with abnormal results according to reference values (95% CI)	% patients with markedly abnormal results according to action limits (95% CI)
Patients	1484	64% (61-66%)	21% (19-23%)
Sex			
men	591 (40%)	63% (60-67%)	18% (15-21%) *
women	893 (60%)	64% (61-67%)	23% (20-26%)
Age, years			
<40	343 (23%)	48% (42-53%) ***	10% (7-13%) ***
40-60	582 (39%)	60% (57-64%)	16% (13-20%)
> 60	559 (38%)	77% (74-81%)	33% (29-37%)
Reason for ordering tests	16 missing		
confirm / determine treatment	167 (11%)	68% (61-75%) ***	30% (23-37%) **
exclude / physician's uncertainty	651 (44%)	63% (59-66%)	20% (17-23%)
reassure / patient's request	207 (14%)	53% (46-60%)	11% (7-15%)
monitoring (screening / check-up)	434 (29%)	69% (65-73%)	24% (20-27%)
other reasons	9 (1%)	33% (7-69%)	11% (0-48%)
Estimated pretest probability [‡]	28 missing		
definitely no disease	97 (9%)	58% (48-68%) ***	12% (6-19%) ***
probably no disease	353 (34%)	54% (48-59%)	14% (10-17%)
possibly disease	301 (29%)	66% (60-71%)	22% (17-27%)
probably disease	218 (21%)	69% (63-75%)	28% (22-33%)
definitely disease	53 (5%)	70% (57-82%)	40% (26-53%)
Presumed diagnosis [‡]	38 missing		
A97, P, Z	173 (17%)	58% (51-66%)	16% (10-21%)
other ICPC codes	839 (83%)	62% (59-65%)	21% (18-24%)

* $P < 0.05$ ** $P < 0.01$ *** $P < 0.001$; [†] Limited to the 14 tests of table 4.1; [‡] The physicians did not estimate a pretest probability nor mention a presumed diagnosis if they ordered the tests for screening or check-up.

Discussion

This study shows that many patients have abnormal laboratory test results, even when the tests are ordered in a situation of low pretest probability, such as to reassure patients (53.1%) or to exclude disease (62.5%). A large proportion of the abnormal test results were only mildly abnormal. Given the low pretest probability and the statistical definition of reference values, there is a fair chance that the abnormalities may not have clinical significance in terms of diagnosis or therapy¹³. This underlines the necessity for physicians to carefully consider their orders for laboratory tests. It also raises the question what exactly physicians do with these results. Future research should therefore attempt to explore how physicians interpret these test results, and what the diagnostic and therapeutic consequences are.

The total proportion of patients with an abnormal test result in our study was 73.2%. On the whole, our results are in line with those of others in terms of the high frequency of abnormal results found^{10,11,14,15}. However, the test for ferritin was abnormal for comparatively large numbers of patients. This may be due to the fact

that many laboratories in the Netherlands only report ferritin when hemoglobin is too low. The low number of abnormal test results when action limits are applied may relate to the fact that in these cases the statistical chance for finding an abnormality is very limited.

Strong points of this study were that it included many physicians and that they were a representative sample of those in the Netherlands. Also, the data were prospectively collected and we tried to prevent selection bias by careful instruction of the physicians. However, Dutch primary care physicians tend to order fewer laboratory tests than those in other countries¹⁶. In countries where physicians tend to order more tests, there will probably be even more patients with abnormal results, underlining the need for rational test ordering and a better understanding of the influence of results on further diagnosis and therapy.

Since the action limits that were specified by the expert group might be criticized, we also examined the possibility of defining action limits by using 99% reference values, critical differences and a computerized expert system (Valab)¹⁷. However, none of these methods allowed us to specify action limits for our data corresponding to clinically significant abnormalities. Furthermore, varying our action limits caused only slight changes in the number of markedly abnormal results, which would not have changed our conclusions. We therefore think that the use of the action limits defined by the expert group was the best option for this study.

The widespread test ordering in low pretest probability may be discussed. It usually adds little information¹⁸ and it increases costs while many test orders in low pretest probability may be considered unnecessary^{1,2,19-26}. Also, some evidence indicates that it may give rise to diagnostic problems due to the frequent occurrence of marginal abnormal results^{1,2,5}. However, at the same time, sporadic but important laboratory abnormalities may be found, despite the low pretest probability. So, reducing test ordering in low pretest probability may reduce costs, but it must be done carefully to avoid underdiagnosis.

A way to overcome the problem of misleading test results may be by using decision limits instead of reference values. Decision limits are based on consensus linking test results to clinical outcomes. A problem is that with different analytical systems used by laboratories, the decision limits may also vary, just like reference values may vary. Also, decision limits vary for specific clinical cases. This is complex to present and may negatively influence the presentation of laboratory reports. In fact, there is still much discussion in clinical laboratory literature, and in other areas such as diagnostic imaging, with respect to appropriate helpful methods for interpretation of test results, as all methods meet specific difficulties^{27,28}.

In their training, physicians often learn to interpret results by means of computations involving pretest probability, sensitivity and specificity, which allow the posttest probability to be calculated^{29,30}. This may, however, be difficult to apply in the complex environment of routine practice, where the results of several tests are usually reported, and where several test results may be abnormal³¹. Greater insight

into the principles that underlie the interpretation of results in routine care might help to develop better methods to support physicians in interpreting abnormal results.

In conclusion, physicians are likely to be confronted with many marginally abnormal test results, even if tests were ordered to reassure or in case of low pretest probability. Given the low pretest probability and the statistical definition of reference values, there is a fair chance that these abnormalities have no clinical significance. Therefore, physicians should carefully consider whether ordering a test is necessary. Furthermore, because the consequences of these abnormalities are unclear with respect to diagnosis and further management, future research should help to better understand physicians' interpretation of test results and its impact on diagnosis and management.

References

1. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician* 1996;25:971-3.
2. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health* 2002;23:23-44.
3. 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. *BMJ* 2001;322:276-8.
4. 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;15:264-5.
5. Aiarzaguena JM, Grandes G, Salazar A, Gaminde I, Sanchez A. The diagnostic challenges presented by patients with medically unexplained symptoms in general practice. *Scand J Prim Health Care* 2008;26: 99-105.
6. van der Weijden T, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002;52: 974-80.
7. van Bokhoven MA, Pleunis-van Empel MC, Koch H, Grol RP, Dinant GJ, van der Weijden T. Why do patients want to have their blood tested? A qualitative study of patient expectations in general practice. *BMC Fam Pract* 2006;7:75.
8. Van Boven K, Dijksterhuis P, Lamberts H. Defensive testing in Dutch family practice. Is the grass greener on the other side of the ocean? *J Fam Pract* 1997;44:468-72.
9. Patt CH, Yoo HY, Dibadj K, Flynn J, Thuluvath PJ. Prevalence of transaminase abnormalities in asymptomatic, healthy subjects participating in an executive health-screening program. *Dig Dis Sci* 2003;48:797-801.
10. Mold JW, Aspy CB, Lawler FH. Outcomes of an insurance company-sponsored multichannel chemistry screening initiative. *J Fam Pract* 1998;47:110-7.
11. Phillips WR, Thompson DJ. Multi-channel laboratory testing and the unexpected abnormal result: a statistical myth corrected. *N Z Med J* 1981;94:462-4.
12. Houben PHH, van der Weijden T, van Bokhoven MA, Droog A, Winkens RA, Grol R. Overwegingen van huisartsen bij het interpreteren van laboratoriumonderzoek; een kwalitatief onderzoek. (How general practitioners interpret laboratory test results; a qualitative interview study) English summary. *Huisarts Wet* 2005;48:326-32.
13. Brigden ML, Heathcote JC. Problems in interpreting laboratory tests. What do unexpected results mean? *Postgrad Med* 2000;107:145-6, 51-2, 55-8 passim.
14. Hubbell FA, Frye EB, Akin BV, Rucker L. Routine admission laboratory testing for general medical patients. *Med Care*. 1988;26:619-30.
15. Rutten GEHM, De Grauw WJC, Nijpels G, Goudswaard AN, Uitewaal PJM, Van der Does FEE, et al. NHG-standaard Diabetes Mellitus type II. *Huisarts Wet* 2006;49:137-52.
16. 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;45:21-5.
17. Franzini C. Relevance of analytical and biological variations to quality and interpretation of test results: examples of application to haematology. *Ann Ist Super Sanita* 1995;31:9-13.
18. De Ruijter W, Assendelft WJ, Macfarlane PW, Westendorp RG, Gussekloo J. The additional value of routine electrocardiograms in cardiovascular risk management of older people. *Scand J Prim Health Care* 2008;26:147-53.
19. van Wijk M, Mosseveld M, van der Lei J. Design of a decision support system for test ordering in general practice: choices and decisions to make. *Methods Inf Med* 1999;38:355-61.
20. Smith BJ, McNeely MD. The influence of an expert system for test ordering and interpretation on laboratory investigations. *Clin Chem* 1999;45:1168-75.
21. Bugter Maessen AM, Winkens RA, Grol RP, Knottnerus JA, Kester AD, Beusmans GH, et al. Factors predicting differences among general practitioners in test ordering behaviour and in the response to feedback on test requests. *Fam Pract* 1996;13:254-8.
22. Verstappen WH, van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, Grimshaw J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians. A randomized trial. *JAMA* 2003;289:2407-12.

23. Lyon AW, Greenway DC, Hindmarsh JT. A strategy to promote rational clinical chemistry test utilization. *Am J Clin Pathol.* 1995;103:718-24.
24. Winkens RA, Pop P, Bugter Maessen AM, Grol RP, Kester AD, Beusmans GH, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet.* 1995;345:498-502.
25. van Walraven C, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits (see comments). *JAMA.* 1998;280:550-8.
26. Zaat JO, Schellevis FG, van Eijk JT, van der Velden K. Do out-of-office laboratory tests affect diagnoses in general practice? *Scand J Prim Health Care.* 1995;13:46-51.
27. Henny J. Interpretation of laboratory results: the reference values, a necessary evil? *Clin Chem Lab Med.* 2007;45:939-41
28. Espeland A, Baerheim A. General practitioners' views on radiology reports of plain radiography for back pain. *Scand J Prim Health Care.* 2007;25:15-9.
29. Puhan MA, Steurer J, Bachmann LM, ter Riet G. A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Ann Intern Med.* 2005;143:184-9.
30. Gigerenzer G, Hoffrage U. How to improve bayesian reasoning without instruction: frequency formats. *Psychol Rev.* 1995;102:684-704.
31. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med.* 1998;104:374-80.

Chapter 5

**Pretest expectations strongly influence
interpretation of abnormal laboratory results and
further management**

Abstract

Background

Abnormal results of diagnostic laboratory tests can be difficult to interpret when disease probability is very low. Although most physicians generally do not use Bayesian calculations to interpret abnormal results, their estimates of pretest disease probability and reasons for ordering diagnostic tests may - in a more implicit manner - influence test interpretation and further management. A better understanding of this influence may help to improve test interpretation and management. Therefore, the objective of this study was to examine the influence of physicians' pretest disease probability estimates, and their reasons for ordering diagnostic tests, on test result interpretation, posttest probability estimates and further management.

Methods

Prospective study among 87 primary care physicians in the Netherlands who each ordered laboratory tests for 25 patients. They recorded their reasons for ordering the tests (to exclude or confirm disease or to reassure patients) and their pretest disease probability estimates. Upon receiving the results they recorded how they interpreted the tests, their posttest probability estimates and further management. Logistic regression was used to analyse whether the pretest probability and the reasons for ordering tests influenced the interpretation, the posttest probability estimates and the decisions on further management.

Results

The physicians ordered tests for diagnostic purposes for 1253 patients; 742 patients had an abnormal result (64%). Physicians' pretest probability estimates and their reasons for ordering diagnostic tests influenced test interpretation, posttest probability estimates and further management. Abnormal results of tests ordered for reasons of reassurance were significantly more likely to be interpreted as normal (65.8%) compared to tests ordered to confirm a diagnosis or exclude a disease (27.7% and 50.9%, respectively). The odds for abnormal results to be interpreted as normal were much lower when the physician estimated a high pretest disease probability, compared to a low pretest probability estimate (OR=0.18, 95% CI=0.07-0.52, $P<0.001$).

Conclusion

Interpretation and management of abnormal test results were strongly influenced by physicians' estimation of pretest disease probability and by the reason for ordering the test. By relating abnormal laboratory results to their pretest expectations, physicians may seek a balance between over- and under-reacting to laboratory test results.

Background

Laboratory tests are frequently ordered in routine primary care as part of the diagnostic process, even though the physician's pretest expectation may often be that the probability of disease is low, and they often order tests for other than purely medical reasons, such as patient reassurance¹⁻². As a consequence of the statistical definitions used for the reference values for laboratory tests, abnormal results are frequent, even in healthy individuals³. For example, in a screening programme with healthy individuals, a battery of 8 blood chemistry tests yielded at least one abnormal result for 20.6% of the individuals⁴. Abnormal results may therefore be difficult to interpret, certainly in the light of the low probability of serious disease in the primary care population⁵.

When physicians interpret test results and make plans for further management, it would be interesting to know to what extent they take their pretest expectations into account. For example, many physicians have difficulty performing Bayesian calculations to interpret test results, as such calculations may be complex and are often not easily applicable to situations where several diagnostic hypotheses are considered and several tests are ordered⁶. However, physicians' pretest expectations, such as their estimates of pretest probability and their reasons for ordering diagnostic tests (for example to exclude or confirm disease or to reassure patients) may influence test result interpretation and management, though perhaps not in a direct Bayesian fashion but probably in a more implicit way. Although the influence of physicians' pretest expectations on the ordering of tests has been extensively studied, little is known about the influence of their pretest expectations on the interpretation of test results and further management in routine care, as research on this subject has been scarce^{7,8}. A better understanding of this influence may help to improve test interpretation and management. The objective of this study was therefore to examine the influence of physicians' pretest expectations in terms of estimated pretest disease probability, and their reasons for ordering diagnostic tests, on the subsequent interpretation of the results of these diagnostic tests and further management.

Methods

Design and setting

We conducted a prospective study among primary care physicians and their patients in 7 rural, suburban and urban areas in the south of the Netherlands, in 2004/05. Each participating physician was instructed to record data on 25 adult patients for whom they had decided to order laboratory tests during the consultation. To prevent selection bias, they were instructed to include the first 25 patients for whom laboratory tests were ordered, without any further selection. Physicians working part-

time included a smaller number of patients, proportional to the number of hours a week they worked. Patients were asked to give informed consent. The Maastricht Medical Ethics Committee approved the study (reference number MEC 03-195-1).

Measurements

The physicians recorded data both when they ordered the laboratory tests and when they received the test results, using forms (see appendices) that were specifically designed for the study and took about 2 minutes to complete. The forms had been pilot-tested and evaluated as regards validity, reliability and user convenience in an iterative process among a sample of ten primary care physicians and a questionnaire expert.

Variables, pretest expectations

Reason for ordering tests

We distinguished nine reasons for test ordering, which were chosen on the basis of a qualitative interview study among primary care physicians about ordering and interpreting laboratory tests⁹. Physicians recorded the most important reason for ordering the investigations by ticking one of nine check-off boxes. We summarized these into five categories: (1) to exclude disease and reduce the physician's own uncertainty, (2) to confirm diagnosis and to determine treatment, (3) to reassure patients and at patients' request, (4) to screen for hypertension/cholesterol/diabetes and check-up for a known disorder and (5) other reasons.

Pretest estimate of disease probability

The form that the physicians had to complete asked: 'Do you suspect that the patient has a disease?'. The physicians answered on a 5-point Likert scale: 'definitely not', 'probably not', 'maybe', 'probably yes' and 'definitely yes'.

Variables, outcomes

Interpretation of the laboratory results

When the physicians received the results we asked them: 'How do you interpret these results for this patient?' The physicians answered on a 3-point scale; 'normal', 'possibly abnormal' or 'clearly abnormal'.

Posttest estimate of disease probability

The form that the physicians had to complete asked: 'Do you suspect that the patient has a disease?'. The physicians answered on a 5-point Likert scale: 'definitely not', 'probably not', 'maybe', 'probably yes' and 'definitely yes'.

Management

We distinguished nine management items, and physicians were instructed to select one or two items in check-off boxes. We classified these items into passive and active management items. Passive management items were 'reassurance/explanation', 'expectative/wait-and-see', 'advice (about lifestyle, complaints, etc.)', and 'instructions'. Active management items were 'additional investigations (laboratory, imaging, etc.)', 'new/follow-up appointment', 'medication (start, stop, change)', referral (specialist, other health care provider) and 'other management'. We defined the management as active if at least one of the checked boxes was an active management item.

Analysis

Only patients for whom the physician ordered laboratory tests for diagnostic purposes (reasons 1-3) were included in the analysis. We excluded patients for whom tests were ordered for screening, check-up or other reasons. We included all test results reported by the regional laboratories and defined a patient's results as abnormal if at least one test was outside the laboratories' reference values. We used chi-square tests to identify significant differences in interpretation, posttest probability estimates and management between the various reasons for ordering tests and between the various estimates of pretest probability.

Three logistic regression models were applied to the data (Table 5.4). The first model analysed the influence of patients' age and sex, the reason for ordering tests and the pretest probability on the interpretation of the results. The second model incorporated the previous variables, plus the interpretation of the results, as independent variables, and analysed their influence on the posttest probability estimates. Finally, the third model investigated the influence of all previous variables on the management. To be able to apply logistic regression, we dichotomized the dependent variables, distinguishing the categories 'normal' and 'possibly or clearly abnormal' for the interpretation of results, and the categories 'low probability' (definitely not/probably not) and 'high probability' (maybe/probably yes/definitely yes) for the posttest probability. We considered P-values smaller than or equal to 0.05 to be significant. We checked for multicollinearity (condition index >30 and variance decomposition proportion (VDP) >0.5) and tested the goodness-of-fit using the Hosmer and Lemeshow test. All analyses were performed with SPSS version 16.0.

Results

Eighty-seven primary care physicians participated, and together they included 1775 patients (Table 5.1). Laboratory tests were ordered for diagnostic reasons for 1253 (71%) patients. We received no laboratory results for 7.2% of these patients, the primary reason being that patients failed to visit the laboratory (29%). The laboratories reported 11,548 tests for the remaining 1,163 patients, a mean of 9.9 tests per patient. The most common reason for ordering tests was to exclude disease (62%). Tests for reassurance were ordered for 20% of the patients. The estimated pretest disease probability was low for 43% of the patients (Table 5.2). There were 742 patients (64%) with a laboratory result including one or more abnormal tests.

Table 5.1 Characteristics of primary care physicians and patients.

Primary care physicians		N=87 (NIVEL [*])
Sex	man	68% (67%)
	woman	32% (33%)
Age, years	<50	68% (55%)
	>50	32% (45%)
Experience, years [†]	<15	41%
	>15	59%
Working	full-time	52% (54%)
	part-time	48% (46%)
Patients		N=1253
Sex	man	38%
	woman	62%
Age, years	18-40	33%
	40-60	37%
	60+	30%

^{*} The Netherlands Institute for Health Services Research (www.nivel.nl) documents data on all Dutch primary care physicians (N=8408, data 2005); [†] No Nivel data available

Table 5.2 Reasons for ordering laboratory tests and pretest probability estimates.

Reason for ordering lab tests	N=1147 (16 missing)
reassure patient	226 (20%)
exclude disease	708 (62%)
confirm diagnosis	213 (19%)
Estimate of pretest disease probability	N=1138 (25 missing)
definitely no disease	114 (10%)
probably no disease	377 (33%)
maybe	329 (29%)
probably disease	252 (22%)
definitely disease	66 (6%)

The physicians interpreted the abnormal laboratory results for these 742 patients as normal in 48% of the cases, while their estimation of the posttest probability was low in 49.5% of the cases, and their management consisted of 'no action' for 49.2% of these patients. The percentage of patients whose abnormal results were interpreted as normal was significantly larger if tests were ordered to reassure (65.8%) compared with other reasons (50.9% and 27.7%, $P<0.001$) and was significantly larger if the pretest probability was estimated to be low (66.1%) compared with high pretest probabilities (19.6%, $P<0.001$). Similar significant relations were found for the posttest probability estimates and the management (Table 5.3). If tests were ordered for reassurance or if the physicians' pretest probability estimate was low, the interpretation for patients (comparable in terms of age and sex) having only normal results was 'normal' in 100% of the cases. The posttest probability estimates were low in 100% and 96.6% of the cases, respectively, while the management was 'no action' in 88.9% and 91.2% of the patients, respectively.

Table 5.3 Interpretation, posttest disease probability estimates and management after receiving abnormal laboratory results.

	Test interpretation = normal	Posttest probability = no disease	Management = no action
Reason for ordering lab tests			
reassure patient	75 (65.8%)*	90 (76.9%)*	85 (75.2%)*
exclude disease	220 (50.9%)	224 (51.1%)	214 (49.4%)
confirm diagnosis	41 (27.7%)	34 (23.0%)	43 (28.9%)
Pretest probability estimate			
definitely no disease	37 (66.1%)*	48 (84.2%)*	42 (76.4%)*
probably no disease	139 (67.1%)	150 (71.8%)	137 (66.8%)
maybe	95 (45.2%)	101 (47.6%)	91 (43.1%)
probably disease	55 (31.4%)	43 (24.0%)	58 (32.8%)
definitely disease	9 (19.6%)	6 (13.0%)	14 (29.8%)

* chi-square test, $P<0.001$

If tests were ordered for reassurance, the percentage of patients with abnormal results being offered further diagnostic investigations was 8.8%, while none of the patients whose laboratory results were normal were offered further investigations. Of the patients with a low pretest probability and abnormal results, 11.1% were offered further investigations by their physician, while 3.7% of the patients with a low pretest probability and normal results were offered further investigations.

Table 5.4 shows the results of the logistic regression analysis. There was no multicollinearity. Compared to a low pretest probability estimate, a high estimate decreased the likelihood that abnormal results were interpreted as normal (OR=0.18, 95% CI=0.07-0.52, $P<0.001$) and also decreased the likelihood of a low posttest probability estimate (OR=0.04, 95% CI=0.01-0.23, $P<0.001$). The physicians were also less likely to interpret abnormal tests as normal if the laboratory tests were ordered

to confirm a diagnosis, compared to those ordered to exclude disease (OR 0.59, CI 0.37-0.93, $P=0.067$). The likelihood of passive management ('no action') increased if tests were ordered for reassurance, compared to those ordered to exclude disease (OR 2.25, CI 1.08-4.66, $P=0.06$).

Table 5.4 Influence of pretest expectations on interpretation, posttest probability estimates and management after abnormal results.

N=742	Test interpretation = normal [†] odds ratio (95% CI)	Posttest probability = no disease [‡] odds ratio (95% CI)	Management = no action [§] odds ratio (95% CI)
Reason for ordering lab tests	$P=0.067$	$P=0.52$	$P=0.06$
exclude disease	1	1	1
reassure patient	1.1 (0.63-1.83)	1.46 (0.67-3.17)	2.25 (1.08-4.66)
confirm diagnosis	0.59 (0.37-0.93)	0.82 (0.42-1.59)	0.80 (0.46-1.40)
Pretest probability estimate	$P<0.001$	$P<0.001$	$P=0.19$
definitely no disease	1	1	1
probably no disease	1.13 (0.56-2.26)	0.35 (0.11-1.08)	1.67 (0.64-4.32)
maybe	0.49 (0.24-1.03)	0.20 (0.06-0.63)	1.11 (0.41-3.00)
probably disease	0.33 (0.15-0.73)	0.07 (0.02-0.23)	1.75 (0.59-5.17)
definitely disease	0.18 (0.07-0.52)	0.04 (0.01-0.23)	2.88 (0.79-10.59)
Test interpretation	N/A [*]	$P<0.001$	$P<0.001$
normal		1	1
possibly abnormal		0.09 (0.06-0.15)	0.20 (0.10-0.40)
abnormal		0.01 (0.005-0.02)	0.43 (0.25-0.72)
Posttest probability estimate	N/A [*]	N/A [*]	$P<0.001$
definitely no disease			1
probably no disease			0.48 (0.26-0.89)
maybe			0.14 (0.07-0.32)
probably disease			0.07 (0.03-0.16)
definitely disease			0.09 (0.04-0.23)

* N/A not applicable; [†] Goodness-of-fit: chi-square=5.29, $P=0.73$, [‡] chi-square=7.12, $P=0.52$, [§] chi-square=5.70, $P=0.68$

Discussion

The results show that the interpretation of test results, posttest disease probability estimates and management were significantly influenced by the physicians' pretest expectations. If the pretest probability was low or when tests were ordered at patients' request or to reassure them, the physicians tended to interpret abnormal results as normal and not to initiate further action. On the whole, this may be a correct decision, since many laboratory abnormalities will not be clinically relevant if the pretest probability is low. Physicians may use their pretest expectations to seek a balance between over- and under-reacting to laboratory test results.

To our knowledge, research about physicians' routine interpretation of laboratory results is still scarce. This study attempted to examine what they do with the results of laboratory tests. Strong points of this study were that it included many physicians and

patients, that the data were prospectively collected and that we tried to prevent selection bias by instructing the physicians to include consecutive patients for whom laboratory tests were ordered.

A disadvantage of our method was the heterogeneity in terms of laboratory tests, abnormal results and diagnoses. This means that interpretation and management cannot be related to a specific test, abnormality or diagnosis. In the context of this study, however, it would have been unrealistic to reduce clinical variation to a minimum and thus force the physicians into a standardized study, since the primary goal was to examine whether pretest expectations influence interpretation, posttest probability estimates and management in day-to-day care. It could be interesting in future studies to examine in more detail how specific tests influence further diagnosis and management. Another limitation is that the abnormal results in the group of patients with a low pretest probability may have been less abnormal than those for patients with a high pretest probability. Such differences in the level of abnormality of test results may have influenced physicians' interpretation of the test results and their further management. This influence was difficult to correct for in our analyses, as many different laboratory tests were ordered. Future studies may address more specifically the influence of the level of abnormality of test results on interpretation and further management.

Finally, as each physician included several patients, there may have been a certain clustering of specific interpretations and behaviour at the level of the physician. We have not analyzed this, as the focus of the study was to explore how pretest expectations influence the interpretation of results at the level of individual patients, and we did not intend to explore the differences in interpretation between physicians. We recruited a large group of practitioners (87) to ensure external generalizability of our findings.

Our study found that the pretest disease probability strongly influenced the physicians' interpretation of the laboratory results and their posttest probability estimates. This seems in line with Bayesian theory, which shows that the significance of a particular test result depends on pretest probability¹⁰. But it has also been pointed out that physicians are often not very proficient at the calculations that this theory requires, and they do not routinely use these calculations^{6,11}. There may be a gap between physicians' performance in terms of these calculations and the way they interpret laboratory results in routine care. This discrepancy might be addressed in future research.

Furthermore, the magnitude of abnormality of a test may be an important factor in the interpretation of results. When test results are dichotomized into normal and abnormal, important information may be lost. It may therefore be useful not to dichotomize test results, but to use, for example, likelihood ratios instead. These likelihood ratios may help physicians come to a more appropriate interpretation of test results¹², although it remains unclear if they are really helpful in routine practice^{13,14}.

Since many of the abnormal laboratory results hardly affected posttest probability estimates and management in our study, physicians should carefully consider if it was useful to order the tests in the first place. Also, the physicians ordered further investigations for nearly 10% of the patients for whom the original tests had been ordered for reassurance. It may be doubted if this was necessary, since it has been shown that investigations may also have negative consequences, such as an unjustified cascade of further investigations to explain unexpected abnormalities^{15,16}. In view of the number of patients in this study who were offered further investigations, future research should examine how often negative consequences of laboratory testing, such as cascade processes, occur.

Conclusions

Physicians' interpretation of laboratory results and further management after receiving the results of laboratory tests is clearly influenced by their pretest expectations. Physicians may use these expectations to seek a balance between over- and under-reacting to laboratory results. Our findings help to understand interpretation and use of laboratory results in day-to-day care. However, further research into the interpretation and use of laboratory results, including different levels of abnormality, is necessary and might in the future help to improve test interpretation and management.

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;45:21-5.
2. van der Weijden T, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002;52: 974-80.
3. Phillips WR, Thompson DJ. Multi-channel laboratory testing and the unexpected abnormal result: a statistical myth corrected. *N Z Med J* 1981;94:462-4.
4. Mold JW, Aspy CB, Lawler FH. Outcomes of an insurance company-sponsored multichannel chemistry screening initiative. *J Fam Pract* 1998;47:110-7.
5. Brigden ML, Heathcote JC. Problems in interpreting laboratory tests. What do unexpected results mean? *Postgrad Med* 2000;107:145-6, 151-142, 155-148 passim.
6. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104:374-80.
7. Verstappen WH, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract* 2004;21:387-95.
8. Holtgrave DR, Lawler F, Spann SJ. Physicians' risk attitudes, laboratory usage, and referral decisions: the case of an academic family practice center. *Med Decis Making* 1991;11:125-30.
9. Houben PHH, van der Weijden T, van Bokhoven MA, Droog A, Winkens RA, Grol R. Overwegingen van huisartsen bij het interpreteren van laboratoriumonderzoek; een kwalitatief onderzoek. *Huisarts Wet* 2005;48:326-32.
10. Gigerenzer G, Hoffrage U. How to improve bayesian reasoning without instruction: frequency formats. *Psychol Rev* 1995;102:684-704.
11. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002;324:824-6.
12. Brown MD, Reeves MJ. Evidence-based emergency medicine/skills for evidence-based emergency care. Interval likelihood ratios: another advantage for the evidence-based diagnostician. *Ann Emerg Med*. 2003;42:292-7.
13. Puhan MA, Steurer J, Bachmann LM, ter Riet G. A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Ann Intern Med* 2005;143:184-9.
14. Sonis J. How to use and interpret interval likelihood ratios. *Fam Med*. 1999; 31:432-7.
15. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health* 2002;23:23-44.
16. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician* 1996;25:971-3.

Chapter 6

Reasons for ordering spinal X-ray investigations: how they influence general practitioners' management

PHH Houben, T van der Weijden, J Sijbrandij, RPTM Grol, RAG Winkens

Can Fam Physician 2006;52:1266-7

Abstract

Objective

To explore whether general practitioners (GPs) who request spinal X-ray investigations are less likely to refer, prescribe medication, or order additional tests when they intend only to reassure patients rather than to exclude or confirm disease.

Design

Cross-sectional survey. Physicians recorded reasons for requesting X-ray examinations and differential diagnoses on the request form. When they received a radiologist's report on a spinal X-ray test, GPs also received a questionnaire on their subsequent management.

Setting

Maastricht region of the Netherlands.

Participants

All 90 GPs in the Maastricht region who referred patients for spinal X-ray examinations during 19 consecutive weeks.

Main outcome measures

Reason for requesting spinal X-ray investigations, results, and subsequent management.

Results

We received completed questionnaires on 386 patient contacts. Physicians were significantly less likely to refer, prescribe medication, or order additional tests when they intended only to reassure patients rather than to exclude or confirm disease (odds ratio 0.3, 95% confidence interval 0.1 to 0.6). This effect was independent of the results of the X-ray examinations.

Conclusion

Physicians seem to be aware that they have requested spinal X-ray examinations only to reassure patients, and they account for this in their management. These findings support debate on the negative consequences of requesting tests for other than purely medical reasons. Concerns about unjustified actions based on the results of X-ray examinations that were requested primarily to reassure patients should probably have only a minor role in the debate about unnecessary testing.

Introduction

Imaging of the cervical, thoracic, or lumbar spine is frequently ordered in general practice¹. Despite evidence-based guidelines on the usefulness of spinal X-ray films for diagnosis or therapy, general practitioners (GPs) continue to request them contrary to recommendations¹⁻⁴. Up to 50% of requests for lumbar spine radiography could be unnecessary according to clinical guidelines⁵⁻⁷. Several studies showed that GPs have many reasons for these requests; reassuring patients is apparently particularly important grounds for such requests^{8,9}.

It is important to know the effect of requesting spinal X-ray films on GPs' management. If they request the tests, their management might be guided mainly by the results. This is important because experts on medical decision making warn that, in interpreting test results, it is essential to look not only at the results themselves but also at GPs' estimate of disease probability¹⁰. If disease probability is not included in interpreting results, the importance of abnormal findings could be overestimated¹¹. There have been case reports of GPs' requesting laboratory investigations to reassure patients and receiving unexpected abnormal results that led to unjustified additional tests to search for the causes of the abnormalities^{12,13}. When requesting spinal X-ray films in daily practice, the very fact that GPs intend only to reassure might increase their reluctance to prescribe medication or refer patients, regardless of the results. Such reluctance might be caused by lack of interest in the results or by a perceived incongruence between clinical symptoms and X-ray findings¹⁴.

Although the intention to reassure patients might be expected to affect GPs' management, there is hardly any evidence supporting such an expectation. This study tested the hypothesis that, when they request spinal X-ray films, GPs are less likely to pursue active management (referral to specialists, prescribing medication, physical therapy, or additional investigations) when they intend only to reassure patients rather than to exclude or confirm disease.

Methods

Study design

We conducted a cross-sectional survey of all 90 GPs in the region around the Dutch town of Maastricht. The GP population in this region does not differ substantially from the overall population of GPs in the Netherlands¹⁵. In this region, GPs order tests without referring patients to outpatient clinics and without having to consult specialists. All requests are handled by the Diagnostic Centre at the Maastricht University Hospital. X-ray films are seen and interpreted by hospital radiologists, and reports are sent to the GPs who requested the examinations.

Data collection

To ensure that data collection would be feasible and not put undue strain on GPs, we chose one specific, infrequently ordered test: X-ray examination of the lumbar, thoracic, or cervical parts of the spine. All requests for such tests ordered during 19 consecutive weeks were selected. Physicians were already in the habit of recording the reason for their requests and the differential diagnoses on the request forms. When results were returned to GPs, a 1-page inquiry was attached. It contained the question, "What was your management after you received the result of the X-ray test?" Physicians were asked to tick off any management maneuvers they had used and to add comments. The questionnaire had been pilot-tested in 2 practices. Telephone calls were used to remind GPs to complete and return the forms. Maastricht University's Medical Ethics Review Committee approved the study.

Variables

Physicians' decision to pursue active management was the dependent variable. Active management was defined as starting or changing medication, repeat testing, offering physical therapy, or consulting a specialist. Physical therapy included physiotherapy, exercise therapy, and manual therapy. The reason for the request for spinal X-ray films was the independent variable. It was classified into 5 categories: to exclude disease, to confirm disease, for follow-up, to reassure a patient, and requested by a health professional other than GP. The category "to reassure a patient" included physicians' requests and X-ray films ordered at patients' request.

We regarded the results of X-ray examinations as a confounder and corrected for their influence on management in the analysis. Results were classified into 3 categories by 3 of the authors using radiologists' reports: normal (no abnormalities mentioned), minor degenerative changes (minor degenerative changes and minor scoliosis or kyphosis), and abnormal (eg, appreciable degenerative changes, fractures, or osteoporosis) (Table 6.1). We also regarded the differential diagnosis as a confounder. It was coded into 2 categories: less serious and more serious. Coding was done by 2 of the authors. Patients' age and sex were also considered as potential confounders in the analysis.

Analysis

Analysis was done using SPSS 11.0 for Windows. Descriptive cross-tabulations were used to analyze the influence of the reason for the request and the test result on decisions about management. Chi-square tests were performed to calculate significant differences. Variables were checked for multicollinearity. Multiple logistic regression analysis, using the enter method, was used to compare the categories "to confirm disease" and "to reassure a patient" from the variable "reason for request" with the category "to exclude disease". The categories "for follow-up" and "request

by a health professional other than GP” were omitted from the analysis (25 cases). Cases with missing data were included in the analysis. The product of the variables “reason for request” and “result” was included in the initial model to test for effect modification. It turned out not to be significant and was omitted from the final analysis.

Table 6.1 Differential diagnoses mentioned on request forms.

Diagnoses	1 st diagnosis	2 nd diagnosis	3 rd diagnosis	Classification*
Non-specific [†]	126	12	2	Not very serious
Degenerative changes	108	24	1	Not very serious
Pain	31	3	0	More serious
Fracture	26	7	0	More serious
Osteoporosis	16	15	2	More serious
No diff. diagnosis mentioned	16	0	0	Missing
Other	63	26	6	More serious
Total	386	87	11	

* If at least 1 of 3 differential diagnoses was a more serious diagnosis, the differential diagnosis variable was coded as more serious; [†] The GP mentioned no differential diagnosis but asked for abnormalities (eg, pathology, bone abnormalities, musculoskeletal disease); [‡] All other diagnoses (eg, spondylolisthesis or radicular syndrome).

Results

All 90 GPs associated with the Diagnostic Centre requested X-ray tests during the study period. Among the 98% of questionnaires (N=386) returned, 58% involved female patients, 22% involved patients older than 65 years, and 33% involved patients younger than 40 years. The most frequently missing item among the responses was the reason for requesting X-ray investigation (Table 6.2). In descending order, 63% of X-ray tests were requested to exclude disease, 15% to reassure a patient, and 14% to confirm disease. More than 70% of the differential diagnoses mentioned on the request forms concerned disorders that were not serious. As to the results, 137 (36%) were normal, 185 (48%) showed minor degenerative changes, and 64 (17%) were abnormal.

Table 6.2 Reasons for requesting spinal X-ray films.

Reason for request*	Percentage stating the reason
To exclude	63.2%
To reassure	14.9%
To confirm	14.0%
Other health professional’s request	4.4%
For follow-up	3.5%

* N=315, 71 responses missing

Table 6.3 shows the relation between reasons for requests and management by GPs. X-ray investigations requested only to reassure patients led to active management of 11 patients (23%). This was fewer than the corresponding number of patients actively managed when X-ray tests were requested to exclude (109, 55%) or to confirm (24, 55%) disease ($P<0.001$).

Table 6.3 Reasons for requesting an X-ray and management, $P<0.001$, 2 degrees of freedom.

Reason for request	Active management [*]	No active management	Physical therapy [†]	Consulting specialist [†]	Starting or changing medication [†]	Repeated testing [†]
To exclude (n=199)	109 (55%)	90 (45%)	63 (32%)	32 (16%)	14 (7%)	7 (4%)
To confirm (n=44)	24 (55%)	20 (45%)	15 (34%)	5 (11%)	5 (11%)	0 (0%)
To reassure (n=47)	11 (23%)	36 (77%)	6 (13%)	3 (6%)	1 (2%)	2 (4%)

^{*} Active management includes physical therapy, consulting specialists, starting or changing medication and repeat testing. Because GPs could choose more than 1 item per patient, the sum of these items might exceed the number of patients actively managed; [†] Not tested for significance.

Results were regarded as a confounder in this study, but they also influenced management (Table 6.4). Active management was more likely to be pursued ($P<0.001$) when results showed minor degenerative changes (52%) or were abnormal (63%) than when they were normal (34%). Results showing minor degenerative changes were more likely to lead to referrals for physical therapy than abnormal or normal results were.

Table 6.4 Test results and management decisions, $P<0.001$, 2 degrees of freedom.

Results of X-ray tests	Active management [*]	No active management	Physical therapy [†]	Consulting specialist [†]	Starting or changing medication [†]	Repeated testing [†]
Normal (n=137)	47 (34%)	90 (66%)	32 (23%)	12 (9%)	3 (2%)	1 (1%)
Minor degenerative changes (n=185)	96 (52%)	89 (48%)	66 (36%)	22 (12%)	5 (3%)	7 (4%)
Abnormal (n=64)	40 (63%)	24 (37%)	12 (19%)	16 (25%)	15 (23%)	2 (3%)

^{*} Active management includes physical therapy, consulting specialists, starting or changing medication and repeat testing. Because GPs could choose more than 1 item per patient, the sum of these items might exceed the number of patients actively managed; [†] Not tested for significance.

Table 6.5 shows the results of multiple logistic regression analysis. The GPs were significantly less likely to pursue active management (odds ratio (OR) 0.3, 95% confidence interval (CI) 0.1 to 0.6) when they had intended to reassure a patient than when they had intended to exclude disease. Requests intended to confirm disease had the same influence as those intended to exclude disease. This effect on management was independent of X-ray results. Results were the only confounder that also had an independent influence on management. When results showed minor degenerative changes, active management was 2.4 times more likely (95% CI 1.3 to 4.6) than when results were normal. Active management was 5.9 times more likely (95% CI 2.3-15.0) when results were abnormal.

Table 6.5 Multiple logistic regression analysis.

Variable*	Adjusted odds ratio for active management	95 % confidence interval
Reason for request		
to exclude disease	1	
to confirm diagnosis	0.8	0.4-1.7
to reassure a patient	0.3 [†]	0.1-0.6
X-ray result		
normal	1	
minor degenerative changes	2.4 [†]	1.3-4.6
abnormal	5.9 [‡]	2.3-15.0
Differential diagnosis		
more serious vs. less serious	1.2	0.7-2.1
Age, years		
<40	1	
40-65	0.6	0.3-1.2
>65	0.6	0.2-1.4
Sex		
female vs. male	0.9	0.5-1.5

* n=281; 80 cases were missing, and 25 excluded; [†] P<0.01, [‡] P<0.001

Discussion

Summary of main findings

Results of this study confirmed our hypothesis: when spinal X-ray investigations were requested with the intention of reassuring patients, the adjusted OR (i.e., irrespective of the result) for pursuing active management was 0.3 compared with when tests were ordered to exclude disease. Apparently then, GPs' decisions to pursue active management are influenced by the reasons for requests.

Strengths and limitations of the study

As far as we know, this is the first study to determine the influence on GPs' routine management of ordering spinal X-ray investigations with the intention of reassuring

patients. The strength of this study is that it provides insight into the way in which requests that are probably not in accordance with evidence-based guidelines influence management. It appears that, even when results are abnormal, GPs are less likely to pursue active management when the intent was to reassure patients. We think, therefore, that the results of this study are valuable for those seeking to improve the requesting of spinal X-ray tests.

The limitations of cross-sectional designs are generally known. For our results, these limitations mean that, although we found a relation between the reasons for requesting the tests and subsequent management, we cannot prove a causative relationship.

The response rate was surprisingly high. This might be explained by the existing close cooperation between the GPs and the Diagnostic Centre. Physicians in Maastricht have attended several training sessions on test ordering organized by the Diagnostic Centre and are accustomed to speak openly with colleagues about requesting diagnostic tests.

We recognize that other variables influence complex decision processes like management choice, such as patients' symptoms, pain, or worries. Because assessing such symptoms would have taken a great deal of time and effort from participating GPs, we chose to focus on the few variables that we regarded as key determinants of management. Future studies could explore the role of other variables that might confound the relation between reasons for requests and management.

Comparison with other studies

Results of X-ray tests were in accordance with those found in other primary care studies. In a randomized controlled trial, results of lumbar spine X-ray tests showed no abnormalities in 31% of patients¹⁶. In another study, 37% of radiologic findings on lumbar X-ray films were normal, and 36% showed minor changes⁵. These results are roughly comparable with our results.

A study in a Dutch population reported that X-ray investigations were ordered at patients' request or for reassurance in 38% of cases¹⁷. Little and colleagues⁶ found that 28% of all requests were made for psychosocial reasons. We found only 14.9% of requests were made to reassure. We recorded only the primary reason, however, while the other studies recorded secondary reasons also. Reassurance is often a secondary reason, and our figure of almost 15% would probably have been higher if secondary reasons had been taken into account.

Implications for future research and clinical practice

Many requests for spinal X-ray films are intended to reassure patients, and there could be concerns about unjustified active management after results are received. Our study found that GPs take their reason for requesting X-ray films into account when deciding on management. Therefore, concerns about unjustified active

management should probably play only a minor role in the debate about requesting unnecessary tests.

Nevertheless, there are still many reasons to encourage GPs to adhere closely to national guidelines on requesting spinal X-ray investigations. For example, patients with low back pain who undergo X-ray tests for reassurance are more likely to consult their GPs in the subsequent months¹⁶.

Conclusion

Up to 50% of all requests for lumbar spine radiography could be unnecessary according to clinical guidelines. While reassuring patients is an important motive for requesting spinal X-ray tests, we did not know how such requests influenced further management. This study shows that, when GPs request spinal radiography to reassure patients, active management, such as prescribing medication and referral to specialists or physical therapy, is about 3 times less likely ($P<0.01$) than when they request radiography to confirm or exclude disease. This effect was found regardless of results.

Apparently, GPs take their reasons for requesting radiography into account when deciding on active management. This study increases our insight into the consequences of requesting tests for other than purely medical reasons and thus provides valuable information to those involved in improving physicians' use of radiographic services.

References

1. Oakeshott P, Kerry SM, Williams JE. Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br J Gen Pract* 1994;44:197-200.
2. Royal College of Radiologists Working Party. Influence of Royal College of Radiologists' guidelines on referral from general practice. *BMJ* 1993;306:110-1.
3. De Vos Meiring P, Wells IP. The effect of radiology guidelines for general practitioners in Plymouth. *Clin Radiol* 1990;42:327-9.
4. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;342:1317-22.
5. Halpin SF, Yeoman L, Dundas DD. Radiographic examination of the lumbar spine in a community hospital: an audit of current practice. *BMJ* 1991;303:813-5.
6. 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;15:264-5.
7. Van Boven K, Dijksterhuis P, Lamberts H. Defensive testing in Dutch family practice. Is the grass greener on the other side of the ocean? *J Fam Pract* 1997;44:468-72.
8. Espeland A, Baerheim A, Albrektsen G, Korsbrekke K, Larsen JL. Patients' views on importance and usefulness of plain radiography for low back pain. *Spine* 2001;26:1356-63.
9. Van der Weijden T, Van Bokhoven MA, Dinant GJ, Van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract* 2002;52:974-80.
10. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med* 1998;73:538-40.
11. Heller RF, Sandars JE, Patterson L, McElduff P. GPs' and physicians' interpretation of risks, benefits and diagnostic test results. *Fam Pract* 2004;21:155-9.
12. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician* 1996;25:971-5.
13. Brigden ML, Heathcote JC. Problems in interpreting laboratory tests. What do unexpected results mean? *Postgrad Med* 2000;107:145-6,151-2,155-8,161-2.
14. Symmons DP, Van Hemert AM, Vandenbroucke JP, Valkenburg HA. A longitudinal study of back pain and radiological changes in the lumbar spines of middle aged women. II. Radiographic findings. *Ann Rheum Dis* 1991;50:162-6.
15. Bugter-Maessen AM, Winkens RA, Grol RP, Knottnerus JA, Kester AD, Beusmans GH, Pop P. Factors predicting differences among general practitioners in test ordering behaviour and in the response to feedback on test requests. *Fam Pract* 1996;13:254-8.
16. Kendrick D, Fielding K, Bentley E, Kerslake R, Miller P, Pringle M. Radiography of the lumbar spine in primary care patients with low back pain: randomized controlled trial. *BMJ* 2001;322:400-5.
17. Dijksterhuis PH, Van Boven C. De schatbare waarde van aanvullend onderzoek in de huisartspraktijk. Lelystad, NL: Meditekst; 1993.

Chapter 7

**Cascade effects of laboratory testing are found to be
rare in low disease probability situations:
prospective cohort study**

PHH Houben, T van der Weijden, RAG Winkens, RPTM Grol

J Clin Epidemiol. 2010;63:452-8

Abstract

Objectives

(1) To investigate the frequency of cascades of further diagnostic investigations and referrals after abnormal laboratory results in situations of low disease probability; (2) to investigate pretest and posttest determinants; and (3) to describe the cascades that occur.

Study Design and Setting

Prospective cohort study in primary care in the Netherlands. Numbers of investigations/referrals were recorded during 6 months of follow-up for 256 patients with normal and abnormal laboratory results. The influences of the reason for ordering tests, interpretation of results and pretest/posttest disease probability were examined.

Results

After receiving the laboratory results, the physicians ordered further investigations for 22 (17.3%) patients with abnormal results, and for 2 (1.6%) patients with normal results ($P<0.001$). They referred 12 (9.4%) patients with abnormal results and 8 (6.2%) patients with normal results ($P=0.33$). Six patients had 2 investigations and/or referrals, and 1 patient had 3 referrals. There were significantly more investigations/referrals for results interpreted as abnormal ($P=0.004$) and for cases with a high posttest disease probability ($P=0.001$).

Conclusion

This study suggests that cascade processes after laboratory testing in situations of low disease probability are limited in magnitude and frequency. Improving interpretations may help improve the appropriateness of further investigations and referrals.

Introduction

Laboratory tests are frequently ordered in routine primary care as part of the diagnostic process¹. As a consequence of the statistical definition used for the reference values for laboratory tests, there are often abnormal results, even in healthy individuals². For example, in a screening program with healthy individuals, a battery of 8 blood chemistry tests showed at least one abnormal result for 20.6% of the individuals³. Abnormal results may therefore be difficult to interpret, certainly in the light of the low probability of serious disease in the primary care population⁴. Sometimes, abnormal results may lead to a process where referrals to specialists or an increasing number of diagnostic investigations are initiated because of uncertainty or to explain the initial abnormality, a process sometimes referred to in the literature as the 'cascade effect'⁵⁻⁷. Some authors argue that such cascade processes may have adverse effects, such as increased patient anxiety, risks for patients if they are exposed to invasive test procedures, or health costs due to unnecessary tests or procedures⁷. Some authors therefore recommend reticence in using diagnostic testing in patients with low pretest probabilities, to prevent such cascade processes^{8,9}. However, research in this area has been scarce and there is a lack of information on both the actual frequency and the determinants of such cascades. Better understanding may also help the discussion on optimal test ordering and test interpretation.

We therefore investigated whether, in situations of low pretest disease probability, a cascade process is more likely to occur after abnormal laboratory results than after normal results. We also studied whether pretest determinants (reason for ordering tests, diagnosis) or posttest determinants (interpretation of tests, estimates of disease probability, diagnosis) increase the risk of such cascade processes, and we qualitatively describe the cascades that occurred.

Methods

Design and setting

We used a subset of data from a prospective cohort study on laboratory test interpretation and management. Primary care physicians for this cohort study were selected in 2004/05, with the assistance of 7 laboratories in rural, suburban and urban areas in the south of the Netherlands. The 664 primary care physicians in these areas were invited to participate by postal mail. Physicians who did not respond were contacted by phone. Based on our calculation of the number of GPs needed for the prospective cohort study (whose results are presented elsewhere), further inclusion of physicians was stopped when 93 physicians had agreed to participate. Eventually,

6 physicians did not complete the data collection, so 87 primary care physicians participated.

Each participating physician was instructed to record data on 25 adult patients for whom they had decided to order laboratory tests during the consultation. To prevent selection bias, they were instructed to include the first 25 patients for whom laboratory tests were ordered, without any further selection. Physicians working part-time included a proportion of the patients, corresponding to their working week. The Medical Ethics Committee of Maastricht University approved the study. There was no external funding. Patients were asked to give informed consent.

We selected those patients from the cohort who met 3 criteria: (1) the physician had estimated a low pretest disease probability, (2) the physician had ordered the tests for diagnostic reasons (to exclude disease, to confirm diagnosis or to reassure the patient), and (3) the patient had had one or more abnormal test results (group 1). Each patient in this group was matched in terms of age (in 10-year categories) and sex with a patient also meeting criteria 1 and 2, but with normal test results only (group 2). Then, to keep the workload manageable for the participating physicians, we asked for follow-up information for a maximum of 5 patients per physician, selected at random using a random number generator.

Measurements

The physicians recorded data both when they ordered the laboratory tests and when they received the test results, using forms that were specifically designed for the study and took about 2 minutes to complete. The forms had been pilot-tested and evaluated as regards validity, reliability and user convenience in an iterative process among a sample of ten primary care physicians and a questionnaire expert. After six months, the participating physicians were asked to extract follow-up data from the electronic medical records in which all patients contacts were documented.

Outcome variable

Follow-up investigations/referrals

The physicians recorded if they had ordered investigations (such as laboratory tests, imaging or function tests) and/or if they had referred the patient to a specialist, both when they received the results and at each contact during the 6 months of follow-up. The analysis included only those tests and referrals that related to the diagnoses for which the physicians had initially ordered the laboratory tests, and ignored referrals or investigations for new problems emerging during the follow-up period.

Other variables

Laboratory results

We included all test results reported by the regional laboratories. We defined a patient's results as abnormal if at least one test was outside the laboratories' reference values.

Pretest and posttest estimates of disease probability

At the time of test ordering and when the physicians received the results, we asked: 'Do you suspect a disease?'. The physicians answered on a 5-point Likert scale: 'certainly not', 'probably not', 'maybe', 'probably yes' and 'certainly yes'. We dichotomized this into 2 categories, defined as the physician estimating the probability of disease as low (certainly not/probably not) or high (maybe/probably yes/certainly yes). We selected only patients with a low pretest estimate of disease probability.

Reason for ordering tests

We distinguished nine reasons for test ordering, which were chosen on the basis of a qualitative interview study with primary care physicians on ordering and interpreting laboratory tests¹⁰. Physicians recorded the most important reason for ordering the investigations by ticking boxes: (a) to exclude disease, (b) to confirm diagnosis, (c) at patient's request, (d) to reassure patient, (e) physician's uncertainty, (f) to determine treatment, (g) check-up for known disorder, (h) screening for hypertension/diabetes/cholesterol, and (i) other reasons. To keep the analysis manageable, we summarized these into 5 categories: (1) to exclude disease (a and e), (2) to confirm diagnosis (b and f), (3) to reassure patients (c and d), (4) screening (g and h), and (5) other reasons (i). We selected only patients with reason 1, 2 or 3.

Interpretation of the results

When the physicians received the results we asked them: 'How do you interpret these results for this patient?' The physicians answered on a 3-point scale; 'normal', 'possibly abnormal' or 'clearly abnormal'.

Pretest and posttest diagnosis

Physicians recorded their most likely diagnosis at the time of test ordering and when they received the laboratory results. We coded these diagnoses according to the International Classification of Primary Care (ICPC), defining two groups: (1) The physicians' most likely diagnosis refers to non-somatic disease, i.e. ICPC chapters P (psychological), Z (social) and ICPC code A97 ('no disease'); and (2) all other diagnoses.

Analysis

We counted how often physicians ordered investigations and/or referred to specialist care in both groups, and used the chi-square test to analyze the differences. For the patients with abnormal results, we used the chi-square test to analyze the relation between pretest and posttest considerations (i.e. reason for ordering tests, pretest diagnosis, posttest disease probability, posttest diagnosis, interpretation of the result) on the one hand and the numbers of follow-up investigations/referrals on the other. We also qualitatively described the cases in which cascade processes had occurred.

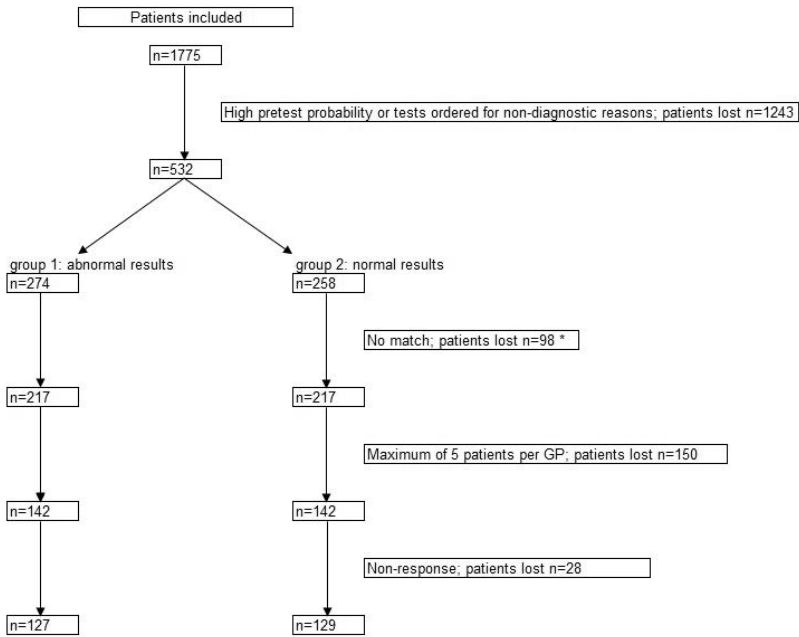
Results

Physicians and patients

Eighty-seven physicians participated in the cohort study. Eight physicians did not provide data at the 6-months follow-up because 1 physician had retired, 3 physicians had no eligible patients and 4 physicians had included their patients anonymously. From the patients included by the remaining 79 physicians, we were able to select 142 patients with an abnormal result who fulfilled our inclusion criteria (group 1), and matched them with 142 patients with a normal result (group 2). Nine physicians (28 patients) did not complete the follow-up forms, so in the end we had complete follow-up forms for 127 patients with an abnormal result and 129 patients with a normal result (Figure 7.1). Physician characteristics are described in Table 7.1. The groups were comparable in terms of patients' age, sex and physicians' pretest considerations (Table 7.2).

Referrals and investigations during the 6-months of follow-up

The physicians ordered investigations for 22 (17.3%) patients with abnormal results, compared to only 2 (1.6%) patients with normal results ($P < 0.001$; Table 7.3). The physicians referred 12 (9.4%) patients with abnormal results to specialist care, compared to 8 (6.2%) patients with normal results ($P = 0.33$). During the follow-up period, physicians ordered more than one investigation or referral only for a few patients with abnormal results: 6 (4.7%) patients had 2 follow-up investigations/referrals, and 1 (0.8%) patient in group one was referred 3 times. The physicians ordered no more investigations or referrals for patients with normal laboratory results.



* A relatively rare number of patients could not be matched (n=98), as patients of older age were overrepresented in group 1, and patients of younger age were overrepresented in group 2.

Figure 7.1 Inclusion flowchart.

Table 7.1 Physicians' characteristics.

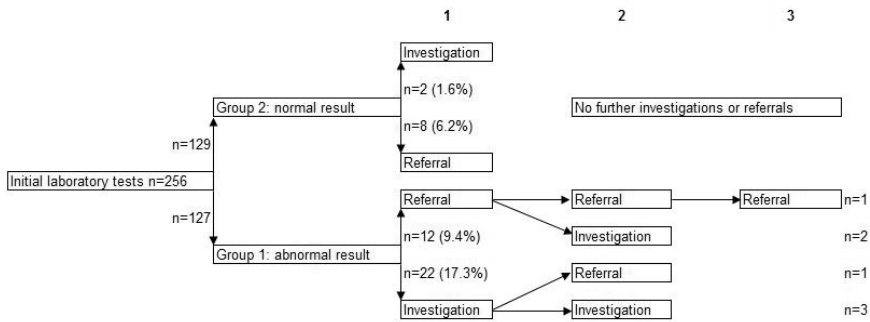
Physicians n=70		Percentage (NIVEL 2005 [*])
Sex	man	73%(67%)
	woman	27%(33%)
Age, years	<40	16%(21%)
	40-55	71%(58%)
	>55	13%(21%)
Experience, years	<10	16% [†]
	10-20	39%
	>20	46%
Working	full-time	54%(54%)
	part-time	46%(46%)

^{*} The Netherlands institute for health services research (NIVEL: www.nivel.nl) documents data on Dutch primary care physicians (N=8408); [†] No NIVEL data available.

Table 7.2 Patient characteristics and pretest considerations.

	Group 1: abnormal results n=127	Group 2: normal results n=129	Chi-square
Mean age	44.5 yr (SD: 14.8)	44.4 yr (SD: 15.5)	
Sex			
man	32%	33%	
woman	68%	67%	
Reason for ordering tests			
exclude disease	54%	51%	<i>P</i> =0.60
confirm diagnosis	2.4%	4.7%	
reassure patient	44%	44%	
Pretest ICPC diagnosis			
A97, P, Z	33%	33%	<i>P</i> =0.89
Other ICPC codes	67%	68%	

Table 7.3 Referrals and investigations during 6 months of follow-up.



Determinants of referrals/investigations for patients with abnormal results

If the physicians had ordered the laboratory tests with the intention to reassure the patient, they ordered fewer follow-up investigations/referrals, although this relation was not significant (*P*=0.25; Table 7.4). Neither the pretest nor the posttest diagnosis was of significant influence on the numbers of follow-up investigations/referrals. The physicians interpreted 67% of the abnormal results in group 1 as normal and they estimated a low posttest probability for 74% of the patients with abnormal results. They ordered significantly fewer follow-up investigations/referrals if they interpreted the results as normal (*P*=0.004) and if they estimated a low posttest probability (*P*=0.001).

Table 7.4 Pretest and posttest determinants of follow-up investigations/referrals for patients with abnormal results.

	Percent of referrals/ investigations (no.)	Percent with 1 referral/ investigation (no.)	Percent with >1 referrals/ investigations (no.)	Total	Chi-square
Pretest determinants					
Reason for ordering tests					
exclude disease	48% (45)	63% (17)	86% (6)	54% (68)	$P=0.25$
confirm diagnosis	2.2% (2)	3.7% (1)	0% (0)	2.4% (3)	
reassure patient	50% (46)	33% (9)	14% (1)	44% (56)	
Pretest ICPC diagnosis					
A97, P, Z	33% (30)	35% (9)	29% (2)	33% (41)	$P=0.96$
other ICPC codes	67% (60)	65% (17)	71% (5)	67% (82)	
Posttest determinants					
Interpretation of the results					
normal	77% (69)	44% (12)	29% (2)	67% (83)	$P=0.004$
doubtful	11% (10)	30% (8)	29% (2)	16% (20)	
abnormal	12% (11)	26% (7)	43% (3)	17% (21)	
Posttest estimate of disease probability					
low	82% (75)	59% (16)	29% (2)	74% (93)	$P=0.001$
high	18% (16)	41% (11)	71% (5)	26% (32)	
Posttest ICPC diagnosis					
A97, P, Z	42% (37)	33% (9)	29% (2)	39% (48)	$P=0.60$
other ICPC codes	58% (51)	67% (18)	71% (5)	61% (74)	

Qualitative description of the cascades

Within the group of seven patients for whom the physician ordered investigations or referrals more than once, two subgroups could be distinguished on basis of their laboratory results (Table 7.5). Subgroup A included 3 (2.4%) patients who had significant abnormal results. For these patients the physician immediately ordered more investigations or referrals (cases 1-3). Subgroup B included 4 (3.1%) patients who had marginally abnormal results. One patient had additional investigations and a referral several months after the initial consultation (case 4). One patient wanted to be referred to a rheumatologist (case 5). One patient had further investigations for the initial complaints (ECG) and had an ultrasound for an abnormal ALAT finding (case 6). The final patient in subgroup B was thought to suffer from hyperventilation, and had an elevated ESR (28 mm, reference 0-8). This patient was referred to three specialists, but no abnormalities were found (case 7).

Table 7.5 Patients with more than one investigations/referrals.

Subgroup A: clearly abnormal laboratory results

Case 1: #3001, male 72; initial diagnosis was 'thoracic pain/infection lungs', patient was known with decreased kidney function, and the physician ordered lab tests to exclude disease. Abnormal results were: ESR 100 mm (1-12), creatinine 1.57 mg/dl (0.62-1.39), Hb 12.4 g/dl (13.7 -17.7), blood was slightly hemolytic and macrocytic. Physician interpreted results as abnormal and estimated high disease probability. Her diagnosis became 'lung disease' and she ordered more investigations. Patient was treated with antibiotics and physician later ordered lab tests for check-up. Results had improved.

Case 2: #8106, female 39; initial diagnosis was 'hyperventilation', patient was known with hypothyroidism, and the physician ordered lab tests and ECG to exclude disease. Abnormal results were: creatinine 1.09 mg/dl (0.62-1.07), TSH 17.68 mE/l (0.4-5.5). ECG was also abnormal. Physician interpreted lab results as normal and estimated high disease probability. His diagnosis became 'hyperventilation/stress and poorly managed hypothyroidism'. Patient was immediately referred to a cardiologist (who found no abnormalities). Five months later the patient was referred to an endocrinologist.

Case 3: #1207, female 72; initial diagnosis was 'weakness due to viral infection', patient was known with thalassemia, and the physician ordered lab tests to exclude disease. Abnormal results were: ALAT 59 U/l (<40), ESR 42 mm (1-20), Hb 11,0 g/dl (12.1-16.1), MCV 71.4 fl (80-100), toxic granulation, atypical lymphocytes. Physician interpreted this as abnormal and estimated high disease probability. His diagnosis became 'infection or malignancy' and he ordered thoracic x-ray, abdominal ultrasound and more lab. This showed hypovitaminosis B12, x-ray was normal, ultrasound showed cholelithiasis. B12 was supplemented and later physician ordered lab tests for check-up. Results had improved.

Subgroup B: marginal abnormal laboratory results

Case 4: #2607, Male 72: initial diagnosis was 'functional complaints/malignancy' and the physician ordered lab tests to exclude disease. Abnormal results were: ESR 16 mm (1-15). Physician interpreted results as normal, estimated low disease probability, and his diagnosis became 'functional complaints/old age'. Four months later, patient experienced palpitations and was referred to a cardiologist. Six weeks later, the physician found lymph node enlargement and ordered lab tests.

Case 5: #4009, female 43; initial diagnosis was 'tired/mononucleosis infectiosa/depression' and the physician ordered lab tests to reassure the patient. Abnormal results were: ESR 13 mm (0-8), Hb 11.9 g/dl (12.6-16.3). Physician interpreted results as abnormal and estimated high disease probability, and his diagnosis became 'anemia/joint problems'. Patient wanted to go to a rheumatologist and was referred. Six months later, physician ordered lab tests for check-up.

Case 6: #714, Male 35: initial diagnosis was 'tiredness' and the physician ordered lab tests to exclude disease. Abnormal results were: ALAT 46 U/l (0-45). Physician interpreted results as possibly abnormal and estimated high disease probability. His diagnosis became 'tiredness/ slight liver function abnormality'. Patient also experienced palpitations and physician ordered 24-hour ECG. ECG was normal and physician ordered abdominal ultrasound for the ALAT abnormality (no abnormalities found).

Case 7: #4011, male 42: initial diagnosis was 'hyperventilation' and the physician ordered lab tests to exclude disease. Abnormal results were: glucose 126 mg/dl, ESR 28 mm (0-8), Hb 13.7 g/dl (14.0-17.7). Physician interpreted results as possibly abnormal and estimated low disease probability. Diagnosis was now 'impaired glucose tolerance/ hyperventilation'. Patient was referred to pulmonologist. His findings were normal. Two months later, the complaints occurred again and patient was referred to a cardiologist. His findings were also normal. One month later, patient experienced dizziness and was referred to an ENT specialist (all specialists reported normal findings).

ESR, erythrocyte sedimentation rate; Hb, hemoglobin; ECG, electrocardiogram; TSH, thyroid stimulating hormone; ALT, alanine aminotransferase; MCV, mean corpuscular volume; ENT, ear, nose, and throat.

Discussion

The literature describes cascade processes as processes in which an initial abnormal test result induces a growing number of subsequent referrals and investigations. To our knowledge, our study is the first to prospectively investigate the occurrence of such cascade processes in primary care. The results show that the physicians generally seemed to use a conservative, wait-and-see approach, even when laboratory results were abnormal. Cascade processes only occurred for a very limited number of patients: only 3.1% of patients (cases 4-7) with a minor abnormality had more than one investigation and/or referral. These findings suggest that cascade processes after laboratory testing are rare.

The physicians seemed to take a careful approach in response to abnormal results, and only for a limited number of patients did they choose to order more investigations. This was significantly related to the physicians' interpretation of the results. However, the literature shows that physicians' interpretation of results may show considerable variation. For example, physicians may differ in their estimates of what is a clinically important change in a laboratory test result, and they may use different laboratory values for diagnosis and workup with tests such as hemoglobin or ESR¹¹⁻¹³. A better understanding of how results are interpreted may therefore help to improve the ordering of diagnostic tests.

A few comments on the study's methods are in order. Firstly, the overall incidence of cascades was low. The group of patients with normal results even had no cascades at all. This might limit the generalizability, and future studies therefore may need to include more patients to find more cascades processes.

Secondly, although the primary care physicians in our study were comparable to the overall population of Dutch primary care physicians, it must be noted that Dutch primary care physicians tend to order fewer laboratory tests than those in other countries¹. Differences between countries in terms of medical culture, such as differences in dealing with uncertainty or risks of liability, may cause differences in the numbers of laboratory tests ordered. In those countries where physicians tend to order more tests, the frequency of cascade processes may differ from that found in our study.

Thirdly, the assessments in our study were limited to investigations and referrals ordered by the patients' primary care physician, because our primary interest was their management in situations of low pretest disease probability. We had no information about investigations ordered after the physician had referred the patient, as this was beyond the scope of our research. Our conclusions are therefore limited to the primary care physicians' actions, and our results must be interpreted with some caution, and need to be validated in other settings.

As far as we know, our study is the first to investigate the occurrence of cascade processes, and a major strength of this study was its prospective design: the physicians prospectively included and followed up patients for six months and filled in

the forms when they ordered the tests and when they received the results, while follow-up data was prospectively documented in the medical records.

We investigated laboratory tests commonly used in primary care. Other tests, such as spinal imaging, may have a greater likelihood of generating misleading results¹⁴. Indeed, a randomized trial of spinal imaging in low-risk patients suggested (although the findings were not statistically significant) that patients who received MRIs were twice as likely to undergo surgery as patients who received plain x-rays, yet had no better outcomes¹⁵. This may suggest that cascade effects may exist for other tests, but further research is needed.

To conclude, physicians seem to use a careful approach when responding to abnormal results in a low pretest probability situation: cascade processes in our study were rare, and the majority of patients had no further investigations or referrals at all. We found a significant relation between the interpretation of results of investigations and further actions. Therefore, future research should closely examine physicians' considerations when interpreting results, which may help improve the ordering of diagnostic tests.

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;45:21-5.
2. Mold JW, Aspy CB, Lawler FH. Outcomes of an insurance company-sponsored multichannel chemistry screening initiative. *J Fam Pract* 1998;47:110-7.
3. Phillips WR, Thompson DJ. Multi-channel laboratory testing and the unexpected abnormal result: a statistical myth corrected. *N Z Med J* 1981;94:462-4.
4. Brigden ML, Heathcote JC. Problems in interpreting laboratory tests. What do unexpected results mean? *Postgrad Med* 2000;107:145-6, 51-2, 55-8 passim.
5. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician* 1996;25:971-3.
6. Mold JW, Stein HF. The cascade effect in the clinical care of patients. *N Engl J Med* 1986;314:512-4.
7. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health* 2002;23:23-44.
8. Verstappen WH, van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, Grimshaw J, Grol RP. Effect of a practice-based strategy on test ordering performance of primary care physicians. A randomized trial. *JAMA* 2003;289:2407-12.
9. Winkens RAG, Dinant GJ. Evidence base of clinical diagnosis: Rational, cost effective use of investigations in clinical practice. *BMJ* 2002;324:783.
10. Houben PHH, van der Weijden T, van Bokhoven MA, Droog A, Winkens RAG, Grol R. Overwegingen van huisartsen bij het interpreteren van laboratoriumonderzoek; een kwalitatief onderzoek. *Huisarts Wet* 2005;48:326-32.
11. Skendzel LP. How physicians use laboratory tests. *JAMA* 1978;239:1077-80.
12. Daly MP, Sobal J. Anemia in the elderly. A survey of physicians' approaches to diagnosis and workup. *J Fam Pract* 1989;28:524-8.
13. Thue G, Sandberg S, Fugelli P. The erythrocyte sedimentation rate in general practice: clinical assessment based on case histories. *Scand J Clin Lab Invest* 1994;54:291-300.
14. Kendrick D, Fielding K, Bentley E, Kerslake R, Miller P, Pringle M. Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial. *BMJ* 2001;322:400-5.
15. Jarvik JG, Hollingworth W, Martin B, Emerson SS, Gray DT, Overman S, Robinson D, Staiger T, Wessbecher F, Sullivan SD, Kreuter W, Deyo RA. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *JAMA* 2003;289:2810-8.

Chapter 8

Discussion

Discussion

Outline of the study

General practitioners (GPs) order laboratory and imaging tests on a daily basis, and the results of these tests are routinely interpreted and applied for diagnosis and further management. However, due to the often low pretest probabilities of disease in general practice, and the lack of clinical significance of many abnormal test results, the interpretation of test results may be difficult and subsequent decisions may be incorrect.

Test results may be interpreted using Bayes' theorem to calculate posttest probabilities. But instead of using such formal numerical appraisal of test results, GPs may be more likely to interpret test results in the light of general clinical knowledge and pretest expectations such as the presumed diagnosis, the probability of disease and the reasons for test ordering. In view of the lack of data on GPs' current performance and decisions, we investigated the relations between GPs' pretest expectations and the occurrence of abnormal laboratory results, the interpretation of test results, further diagnosis and management.

This chapter starts with a summary of the main findings reported in this thesis. These findings are then discussed in the broader perspective of illness script and decision making theories and in the light of recent publications and methodological considerations. Finally, it discusses the implications for test ordering, for test result interpretation in daily practice and for further research.

Main findings

The laboratories reported a mean of 9.6 laboratory tests per patient (**Chapter 4**). Based on the reference values they used, one-fifth of the reported test results were abnormal, and almost a quarter of these were markedly abnormal. When there was a low estimated pretest probability of disease, or if tests were ordered to reassure patients, nearly half of the patients had one or more abnormal test results. Fewer patients (one in ten) had a markedly abnormal test result. There was a highly significant relation between the estimated pretest probability of disease and the percentage of the patients having abnormal results, with a higher estimated pretest probability resulting in higher numbers of patients with abnormal results.

The qualitative interview study (**Chapter 3**) showed that GPs generally tried to interpret abnormal test results in line with their pretest ideas about their patients' illness, using a wide variety of considerations. These related to aspects of the laboratory tests and aspects of the working hypotheses established before ordering the tests. Particularly important considerations were the GP's pretest estimation of the patient having a particular disorder and the degree of abnormality of a test result. All of the considerations GPs used when interpreting results were particularly evident in problem situations, for instance if tests yielded unexpected abnormal results.

Placing such unexpected abnormal laboratory results in the perspective of patients' overall clinical picture frequently resulted in the GPs not considering another disorder, nor changing their diagnosis.

Indeed, on average, the GPs interpreted abnormal laboratory test results as normal for nearly half of the patients, and undertook no further action (**Chapter 5**). This percentage was even higher if the pretest probability was low or if the tests were ordered to reassure the patient. The pretest expectations, in terms of estimated pretest probability and reason for ordering tests, strongly and significantly influenced the interpretation of the results and the estimated posttest probability. The interpretation of the test results and the estimated posttest probability in turn significantly influenced further management; the combination of test results interpreted as abnormal and a high posttest probability were more likely to lead to active management (such as additional investigations or referral).

This relationship was confirmed by analysing a separate, unrelated dataset, which showed that the management after receiving the results of spinal X-rays was strongly and significantly influenced by the reason for ordering the X-rays (**Chapter 6**). For example, active management was three times less likely if the X-rays had been ordered to reassure patients.

The final finding was that in situations of low pretest probability, abnormal laboratory results generally did not lead to a cascade of additional investigations or referrals (**Chapter 7**). The occurrence of such cascade processes was investigated among 127 patients with an abnormal laboratory result and an estimated low pretest probability. Cascades occurred only for seven patients, and these cascades were limited to two or three further investigations and referrals. In addition, three of these seven patients had markedly abnormal laboratory test results, justifying additional investigations or referrals, so a real (though limited) cascade effect occurred only for four patients.

Interpretation of test results by GPs and illness script theory

Despite the lack of numerical Bayesian calculations, the findings reported in this thesis show that the GPs carefully interpreted the meaning of test results against the reasons for ordering the tests and the background of the patients' clinical picture. These findings appear to fit in well with illness script theory^{1,2}.

Illness script theory describes how doctors often diagnose diseases: during a consultation they perceive clues about the patients' illness, and these clues subsequently activate an illness script. An illness script is a network of previously acquired medical knowledge and experience, which contains the characteristics, such as symptoms and physical signs, of a disease. Illness scripts often emerge as a result of (subconscious) memory association. They help doctors evaluate clinical findings and guide further requests for investigations.

Illness script theory describes how illness scripts are constantly evaluated against the clinical findings. If these clinical findings are compatible with the script, they are called

acceptable values, and such values confirm a script. For example, when hypothyroidism is suspected, an elevated TSH is an acceptable value confirming the hypothyroidism script. The elevated TSH value in a hypothyroidism script is also called the default value for the script. Default values are strong arguments for confirming a script. If the clinical findings do not fit a script, they are called non-acceptable values, and such values cause a script to be rejected. For example, if no disease is expected, a highly elevated ESR is a non-acceptable value, and the no-disease script is rejected¹.

Illness script theory may form the theoretical background against which the findings reported in this thesis can be evaluated, and it may help to understand the interpretation of test results in routine care. However, in the studies reported on in this thesis, the GPs often suspected no disease even though the patient had abnormal laboratory tests. The question is why these abnormal tests had no consequences, and how this fits in with illness script theory.

An explanation may be that two kinds of acceptable values may occur in illness script theory. Firstly, there are acceptable, default values. Default values are strong arguments for confirming a script. Secondly, there are acceptable, but unusual values: in the real diagnostic process, not all of the patients' clinical features usually perfectly match the signs and symptoms in the illness script. These clinical features, which are acceptable but do not match the illness script, are called unusual values. For example, if no disease is expected, a marginally elevated ESR may be an unusual value. Such an unusual value may be acceptable as it may not be strong enough to cause the script to be rejected. Unusual but acceptable values carry less weight than default values^{1,3}.

As the results in this thesis show, abnormal laboratory results when no disease was expected were frequently interpreted by the GPs as normal. These abnormal laboratory results may be a good example of unusual, but acceptable values⁴. However, GPs may differ in the level of unusual values they are prepared to accept, depending on their medical knowledge and clinical experience^{5,6}. For example, Norwegian general practitioners differ in their estimation of the maximum ESR values at which action is necessary⁷. Also, doctors who see more elderly patients may accept lower haemoglobin values before initiating work-up for anaemia⁸. The interpretation of test results may be influenced not only by knowledge and experience, but also by other determinants related to the health professionals, such as uncertainty, risk-taking attitudes and gut feelings^{9,10}.

Summarising, illness script theory may offer a good theoretical basis for our findings that the interpretation of test results is done against the background of the reasons for which the tests were ordered and the patient's clinical picture. It may also provide an explanation for the fact that many marginally abnormal results are accepted and interpreted as normal in situations of low disease probability.

In the literature, test interpretation is usually evaluated using Bayesian decision making theory². Bayes' theorem can be used to calculate the posttest probability of

disease from the pretest probability, the test results and the sensitivity and specificity of the tests^{3,11-13}. Alternatively, the likelihood ratio may be used^{14,15}.

However, these calculations are difficult, and few doctors (3% in a telephone survey) claim to apply formal Bayesian calculations in day-to-day care¹⁷. An important reason mentioned was the impracticality of the Bayesian method. Also, of the doctors who said that they used sensitivity and specificity indexes, many (95%) did not do so in the recommended formal manner¹⁷.

Several aspects may contribute to the complexity of these calculations for use in daily practice. For example, GPs usually consider several diagnoses at the same time, and they often order several tests. Calculating probabilities for all these tests and diagnoses is often not feasible. Also, calculations become more complex when tests are mutually dependent, such as MCV and haemoglobin. The chance that both values are abnormal is not equal to the chance that the independent tests are abnormal. Furthermore, pretest expectations (estimated prevalence of disease), sensitivity and specificity of tests are often not specifically known for the situations where the tests are used. For example, the sensitivity and specificity of the exercise test to diagnose coronary heart disease may differ for men and women and may vary with the systolic blood pressure at baseline¹⁶.

Although explicit use of Bayesian calculations seems to be very limited, one of its basic principles, that the test results are evaluated against the pretest probability seems to be more widely used, although probably in an implicit manner. Our findings show that the GPs' estimation of pretest probability, although it was not an estimation of the probability of a specific disease but a more general estimation of the probability of finding any disease, had an important influence on the further diagnosis and management.

Although the Bayesian model may have shortcomings in describing how GPs interpret test results in daily practice, it may help to direct attention to two major classes of errors in clinical reasoning². The first is that of errors in the estimation of pretest probability. Several experimental studies have shown that doctors' estimations of pretest probability for the same diseases may vary widely¹⁸. The second class of errors is that doctors may incorrectly assess the strength of evidence, for example how reliable a laboratory test is in terms of excluding or confirming a diagnosis¹⁹.

In summary, both illness script theory and decision making theory provide a theoretical framework which may help to understand the interpretation of test results. Illness script theory may be the most useful theory to explain test interpretation in routine care, whereas decision making theory shows how test results should be interpreted. Illness script theory corresponds well with the findings reported in this thesis, which show that interpretation of results in day-to-day care takes place against the background of pretest expectations and the patient's clinical picture. Based on this clinical picture, abnormal laboratory results were often accepted and interpreted as normal. In general, GPs appeared to have few difficulties interpreting abnormal results in situations of low disease probability. Decision making

theory underlines the importance of the pretest probability for evaluating test results and may help to understand possible errors in the interpretation of test results. Illness script theory and decision making theory are not mutually exclusive, but offer two different views on the diagnostic process, and both may help to understand the interpretation of results.

Indications for test ordering

Many studies have shown that tests in routine practice are frequently ordered when this is not necessary from a strictly medical point of view²⁰. It is assumed that such superfluous testing may have several unfavourable side-effects. For instance, abnormal results of unnecessary tests may be difficult to interpret, which may have a negative influence on diagnosis and management and sometimes induce a cascade of further investigations and referrals^{4,21,22}. It is also assumed that the patient may get worried if results are abnormal. And there are, of course, costs involved in these diagnostic tests. However, the results reported in this thesis shed a different light on some of these unfavourable side-effects, which may be smaller than assumed.

Our studies found that the GPs ordered many tests for reasons such as excluding disease and reassuring patients. Our findings showed that many of these tests yielded marginally abnormal results. However, these abnormal results were interpreted by the GPs in the clinical context, and they did not seem to experience problems with the interpretation of these results, and very few cascades of further investigations and referrals occurred. In addition, important laboratory test abnormalities were sometimes found even if the tests were ordered to exclude disease or to reassure the patient. These abnormalities might have important consequences for diagnosis or therapy and might have been missed if the GPs had felt more restricted in ordering tests, inducing under-diagnosing of disease.

The literature review in **Chapter 2** showed that many patients feel reassured if the results of tests are normal. However, some patients remain anxious, which may be due to pre-existent social or psychological morbidity. Furthermore, patients may have difficulties understanding that abnormal results can be clinically irrelevant and they may find it difficult to understand what false-positive results are²³⁻²⁷. This may increase anxiety and may have an adverse influence on further diagnosis and management. Thus, testing may often help to reassure, but may also have negative consequences.

A study of medically unexplained complaints in general practice investigated the costs of a watchful waiting approach compared to the costs of immediate testing²⁸. It found that the total costs of health care, including costs of tests, referrals and medication, were higher for immediate testing, although the difference was not significant. Superfluous testing may thus increase costs, although the literature about the costs of superfluous testing is limited and more research is necessary.

In summary, the results reported in this thesis show that from a doctors' perspective, the unfavourable side effects of superfluous testing may be smaller than is sometimes assumed. These findings may correspond well with GPs' experiences: an intervention study aimed at reducing test ordering found that, despite support from Bayes' theorem, the participating GPs perceived immediate advantages of test ordering, without perceiving any disadvantages²⁹. However, negative consequences of superfluous testing, such as those regarding patients' reassurance and costs, may also exist. Careful considerations are therefore needed in the debate about the burden of superfluous test ordering and its unfavourable side-effects.

Methodological considerations

The studies reported on in this thesis used a broad, explorative approach. Since we wanted to investigate day-to-day care, we included all patients consulting a GP and referred for laboratory testing. As a result, however, there was much heterogeneity in the laboratory tests ordered and in the expected underlying diagnoses. This heterogeneity makes it difficult to apply the results to specific situations, specific disorders or specific laboratory tests. For instance, the results are difficult to apply to GPs' interpretation of results for medically unexplained symptoms (MUS), such as fatigue. But an initially broad explorative approach of the studies seemed appropriate, in view of the limited knowledge about GPs' interpretation of test results in day-to-day care. In this setting, it would have been unrealistic to reduce clinical variation to a minimum. The heterogeneity in terms of tests and diagnoses implies that our findings must be carefully interpreted and adapted to specific situations. It may be interesting to investigate more specific situations, disorders or laboratory tests in future studies.

The measurements reported in **Chapters 4, 5 and 7** were collected for the IDO study, in which each participating doctor was personally visited and instructed to record data on 25 adult patients for whom they had decided to order laboratory tests during the consultation. The doctors recorded data both when they ordered the tests and at the moment when they received the test results. The data collection was successful in terms of the number of participating GPs (87) and patients (1775). However, there was sometimes a gap of more than just a few days between the dates mentioned on the inquiry form when tests were ordered and when the results were received. This suggests that not all GPs recorded data at the moment when they received the results, some recording them at a later moment. This may have biased the findings reported in this thesis regarding GPs' interpretation of results, as these GPs could have had knowledge about the further course of the patients' complaints. However, this bias is presumably small, as many complaints in general practice are self-limiting, and knowledge about the further course will usually not have greatly changed the interpretation of the results. Furthermore, the time it took some GPs to include 25 patients suggests that some GPs might not have included successive patients as instructed, which may have introduced selection bias. Some GPs may only have

included patients during a quiet and easy surgery, or included patients on the basis of their expected willingness to sign informed consent. However, this selection bias was probably small, as it seems unlikely that the decision to select patients was related to the reason for test ordering, the pretest probability or the type of laboratory tests. Thus, although some bias may have been introduced in the data collection, it was presumably small and had no significant effect on the conclusions presented in this thesis.

This research project started with a qualitative study showing several considerations that are important for the interpretation of test results. Important considerations were the GPs' estimation of the risk of disease and the reason for ordering the tests. These two considerations were included in the questionnaire used for the studies in the **Chapters 4, 5 and 7**, but it was not possible to investigate all the other considerations mentioned in the qualitative study. For example, the questionnaire did not include questions about GPs' opinions about test characteristics. This may have influenced the results, although systematic confounding is improbable, as the type and size of this influence may have been different for each test and patient.

The participating GPs were a representative sample of Dutch GPs, who are generally known to order fewer tests than GPs in some other countries³⁰. There may be several explanations for this difference, such as differences in medical culture or defensive attitudes in testing³¹⁻³³. Such differences in defensive attitudes may affect the external generalisability of our findings, for instance in that a more defensive attitude may decrease the doctor's tolerance to abnormal laboratory results. As a consequence, cascade processes to explain laboratory abnormalities may occur more frequently among GPs with a low tolerance for abnormal results. Although it is plausible that principles similar to those we found may also be important in other clinical environments, generalisation of the findings in this thesis to other settings should be done with some caution, and interpretation of results in other clinical environments needs further investigation.

Implications for further research

Many studies have shown that unnecessary tests are sometimes ordered. However, it is not clear how often tests are unjustifiably not ordered. Failing to order necessary tests may lead to under-diagnosis of disease, with potentially serious consequences for the patient. More research should be focused on the frequency of and reasons for failure to order certain tests.

The studies about the interpretation of test results in this thesis were mainly explorative. Future studies may focus on the interpretation of a specific test or on the interpretation of tests for a specific disorder. For example, many tests in general practice are ordered for medically unexplained complaints, and some of these tests can be considered unnecessary. Future studies could investigate the problems that occur when GPs interpret the results of test ordered for such unexplained complaints.

Two other issues, which were only addressed in the review study reported in this thesis, seem relevant for further research: the management of test results and ways of communicating test results to patients. Several studies in other countries than the Netherlands have shown that management of test results in routine practice is often not optimal³⁴. However, little is so far known about Dutch GPs' achievements in terms of managing test results. Further research is therefore recommended, possibly based on the four elementary steps described by Boohaker³⁵: (a) tracking whether results of tests ordered have been received, (b) informing patients about the results, (c) documenting that the patient has been informed, and (d) ensuring that actions are taken if necessary and that follow-up is carried out. When these steps are included in studies, it is not only important to know how often appropriate management of these steps is accomplished, but we should know the consequences and impacts of inappropriate test results management, for example, how often the start of an important therapy is delayed due to an error in test results management.

Another aspect worthy of study is that the decision to order tests is often influenced by patient anxiety and expectations. When the results of their tests prove normal, many patients feel reassured^{36,37}. However, some patients are not reassured by normal results, which may be due to prior anxiety and to illness perceptions, such as concerns about illnesses or symptoms experienced^{38,39}. In addition, patients sometimes have difficulties understanding that abnormal results may have no clinical significance, for instance when results are marginally abnormal. To help GPs discuss test results with patients, more research is needed about patients' perceptions of test results and ways to effectively communicate the meaning of test results.

Implications for education and daily practice

The findings in this thesis may have several implications for education and clinical practice. Firstly, it is important to teach doctors and students about the diagnostic value of tests. This is usually done along the principles of decision making theory. Teaching decision making theory may help students understand the limitations of diagnostic tests and the importance of the pretest probability. Students and doctors may have difficulties understanding and applying the principles of decision making theory, but some approaches may help increase their understanding, such as representing information as frequencies instead of probabilities^{15,19,40}. Nevertheless, calculations based on decision making theory may be too complex for day-to-day use. Teaching based on illness script theory may be more successful for clinical tasks. According to illness script theory, doctors perceive clues about the patients' illness which activate an illness script. Test results are constantly evaluated against this script and may be considered acceptable, i.e. confirming a script, or may be considered unacceptable, causing the script to be rejected. Teaching based on illness script theory may focus on the generation of sufficient suitable illness scripts for a given patient case presented to students. Subsequently, the students could be presented with test

results which they have to evaluate as acceptable or unacceptable according to the illness scripts generated. Teaching from the perspective of illness script theory may thus help to improve interpretation of laboratory tests.

In everyday care, many patients have abnormal laboratory test results, even when tests were ordered for reasons such as patient reassurance. It is important that GPs are aware that many of these abnormalities are due to statistical chance. The abnormalities need to be interpreted against the clinical background and the reasons for ordering the tests, which may help to put abnormalities in perspective and may help to achieve a balance between overreacting and underreacting to abnormal results.

Lastly, many tests ordered in general practice may be considered medically unnecessary from the perspective of medical decision making. It is assumed that such tests may have negative consequences for the interpretation of tests, diagnosis, management, patient anxiety and costs of health care. This might suggest that the ordering of such tests should be discouraged. However, the findings reported in this thesis indicate that some negative consequences may well be smaller than expected. Therefore, a certain degree of tolerance might be in order when GPs order superfluous tests. This perspective on the magnitude of the burden of superfluous test ordering also implies that it may be necessary to reconsider the necessity of costly strategies, such as computerised decision support systems, to improve test ordering behaviour.

References

1. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med.* 2000;75:182-90.
2. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ.* 2002;324:729-32.
3. Kassirer JP. Diagnostic reasoning. *Ann Intern Med.* 1989;110:893-900.
4. Bridgen ML, Heathcote JC. Problems in interpreting laboratory tests. What do unexpected results mean? *Postgrad Med.* 2000;107:145-6, 51-2, 55-8 passim.
5. Thue G, Sandberg S, Fugelli P. Clinical assessment of haemoglobin values by general practitioners related to analytical and biological variation. *Scand J Clin Lab Invest.* 1991;51:453-9.
6. Hani MA, Keller H, Vandenesch J, Sonnichsen AC, Griffiths F, Donner-Banzhoff N. Different from what the textbooks say: how GPs diagnose coronary heart disease. *Fam Pract.* 2007;24:622-7.
7. Thue G, Sandberg S, Fugelli P. The erythrocyte sedimentation rate in general practice: clinical assessment based on case histories. *Scand J Clin Lab Invest.* 1994;54:291-300.
8. Daly MP, Sobal J. Anemia in the elderly. A survey of physicians' approaches to diagnosis and workup. *J Fam Pract.* 1989;28:524-8.
9. Zaat JO, van Eijk JT. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care.* 1992;30:846-54.
10. Stolper E, van Bokhoven M, Houben P, Van Royen P, van de Wiel M, van der Weijden T, et al. The diagnostic role of gut feelings in general practice. A focus group study of the concept and its determinants. *BMC Fam Pract.* 2009;10:17.
11. Knottnerus JA. Diagnostic prediction rules: principles, requirements and pitfalls. *Prim Care.* 1995;22:341-63.
12. Hagen MD. Test Characteristics. How good is that test? *Prim Care.* 1995;22:213-33.
13. Lyman GH, Balducci L. The effect of changing disease risk on clinical reasoning. *J Gen Intern Med.* 1994;9:488-95.
14. Sonis J. How to use and interpret interval likelihood ratios. *Fam Med.* 1999;31:432-7.
15. Puhan MA, Steurer J, Bachmann LM, ter Riet G. A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Ann Intern Med.* 2005 Aug 2;143:184-9.
16. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology.* 1997;8:12-7.
17. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med.* 1998;104:374-80.
18. Heller RF, Sandars JE, Patterson L, McElduff P. GPs' and physicians' interpretation of risks, benefits and diagnostic test results. *Fam Pract.* 2004;21:155-9.
19. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ.* 2002;324:824-6.
20. van Walraven C, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA.* 1998;280:550-8.
21. Vafiadis P. The dilemma of the unexpected result. *Aust Fam Physician.* 1996;25:971-3.
22. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health.* 2002;23:23-44.
23. van Bokhoven MA, Pleunis-van Empel MC, Koch H, Grol RP, Dinant GJ, van der Weijden T. Why do patients want to have their blood tested? A qualitative study of patient expectations in general practice. *BMC Fam Pract.* 2006;7:75.
24. Adelsward V, Sachs L. The meaning of 6.8: numeracy and normality in health information talks. *Soc Sci Med.* 1996;43:1179-87.
25. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making.* 2001;21:37-44.
26. Hamm RM, Smith SL. The accuracy of patients' judgments of disease probability and test sensitivity and specificity. *J Fam Pract.* 1998;47:44-52.
27. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *BMJ.* 2006;333:284.

28. Van Bokhoven M. Blood test ordering for unexplained complaints in general practice. The feasibility of a watchful waiting approach. Maastricht: Maastricht University; 2008:145-56.
29. Van Bokhoven M. Blood test ordering for unexplained complaints in general practice. The feasibility of a watchful waiting approach. Maastricht: Maastricht University; 2008:186-98.
30. 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;45:21-5.
31. Van Boven K, Dijksterhuis P, Lamberts H. Defensive testing in Dutch family practice. Is the grass greener on the other side of the ocean? *J Fam Pract.* 1997;44:468-72.
32. van der Weijden T, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract.* 2002;52: 974-80.
33. Holtgrave DR, Lawler F, Spann SJ. Physicians' risk attitudes, laboratory usage, and referral decisions: the case of an academic family practice center. *Med Decis Making.* 1991;11:125-30.
34. Hickner JM, Fernald DH, Harris DM, Poon EG, Elder NC, Mold JW. Issues and initiatives in the testing process in primary care physician offices. *Jt Comm J Qual Patient Saf.* 2005;31:81-9.
35. Boohaker EA, Ward RE, Uman JE, McCarthy BD. Patient notification and follow-up of abnormal test results. A physician survey. *Arch Intern Med.* 1996;156:327-31.
36. Lucock MP, Morley S, White C, Peake MD. Responses of consecutive patients to reassurance after gastroscopy: results of self administered questionnaire survey. *BMJ.* 1997;315:572-5.
37. Howard L, Wessely S, Leese M, Page L, McCrone P, Husain K, et al. Are investigations anxiolytic or anxiogenic? A randomised controlled trial of neuroimaging to provide reassurance in chronic daily headache. *J Neurol Neurosurg Psychiatry.* 2005;76:1558-64.
38. McDonald IG, Daly J, Jelinek VM, Panetta F, Gutman JM. Opening Pandora's box: the unpredictability of reassurance by a normal test result. *BMJ.* 1996;313:329-32.
39. Donkin L, Ellis CJ, Powell R, Broadbent E, Gamble G, Petrie KJ. Illness perceptions predict reassurance following a negative exercise stress testing result. *Psychol Health.* 2006;21:421-30.
40. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med.* 1998;73:538-40.

Summary

Summary

Chapter 1

This thesis examines how general practitioners (GPs) interpret the results of laboratory and imaging diagnostics and use these to establish a diagnosis and decide on further management. Before we started our study, there had been little research into this process in routine general practice, and little was known about the problems involved. What we did know was that large numbers of diagnostic tests were being ordered without sufficient reason from the point of view of medical decision making theory. Since the pretest probability of disease is often low in general practice, there will be many cases where such tests yield abnormal results that have no clinical relevance, and these test results may lead to undesirable consequences, such as unjustified follow-up diagnostics and overtreatment. This underlines the importance of descriptive research into the interpretation of diagnostic test results and the ensuing management.

Medical decision making theory describes how the results of diagnostics can be interpreted with the help of calculations based on Bayes' Theorem, using the pretest probability of disease, and the sensitivity and specificity of the tests. It is well-known, however, that such calculations are rarely made in routine general practice. Hence, it is likely that diagnostic test results are not usually interpreted with the help of Bayesian calculations, but are rather evaluated in light of the doctor's expectations and intuitions in terms of the estimated probability of disease, the suspected diagnosis and the reason for ordering diagnostic tests. We tried to obtain a better understanding of the way test results are interpreted in routine practice, and explored how much room for improvement there is, by addressing the following research questions:

1. What problems relating to the use of test results have been reported in the literature?
2. What considerations do GPs use when interpreting laboratory results in routine practice?
3. How frequent are mildly and markedly abnormal laboratory results in relation to pretest expectations?
4. How do pretest expectations affect the interpretation of laboratory results and further management?
5. Are GPs less inclined to initiate active management if the reason for requesting a spinal X-ray is to reassure the patient?
6. How common are cascades of further investigations and what determinants increase the risk of such cascades?

Chapter 2

The problems associated with using test results in routine practice occur at three levels: the organizational level, the level of the health care professionals and that of the patients. Since these problems had not been systematically identified, we started a systematic search of the literature to identify as many of them as possible. The results of our search strategy were also used to draw up a research agenda with questions that needed further investigation.

Problems at the organizational level mainly concern the processing of test results. GPs do not always check whether results of tests they order actually come in, and they may not use an effective system to check whether the test results are actually communicated to the patients and whether patients adhere to the agreed follow-up. This may lead to medical errors. Further research would have to focus on investigating the seriousness, nature and consequences of such errors, on establishing the requirements that procedures to process results should meet and on developing interventions to improve these procedures.

Problems at the level of the health care professionals mainly concern the interpretation of test results. GPs sometimes erroneously regard clinically irrelevant abnormalities as important, which may lead to a cascade of further investigations, whereas in other cases they may erroneously regard abnormalities as irrelevant and fail to investigate further. Since there is little information available about the origins of such problems, important research questions are: What are the determinants or principles underlying the interpretation of test results? Can general causes of such errors be identified? What would be effective strategies to improve the interpretation of test results?

A major problem at the patient level is that patients often have difficulty understanding the meaning of test results. They tend to evaluate test results in a dichotomous way: good or bad, and they are often insufficiently aware of the probability of false-positive and false-negative results. In addition, GPs sometimes request diagnostic tests to reassure a patient, but this may be ineffective or even counterproductive for some patients. Further research should examine ways for GPs to clarify test results to their patients and reveal which patients could benefit from having tests done to reassure them and for which patients this would be an ineffective or even adverse approach.

Chapter 3

Medical research literature has devoted much attention to the calculations used in decision theory to interpret test results, even though such calculations are in fact rarely used in routine practice, whereas very little is known about the actual considerations that doctors mainly use in everyday practice to interpret these results. We therefore interviewed 21 GPs about their considerations when interpreting the laboratory results of 191 patients, using a qualitative approach.

The GPs' considerations fell into two categories: considerations relating to the working hypothesis (e.g., whether the working hypothesis concerned a specific disease and the estimated pretest probability of disease) and considerations relating to the laboratory tests (e.g. the unreliability of reference values, previous test results and the degree of abnormality). These considerations were especially important in situations where the test results did not match the working hypothesis, for instance if a test shows a minor abnormality whereas the GP was of the opinion that the patient had no disease. It was in these situations that the GPs frequently referred to considerations like the unreliability of reference values, thus trying to make the test results fit their working hypothesis and putting the abnormal results in the perspective of the overall clinical picture. Hence, abnormal results for patients with a low pretest disease probability usually did not induce the GPs to change their working hypothesis.

Chapter 4

Interpreting abnormal test results may be problematic in situations with a low pretest probability of disease. This is especially true for mildly abnormal results, whereas markedly abnormal results are often easier to interpret. Although research had shown that abnormal test results frequently occur in general practice, it was unknown in what proportion of cases this concerned mild or marked abnormalities, nor what the relation was between the number of abnormalities and the pretest expectations. We therefore studied the numbers of mildly and markedly abnormal laboratory results and the relation with pretest expectations in a study among 1775 patients registered with 87 GPs.

The laboratories reported an average of 9.9 tests when the GPs wanted to exclude disease, against 8.9 tests when the aim was to reassure the patient and 7.9 tests if they were ordered for monitoring purposes. Twenty-one percent of all these tests yielded abnormal results, with 5% markedly abnormal results. Most patients (64%) had at least one abnormal test result, while a minority, though still a substantial proportion (21%), had at least one markedly abnormal result. Many abnormal results were found in situations where the estimated pretest probability of disease was low, or even when the tests had been requested to reassure the patient. About half of the patients for whom tests had been ordered to reassure them had at least one abnormal result, while about one in ten of these patients had at least one markedly abnormal test result.

It is clear, then, that abnormal test results are very common, and many of them are found even in situations with a low pretest probability of disease. In view of the statistical definitions used for reference values, however, there is a reasonable chance that such abnormalities have no clinical relevance, and further research should investigate the consequences of these test results for the diagnosis and further management.

Chapter 5

Even though GPs frequently request laboratory diagnostics in situations with a low pretest probability of disease, many of such tests still yield abnormal results. GPs can interpret such results by putting them in the perspective of pretest considerations such as the reason for requesting the tests and the estimated pretest probability of disease, but it was unknown to what extent such pretest considerations affected the interpretation of abnormal test results. We therefore examined the relation between interpretations of test results and pretest expectations.

The study included 87 GPs, who requested laboratory tests for 1253 patients. The data were analysed using cross-tabulations, chi-square tests and a multivariate logistic regression model. The results of the cross-tabulations showed that abnormal test results were significantly more often evaluated as normal if they had been ordered in a situation of low pretest disease probability than if a high pretest probability had been estimated (66.1% versus 19.6%). Such abnormal results were also significantly more frequently evaluated as normal if the tests had been ordered to exclude a disease than if they had been ordered to establish the presence of a disease (50.9% versus 27.7%). This relation was also significant in the multivariate logistic model. The model also showed that if tests were interpreted as abnormal, there was a significantly higher chance that the posttest probability was estimated to be high. Such a high posttest probability estimate in turn resulted in a significantly higher probability of active management.

These results thus show that GPs attach great value to pretest expectations when interpreting test results and deciding on further management. On the whole, this is probably a good strategy, since many abnormal test results are not clinically relevant, and this approach enables GPs to find a balance between overreacting and underreacting to laboratory test results.

Chapter 6

Spinal X-rays are often ordered without strict medical necessity, according to prevailing guidelines. One of the reasons why GPs request such imaging anyway is to reassure their patients. What consequences do the findings of such examinations have for further management, and do GPs take their reason for requesting the X-rays into account when deciding on further management?

The study involved 90 GPs, who reported their reason for requesting X-rays on the ordering form. After receiving the results, they were asked to answer some questions on the further management they intended to pursue. A multivariate model was used to examine the relation between the reasons for ordering the X-rays and the subsequent management, taking account of the findings on the images, the differential diagnosis and the patient's age.

Thirty-six percent of the patients had no abnormalities on their X-rays, while 48% had mild degenerative abnormalities and 17% had marked abnormalities. The GPs said

they intended to apply a wait-and-see policy for two-thirds of the patients whose X-ray images showed no abnormalities, and for half of the patients whose X-rays showed mild degenerative abnormalities. Fifteen percent of the X-rays had been ordered to reassure the patient. The reason for ordering X-rays had a significant effect on management: if GPs requested spinal X-rays to reassure patients, active management, such as referral, further investigations or prescribing medication, was about three times less likely (after correction for the X-ray findings) than if they had been ordered for other reasons.

Chapter 7

Since reference values for the results of laboratory tests are statistically defined, it is possible that abnormal values (according to the reference) are found for people who are in fact healthy. This means that an initial abnormal test result may erroneously lead to a process of more and more examinations being requested and to referrals, a process also known as a cascade. It was unknown how often such cascade processes occur in general practice and how this is influenced by determinants like the estimated pretest probability of disease and the interpretation of abnormal test results.

We examined this in a study involving 142 patients who had a low pretest disease probability and an abnormal test result and 142 age- and gender-matched patients who also had a low pretest disease probability but had no abnormal test result. We analysed how often follow-up examinations were requested and how many patients were referred. The chi-square test was used to assess whether the number of test requests and the number of referrals were influenced by the reasons for ordering laboratory tests, the estimated posttest probability of disease and the interpretation of the test results.

For the patients with an abnormal test result, after receiving the test results, the GPs requested more than one referral or more than one additional investigation for only 7 patients (5.5%), with a maximum of 3 referrals and additional investigations. Three of the 7 patients had markedly abnormal initial test results, justifying further investigations, whereas the initial test results for the other 4 patients were only mildly abnormal. Our findings show that cascade processes are actually very rare and are limited in magnitude.

Chapter 8

The findings of the various studies reported on in the thesis can be evaluated on the basis of two theories: medical decision making theory and illness script theory, offering two different perspectives on the interpretation of test results. The findings of the studies in this thesis seem to be particularly compatible with the illness script theory.

Illness script theory describes how doctors interpret diagnostic information in everyday practice. In most cases, doctors develop a hypothesis at an early moment in the consultation. Such a hypothesis is called an illness script, and it encompasses all characteristics of a particular illness, including the results of diagnostic tests that are compatible with the relevant clinical entity. The actual results of the diagnostic tests are then compared with the values the doctor expected on the basis of the illness script. If the values fit in, the script is accepted, if not then it is rejected.

Our research shows that GPs sometimes accept a script even though the test results deviate slightly from the expected values. In other words, certain values may not be compatible with the script, but are not considered important enough to reject the script. Such values are apparently regarded as acceptable by GPs. It is suggested that this is because values that are not compatible with a particular script are often less important in terms of script rejection than compatible values are in terms of confirming it. This view appears to be substantiated by the findings reported in this thesis, in that GPs who did not suspect a disease were more likely to interpret abnormal test results as normal, and their suspicion of disease did not increase.

Medical decision making theory describes how the posttest probability of disease can be calculated from the pretest probability, the test result and the test characteristics. Such calculations are not easy, however, because the pretest probability is often not accurately known, and because GPs often order a set of tests rather than one single test. As a result, these calculations are rarely made in routine practice. On the other hand, the theory does help to explain certain errors made when interpreting test results, by emphasizing the importance of accurate estimations of pretest probability and of the predictive value of a test.

Although medical students are often taught to interpret test results on the basis of medical decision making theory, the illness scripts theory may fit in better with the way test results are interpreted in routine practice. Giving more attention to this theory could help students and practising doctors to interpret test results better.

The findings reported in this thesis indicate that GPs are generally well able to handle the results of tests ordered in situations of low pretest probability of disease, or to reassure patients. We found that GPs usually had no problems interpreting results and that cascades of diagnostics are rare. These may be reasons for GPs to continue ordering diagnostic tests even in situations where the guidelines recommend not doing so.

Recommendations for further research made in this thesis focus on three major subjects: the processing of test results in general practice, the interpretation of test results and how well patients understand the meaning of test results. Each of these areas has its own specific questions that future research could focus on, offering opportunities to improve the quality of health care as regards the use of diagnostics in everyday practice.

Samenvatting

Samenvatting

Hoofdstuk 1

In dit proefschrift wordt bestudeerd hoe huisartsen uitslagen van laboratorium en beeldvormende diagnostiek interpreteren en toepassen voor het stellen van een diagnose en het bepalen van verder beleid. Hoe dit in de dagelijkse praktijk gaat was nog maar weinig bestudeerd, en er was weinig bekend over de problemen die hierbij kunnen ontstaan. Het was wél bekend dat veel aanvullende diagnostische tests worden aangevraagd zonder dat dit vanuit een medisch besiskundig oogpunt geïndiceerd is. Doordat de pretest kans op ziekte in de huisartsenpraktijk vaak gering is, zijn de uitslagen van zulke diagnostische tests relatief vaak afwijkend zonder dat dat klinisch relevant is. Deze uitslagen kunnen daardoor leiden tot ongewenste consequenties zoals onnodige vervolgdagnostiek en overbehandeling. Beschrijvend onderzoek naar de interpretatie van uitslagen van diagnostische tests en het verdere beleid is daarom van belang.

Medical decision making beschrijft hoe uitslagen van diagnostiek geïnterpreteerd kunnen worden met behulp van berekeningen volgens de Theorema van Bayes met pretest kans op ziekte, sensitiviteit en de specificiteit. Het is echter bekend dat dit soort berekeningen in de dagelijkse praktijk nauwelijks worden toegepast. Het is dus goed mogelijk dat uitslagen van diagnostiek in de dagelijkse praktijk niet met Bayesiaanse calculaties worden benaderd, maar veel meer worden gezien in het licht van verwachtingen en intuïties ten aanzien van de geschatte kans op ziekte, de vermoedde diagnose en de reden om de diagnostiek aan te vragen. Om het interpreteren van uitslagen in de dagelijkse praktijk beter te begrijpen, en de ruimte te exploreren die er is voor verbetering, zijn de volgende vragen voor dit proefschrift opgesteld:

1. Welke problemen bij het gebruik van testuitslagen worden in de literatuur herkend?
2. Welke overwegingen heeft de huisarts bij het interpreteren van laboratoriumuitslagen in de dagelijkse praktijk?
3. Hoe vaak komen afwijkende laboratoriumuitslagen en ernstig afwijkende laboratoriumuitslagen voor in relatie tot de pretest verwachtingen?
4. Welke invloed hebben de pretest verwachtingen op het interpreteren van laboratoriumuitslagen en het verdere beleid?
5. Zijn huisartsen minder geneigd actief beleid te plannen als een röntgenfoto van de wervelkolom was aangevraagd om de patiënt gerust te stellen?
6. Hoe vaak komt een cascade proces voor en welke determinanten verhogen het risico hierop?

Hoofdstuk 2

De problemen die zich voordoen bij het gebruik van testuitslagen in de dagelijkse praktijk kunnen onderscheiden worden op drie nivo's: organisatorisch, professioneel en patiënt. Deze problemen waren niet eerder systematisch geïnventariseerd en daarom werd door middel van een zoekstrategie in de literatuur getracht zoveel mogelijk van deze problemen te identificeren. Aan de hand van de resultaten van de zoekstrategie werd ook een agenda opgesteld met vragen voor verder onderzoek.

Het bleek dat problemen op organisatorisch nivo zich met name voordoen in de verwerking van testuitslagen. Zo controleren huisartsen niet altijd of uitslagen van aangevraagde tests ontvangen worden, en ontbreekt het aan effectieve controle of patiënten de uitslagen van tests vernemen, en of de afgesproken follow-up nagekomen wordt. Hierdoor kunnen medische fouten in de hand gewerkt worden. Verder onderzoek zou zich moeten richten op de ernst, aard en consequenties van deze fouten, zou eisen moeten opstellen waaraan de verwerking van uitslagen moet voldoen, en zou interventies moeten ontwikkelen om de verwerking van uitslagen te verbeteren.

Problemen op professioneel nivo doen zich met name voor bij het interpreteren van uitslagen. Huisartsen beschouwen klinisch niet relevante afwijkingen soms onterecht als belangrijk waardoor een cascade aan verdere diagnostiek kan ontstaan. Anderzijds beschouwen ze afwijkingen soms onterecht als niet relevant en wordt verder onderzoek juist achterwege gelaten. Inzicht in het ontstaan van deze problemen is echter gering en daarom zijn belangrijke vragen voor verder onderzoek: wat zijn de determinanten/principes die aan het interpreteren van uitslagen ten grondslag liggen? Zijn er bepaalde algemene oorzaken te vinden die fouten veroorzaken? Wat zijn effectieve strategieën om het interpreteren van uitslagen te verbeteren?

Een belangrijk probleem op patiënt nivo is dat het voor patiënten moeilijk is de betekenis van een uitslag goed te begrijpen. Ze zijn geneigd een uitslag dichotoom te beoordelen: goed of fout. Ook schatten ze de kans op een fout-positieve of fout-negatieve uitslag vaak verkeerd in. Daarnaast vragen huisartsen diagnostische tests soms aan om de patiënt gerust te stellen, maar voor sommige patiënten is dit niet effectief of zelfs contra-productief. Verder onderzoek moet helpen duidelijk te krijgen hoe een huisarts uitslagen goed kan bespreken met patiënten en het moet helpen inzicht te krijgen bij welke patiënten het effectief is om onderzoek te doen om gerust te stellen en bij welke patiënten het niet effectief is of zelfs nadelige effecten heeft.

Hoofdstuk 3

In de medische literatuur is er veel aandacht voor besiskundige berekeningen om uitslagen te interpreteren, maar deze berekeningen worden in de praktijk weinig gebruikt. Daarentegen zijn de overwegingen die een belangrijke rol spelen in de dagelijkse praktijk bij het interpreteren van uitslagen grotendeels onbekend. Daarom werden 21 huisartsen volgens een kwalitatieve systematiek geïnterviewd over hun overwegingen bij het interpreteren van laboratoriumuitslagen van 191 patiënten.

De overwegingen die de huisartsen hadden konden worden gegroepeerd in overwegingen die betrekking hadden op de werkhypothese (o.a. of de werkhypothese een specifieke ziekte betrof en de inschatting van de pretest kans op ziekte) en overwegingen die betrekking hadden op de laboratoriumtests (o.a. de onbetrouwbaarheid van referentiewaarden, eerdere laboratoriumuitslagen en de hoogte van een afwijking). Deze overwegingen speelden een belangrijke rol als de werkhypothese en de laboratoriumtests niet in overeenstemming waren zoals wanneer er licht afwijkende uitslagen werden gevonden terwijl de huisarts dacht dat er geen ziekte was. In deze situaties werd er veel gerefereerd aan overwegingen als de onbetrouwbaarheid van referentiewaarden. De huisartsen probeerden daarmee de afwijkende uitslagen in te passen bij de werkhypothese en ze plaatsten afwijkende uitslagen zo in het perspectief van het totale klinische beeld. Hierdoor leidden afwijkende uitslagen bij een lage pretest kans meestal niet tot een verandering van werkhypothese.

Hoofdstuk 4

Het interpreteren van afwijkende laboratoriumuitslagen kan lastig zijn bij een lage pretest kans op ziekte. Met name licht afwijkende laboratoriumuitslagen kunnen in zo'n situatie moeilijk te interpreteren zijn, terwijl dit vaak makkelijker is voor flink afwijkende laboratoriumuitslagen. Hoewel het bekend was dat er veel afwijkende waarden worden gevonden in de huisartspraktijk, was het onbekend hoe vaak dit licht danwel flink afwijkende uitslagen betrof. Ook was de relatie tussen het aantal afwijkingen en de pretest verwachtingen niet bekend. Bij 1775 patiënten van 87 verschillende huisartsen werd daarom onderzocht hoe vaak er afwijkende en flink afwijkende laboratoriumuitslagen waren en hoe de relatie was met de pretestverwachtingen.

Er werden gemiddeld 9,9 tests gerapporteerd als de huisarts ziekte wilde uitsluiten, 8,9 tests als de huisarts wilde gerust stellen en 7,9 tests als de huisarts wilde monitoren. Van al deze tests was 21% afwijkend en 5% flink afwijkend. Het merendeel van de patiënten had op zijn minst één afwijkende test (64%). Een minderheid, maar toch nog een aanzienlijk deel, had op zijn minst één flink afwijkende test (21%). Zelfs als de geschatte pretest kans op een ziekte gering was, of als het onderzoek was aangevraagd om gerust te stellen, werden veel afwijkingen gevonden. Bij geruststelling had ongeveer de helft van de patiënten op zijn minst één afwijkende tests en ongeveer één op de tien patiënten had op zijn minst één flink afwijkende test. Afwijkende uitslagen komen dus heel vaak voor en zelfs bij een lage pretest kans worden veel afwijkingen gevonden. Door de statistische wijze waarop referentiewaarden worden gedefinieerd is er echter een redelijke kans dat veel afwijkingen geen klinische relevantie hebben en is het een vraag voor verder onderzoek welke consequenties deze uitslagen hebben voor de diagnose en het verdere beleid.

Hoofdstuk 5

Ondanks dat huisartsen vaak laboratoriumdiagnostiek aanvragen als de pretest kans op ziekte gering is zijn toch veel laboratoriumuitslagen afwijkend. Huisartsen kunnen deze afwijkingen interpreteren door ze in het perspectief te plaatsen van pretest verwachtingen als de reden om tests aan te vragen en de geschatte pretest kans op ziekte, maar de mate waarin deze pretest verwachtingen een rol spelen bij het interpreteren van afwijkende uitslagen was onbekend. Daarom werd deze relatie tussen het interpreteren van de uitslagen en de pretest verwachtingen onderzocht.

Er waren 87 huisartsen die deelnamen en zij vroegen voor 1253 patiënten laboratorium tests aan. Met behulp van kruistabellen, de chi-kwadraat toets en een Multivariaat logistisch regressie model werden de analyses uitgevoerd. De resultaten van de kruistabellen lieten zien dat afwijkende tests significant vaker als normaal werden beoordeeld als ze waren aangevraagd bij een lage pretest kans op ziekte ten opzichte van een hoge pretest kans op ziekte (66.1 versus 19.6%). Ze werden ook significant vaker als normaal beoordeeld als de tests waren aangevraagd om een ziekte uit te sluiten ten opzichte van tests die waren aangevraagd om een ziekte aan te tonen (50.9 versus 27.7%). Deze relatie was ook in het multivariate model significant. Het multivariate model liet ook zien dat als tests als afwijkend werden geïnterpreteerd dat er dan een significant hogere kans was op een hoog geschatte post-test kans. Een hoge posttest kans gaf een significant hogere kans op een actief beleid.

Deze resultaten laten dus zien dat huisartsen de pretest verwachtingen een belangrijke rol toedichten bij het interpreteren van de uitslagen en de keuze voor het verdere beleid. Dit zal over het algemeen een goede strategie zijn, omdat veel afwijkende uitslagen klinisch niet relevant zijn en huisartsen zo een balans kunnen zoeken tussen het over- en onderreageren op uitslagen van laboratoriumtests.

Hoofdstuk 6

Veel aanvragen van röntgenfoto's van de wervelkolom zijn volgens richtlijnen medisch gezien niet noodzakelijk. Een van de redenen dat huisartsen deze foto's toch aanvragen is om de patiënt gerust te stellen. De vraag is welke consequenties de uitslagen van zulke foto's hebben voor het verdere beleid en of de huisarts hierbij rekening houdt met de reden dat de foto's werden aangevraagd.

Er waren 90 huisartsen die aan dit onderzoek meededen. Op het aanvraag formulier voor de foto's vulden de huisartsen de reden in die ze hadden om het onderzoek aan te vragen. Als de uitslag werd teruggestuurd vulden ze een aantal vragen in over het beleid dat ze verder gingen volgen. Met een multivariaat model werd de relatie onderzocht tussen de reden om aan te vragen en het beleid dat werd gevolgd, waarbij rekening gehouden werd met de uitslagen van de foto's, de differentiaal diagnose en het geslacht en de leeftijd van de patiënt.

De uitslagen lieten bij 36% van de patiënten geen afwijkingen zien, bij 48% milde degeneratieve afwijkingen en bij 17% waren er abnormale bevindingen. Als er op de foto geen afwijkingen werden gevonden dan voerde de huisarts bij twee van de drie

patiënten een afwachtend beleid. Waren er milde degeneratieve afwijkingen, dan koos de huisarts bij de helft van de patiënten voor een afwachtend beleid. Van de röntgenfoto's was 15% aangevraagd om de patiënt gerust te stellen. De reden om de röntgenfoto's aan te vragen had een significante invloed op het beleid: als de foto's waren aangevraagd om gerust te stellen dan was er, gecorrigeerd voor de uitslag van de foto, 3 keer minder vaak een actief beleid, zoals een verwijzing, verder onderzoek of medicatie.

Hoofdstuk 7

Door de statistische wijze waarop referentiewaarden voor uitslagen van laboratoriumonderzoek worden gedefinieerd kunnen er ook bij gezonde personen afwijkende waardes worden gevonden. Hierdoor kan een initiële afwijkende uitslag onterecht leiden tot een proces waarbij steeds meer onderzoek wordt aangevraagd en verwijzingen volgen, ook wel een cascade proces genoemd. Het was niet bekend hoe vaak zo'n proces voorkomt in de huisartspraktijk en wat de invloed is van determinanten als de geschatte pretest kans op ziekte en de interpretatie van afwijkende uitslagen.

Om dit te bestuderen werden 142 patiënten met een lage pretest kans op ziekte én een afwijkende uitslag gematcht op leeftijd en geslacht met 142 patiënten met een lage pretest kans op ziekte en een normale uitslag. Er werd geanalyseerd hoe vaak er in beide groepen vervolg onderzoek werd aangevraagd en hoe vaak er een verwijzing plaatsvond. Met de chi-kwadraat toets werd gekeken of het aantal aanvragen voor onderzoek en het aantal verwijzingen beïnvloed werden door de redenen om laboratoriumonderzoek aan te vragen, de geschatte posttest kans op ziekte en de interpretatie van de uitslagen.

Voor de groep patiënten met een afwijkende uitslag bleek dat na ontvangst van afwijkende laboratoriumuitslagen voor slechts 7 (5,5%) patiënten meer dan één verwijzing of meer dan één aanvraag voor extra onderzoek plaatsvond, met een maximum van 3 verwijzingen en extra onderzoeken. Bij 3 van de 7 patiënten waren de initiële laboratoriumuitslagen fors afwijkend en rechtvaardigden deze verder onderzoek. Bij de andere 4 patiënten waren de initiële laboratoriumuitslagen slechts licht afwijkend. Cascade processen na laboratorium onderzoek komen dus slechts zeer weinig voor en zijn beperkt in omvang.

Hoofdstuk 8

De resultaten van de verschillende onderzoeken uit dit proefschrift kunnen bediscussieerd worden aan de hand van twee theorieën, medical decision making en illness script theorie, die vanuit een verschillend perspectief het interpreteren van uitslagen benaderen. De resultaten van de verschillende onderzoeken in dit proefschrift lijken vooral goed aan te sluiten bij de illness script theorie.

De illness script theorie beschrijft hoe artsen in de dagelijkse praktijk diagnostische gegevens interpreteren. Meestal ontwikkelen artsen al vroeg in een consult een

hypothese, zo'n hypothese wordt een illness script genoemd. Dit illness script herbergt alle eigenschappen van een ziekte, zoals ondermeer de uitslagen van diagnostische tests die passen bij het betreffende ziektebeeld. De werkelijke uitslagen van diagnostische tests worden vervolgens gespiegeld aan de waarden die verwacht zouden worden volgens het illness script. Als de waardes erbij passen dan wordt het script geaccepteerd, past het niet dan wordt het script afgewezen.

Uit ons onderzoek blijkt dat huisartsen het script toch kunnen accepteren als de waarden slechts marginaal afwijken. Er zijn dus ook waardes die weliswaar niet passen bij het script, maar die toch niet sterk genoeg zijn om het af te wijzen. Zulke waardes worden door huisartsen waarschijnlijk als acceptabel beschouwd. Er wordt verondersteld dat dit komt doordat waardes die niet overeenkomen met een script vaak veel minder sterk zijn in het afwijzen ervan, dan de mate waarin de waardes die wel overeenkomen het script bevestigen. Dit lijkt goed aan te sluiten bij de bevindingen in dit proefschrift: Als de huisarts geen vermoeden had op ziekte, dan werden veel afwijkende uitslagen toch als normaal geïnterpreteerd en nam het vermoeden op ziekte niet toe.

Medical decision making beschrijft hoe de posttest kans op ziekte te berekenen is uit de pretest kans op ziekte, de test uitslag en de testkarakteristieken. Deze berekeningen zijn echter lastig omdat de grootte van de pretest kans vaak niet goed bekend is en omdat de huisartsen vaak een set van tests aanvragen in plaats van een test. Ze worden in de dagelijkse praktijk dan ook amper toegepast. De theorie helpt wel om bepaalde fouten bij het interpreteren van uitslagen beter te begrijpen doordat het wijst op het belang van het goed kunnen inschatten van pretest kansen en het goed kunnen inschatten van de voorspellende waarde van een diagnostische test.

Onderwijs over het interpreteren van uitslagen vindt vaak plaats vanuit het perspectief van de medical decision making theorie. Maar de theorie over illness scripts sluit wellicht beter aan bij hoe uitslagen in de dagelijkse praktijk geïnterpreteerd worden. Meer aandacht in het onderwijs hiervoor kan studenten en artsen helpen test uitslagen beter te leren interpreteren.

De resultaten in dit proefschrift wijzen erop dat huisartsen in het algemeen goed om kunnen gaan met uitslagen van tests die zijn aangevraagd bij lage pretest kansen of om de patiënt gerust te stellen. We vonden dat huisartsen doorgaans geen probleem hebben met het interpreteren van uitslagen en dat er weinig cascades aan diagnostiek optreden. Dit kunnen redenen zijn dat huisartsen toch diagnostiek blijven aanvragen in situaties waar richtlijnen het gebruik hiervan ontraden.

Aanbevelingen voor verder onderzoek in dit proefschrift spitsen zich toe op drie belangrijke gebieden: de verwerking van testuitslagen in de praktijk, het interpreteren van testuitslagen en hoe goed patiënten de betekenis van testuitslagen begrijpen. Elk gebied heeft zijn eigen specifieke vragen waar toekomstig onderzoek zich op kan richten en die mogelijkheden bieden de kwaliteit van zorg rondom het gebruik van diagnostiek in de dagelijkse praktijk te verbeteren.

Appendix

Questionnaire

5.b. Do you suspect an innocuous somatic disorder?

- Definitely not
- Probably not
- Maybe
- Probably yes
- Definitely yes

5.c. Do you suspect a serious somatic disorder?

- Definitely not
- Probably not
- Maybe
- Probably yes
- Definitely yes

6. What is currently your most probable diagnosis or working hypothesis?

.....

7.a. Are you considering any other diagnoses? No Yes

7.b. If so, which one(s)?

- 1.
- 2.
- 3.
- 4.

Management

8. What will be your management until the test results become available? Please indicate the main components of your management, by ticking one box in the first column and one in the second.

	policy 1	policy 2
1. Reassuring, explaining	<input type="radio"/>	<input type="radio"/>
2. Expectative, wait-and-see	<input type="radio"/>	<input type="radio"/>
3. Advice (on lifestyle, diet, complaints, etc.)	<input type="radio"/>	<input type="radio"/>
4. Instructions (when to contact me, etc.)	<input type="radio"/>	<input type="radio"/>
5. New/follow-up appointment (telephone, consultation, home visit, etc.)	<input type="radio"/>	<input type="radio"/>
6. Medication (start, stop, adjust)	<input type="radio"/>	<input type="radio"/>
7. Referral or consultation (specialist, therapist, etc.)	<input type="radio"/>	<input type="radio"/>
8. Other, namely	<input type="radio"/>	<input type="radio"/>

Thank you very much for cooperating with this study!

General data

1. *Current date?* ... - ... - 20 ...
2. *Patient's date of birth?* ... - ... -
3. *Patient's sex?* male female

Please read this!

When answering the questions below, please disregard any test results relating to check-ups for the patient's known conditions or to screening for hypertension, diabetes or cholesterol levels. If the test results only relate to these items, please do not continue completing this questionnaire.

Test result

4. *How would you interpret this lab test results for this patient?* Please tick one box. If you prefer to tick multiple boxes please select the most relevant one.

I interpret this test results for this patient as:

- Clearly abnormal
- Possibly abnormal
- Normal

The diagnosis

The following questions relate to all diagnoses you are now considering for this patient.

- 5.a. *Do you now suspect a somatic disorder?*

- Definitely not
- Probably not
- Maybe
- Probably yes
- Definitely yes

- 5.b. *Do you now suspect an innocuous somatic disorder?*

- Definitely not
- Probably not
- Maybe
- Probably yes
- Definitely yes

- 5.c. *Do you now suspect a serious somatic disorder?*

- Definitely not
- Probably not
- Maybe
- Probably yes
- Definitely yes

Publications

Publications

1. Houben PHH, van der Weijden T, van Bokhoven MA, Droog A, Winkens RA, Grol R. Overwegingen van huisartsen bij het interpreteren van laboratoriumonderzoek; een kwalitatief onderzoek. *Huisarts Wet* 2005;48:326-32.
2. Houben PH, van der Weijden T, Sijbrandij J, Grol RP, Winkens RA. Reasons for ordering spinal x-ray investigations: how they influence general practitioners' management. *Can Fam Physician* 2006;52:1266-7.
3. van Bokhoven MA, Houben P, Chavannes N, Nijholt A. The white man has big eyes but does he see? Students' experiences and background information to prepare yourself for an elective in a developing country. Maastricht: Universiteit Maastricht. 2006.
4. Houben PH, van der Weijden T, Winkens RA, Grol RP. Cascade effects of laboratory testing are found to be rare in low disease probability situations: prospective cohort study. *J Clin Epidemiol*. 2010;63:452-8.
5. Stolper E, van Bokhoven M, Houben P, Van Royen P, van de Wiel M, van der Weijden T, et al. The diagnostic role of gut feelings in general practice. A focus group study of the concept and its determinants. *BMC Fam Pract* 2009;10:17.
6. Stolper E, Van Royen P, Van de Wiel M, Van Bokhoven M, Houben P, Van der Weijden T, Dinant GJ. Consensus on gut feelings in general practice. *BMC Fam Pract* 2009;10:66.
7. Houben PH, van der Weijden T, Winkens B, Winkens RA, Grol RP. Pretest expectations strongly influence interpretation of abnormal laboratory results and further management. *BMC Fam Pract* 2010;11:13.
8. Houben PH, Winkens RA, van der Weijden T, Vossen RC, Naus AJ, Grol RP. Reasons for ordering laboratory tests and relationship with frequency of abnormal results. *Scand J Prim Health Care* 2010;28:18-23.
9. Stolper E, van Leeuwen Y, van Royen P, van de Wiel M, van Bokhoven M, Houben P, Dinant GJ. Establishing a European research agenda on 'gut feelings' in general practice. A qualitative study using the nominal group technique. *Eur J Gen Pract* 2010;2.
10. Stolper E, van Bokhoven M, Houben P, Van Royen P, van de Wiel M, van der Weijden T, Dinant GJ. Het 'pluis/niet-pluis'-gevoel van de huisarts. *Ned Tijdschr Geneesk* 2010;154:A1451.

Dankwoord

Dankwoord

Snowboarden is gaaf, maar vooral in het begin maakte ik nogal eens een flinke smak. Nu, eenmaal onder de knie, is het echter heerlijk om op m'n bord door de sneeuw te glijden. Promoveren was minder spannend, maar het had wel wat weg van snowboarden: af en toe vallen, soms hard. Gelukkig stond ik meestal redelijk snel weer op en zijn zware blessures uitgebleven. Langzaam aan leerde ik steeds meer, en nu is er een proefschrift.

Ik heb mijn weg gevonden op de wetenschappelijke piste, maar dat heb ik niet alleen gedaan. Ik promoveer dankzij de hulp en steun van velen. Iedereen die me afgelopen jaren geholpen heeft wil ik hiervoor hartelijk bedanken.

In de eerste plaats wil ik Trudy van der Weijden bedanken, mijn (co)promotor. Je bent het meest betrokken geweest bij mijn begeleiding. Aan onze samenwerking heb ik veel goede herinneringen en ik bewonder je toewijding en inzet. Je hebt veel tijd gestoken in dit project. Het contact met je was persoonlijk en warm. Als ik kritisch was en twijfelde, dan lukte het je meestal om me op te beuren. Je positieve commentaar gaf nieuwe energie. Naast de voortgang van het project, lette je ook op mijn ontwikkeling in onderzoeksvaardigheden door bijvoorbeeld te zorgen voor een gedetailleerd cursusplan. Met alle vragen kon ik bij je terecht, je was altijd te bereiken voor hulp. Bedankt voor al je steun!

Richard Grol was mijn promotor. Internationaal geprezen. Hoofd van het Scientific Institute for Quality of Healthcare. En desondanks toch nog tijd om mij te begeleiden. Je hebt een belangrijk aandeel gehad in het tot stand komen van dit proefschrift. Je deskundigheid was voor mij onmisbaar. Ik bewonder vooral je vermogen om kritisch te kunnen kijken naar de hoofdlijnen. Bijvoorbeeld, in de opbouw van een inleiding of discussie wist je precies de manco's te benoemen. Jouw commentaar en hulp hebben me erg geholpen bij het leren schrijven van wetenschappelijke artikelen.

Ron Winkens was copromotor. Je vond altijd ruimte voor dit project ondanks je vele taken en verantwoordelijkheden als huisarts die ook onderzoek doet. Ik vond het fijn samenwerken met je, je optimisme was voor mij een hele steun. Je ervaring in onderzoek naar diagnostiek en je klinische inbreng als huisarts stonden aan de basis van enkele stukken uit dit proefschrift.

Karin Vaessen, Paddy Hinssen en Marie-Louise Dumont ondersteunden allerlei werkzaamheden van het project. Ze hebben ook geweldig werk verricht door alle laboratoriumuitslagen van bijna 1800 patiënten in SPSS in te voeren. Dat was een geweldige klus die ik alleen nooit had kunnen doen en ik ben jullie dan ook erg dankbaar. Jildou Sijbrandij en Bjorn Winkens wil ik bedanken voor hun deskundige

statistische inbreng. De klinisch chemici André Naus en Renéé Vossen bedank ik voor hun hulp bij het artikel over referentiewaarden en afwijkende uitslagen. Inge, bedankt voor de mooie omslag die je hebt verzorgd. Jan Klerkx zorgde voor de Engelse vertalingen en correcties, Tiny Wouters heeft de lay-out verzorgd. Bedankt!

Voor de huisartsen die deelnamen aan het onderzoek verdienen lof. De dagelijkse praktijk van veel huisartsen is al zo druk en patiëntenzorg en andere verplichtingen eisen veel aandacht. Daar komt de tijd en energie die het kost om patiënten te includeren en vragenlijsten in te vullen nog eens bovenop. Deze huisartsen ben ik dan ook echt dankbaar voor hun inzet bij onze dataverzameling. De namen van deze huisartsen staan aan het eind van dit dankwoord vermeld.

Naast hen hielpen ook veel anderen mee aan de dataverzameling: dhr. Kusters, dhr. Pennings, GERALINE Leusink, Hans Vlek, Hans van Duijnhoven, Dirk Wijkel, Rob Jansen, Wim Verstappen, Renéé Vossen, Frank Somers en alle medewerkers van de volgende lokaties: het Diagnostisch Centrum in Eindhoven en de laboratoria van het Jeroen Bosch Ziekenhuis in 's Hertogenbosch, het St. Jans Gasthuis in Weert, het Elkerliek ziekenhuis in Helmond, het St. Anna ziekenhuis in Geldrop, het Maasland Ziekenhuis in Sittard en het AtriumMC in Heerlen: allen bedankt voor jullie hulp!

Geert-Jan Dinant wil ik bedanken voor alles wat hij heeft gedaan voor de Aiotho's. Dankzij hem was het mogelijk de opleiding tot huisarts te combineren met promotieonderzoek. Daarnaast leidde hij de Primary Health Care groep (PHC). Voor de PHC begeleidde ik tal van studenten voor hun stages in het buitenland. Ik kreeg de kans verder te gaan in het studentenonderwijs dankzij Paul Zwietering. Daar ben ik hem dankbaar voor want ik heb nu veel plezier in mijn nieuwe baan als docent bij de VU.

De laatste jaren deelde ik met veel plezier de kamer met Roelf en Annemiek. We spraken over de toppen en dalen van ons onderzoek. We bezochten elkaars praktijken, gingen samen eten en we hebben heel veel gelachen. Ik heb veel goede herinneringen aan jullie: Roelf dacht diep na, dan was het stil. Als ik dacht dat hij me niet had gehoord en ik weer verder ging werken, kwam er opeens nog een antwoord. Met Annemiek was het altijd gezellig. We hebben veel gedaan voor de PHC, zoals het boek dat we samen met Niels en Loes hebben geschreven. Loes was begeleidster van mijn wetenschapsstage eind 2000. Dankzij jou ben ik onderzoek gaan doen. Afgelopen jaren heb ik veel en fijn met je samengewerkt: onze projecten lagen inhoudelijk dicht bij elkaar. Babette Doorn en de andere redactieleden wil ik bedanken voor het plezier dat ik had bij het maken van Op Eén Lijn.

Naast onderzoeker was ik ook huisarts in opleiding. Mijn beide huisartsopleiders wil ik bedanken. Bij Rob de Jonge deed ik mijn eerste ervaringen op als huisarts. Met al je

kunde heb je mij, jong en erg onervaren, goed geïntroduceerd in de huisartsgeneeskunde. Paul Schaffer was mijn opleider in het derde jaar. Ik heb veel bewondering en respect voor je vaardigheden als huisarts en opleider. Je hebt me een spiegel voorgehouden waardoor ik dichter bij mezelf kwam en waardoor ik heb geleerd mijn gevoel meer te laten spreken. Daar ben ik ontzettend blij mee, nog steeds iedere dag als ik werk in de praktijk. Ik ben jullie dankbaar voor de flexibiliteit waardoor het mogelijk was onderzoek en opleiding te combineren.

Sinds 2006 werkte ik als huisarts samen met Johan de Paepe. Je bent een fantastische kerel: je geduld, je vermogen om te luisteren en de goede band die je met je patiënten onderhoudt. De praktijk met jou en Marja, Anita en Lenie was laagdrempelig en persoonlijk voor onze patiënten. Het verliep wel eens chaotisch, maar ik heb met veel plezier met jullie samengewerkt. Jullie gaven mij m'n eigen plek. We hebben vaak en veel gelachen en we hebben veel plezier gehad in Glasgow en de Eifel. Aan onze samenwerking bewaar ik mooie herinneringen. Bedankt!

Ons groepje Aiotho's in Maastricht was een plezierige club: Katinka, Rogier, Niels, Annemiek, Roelf, Loes, Patrick, Marjolein en Wemke. We deden samen de postgrade cursussen, hadden een maandelijkse refereer club en gingen op het werk en privé veel met elkaar om. Deze club gaf elkaar veel steun. Sinds 2006 doe ik supervisie met Koen, Ilse, Loes, Katinka en Felicitas. We hebben Marie-José als begeleidster. De supervisie helpt me gestructureerd te reflecteren op mezelf. Het is voor mij een belangrijke hulp om een goede en gelukkige huisarts, onderzoeker en docent te zijn.

Mijn allerliefste, beste, oude en nieuwe vrienden waren een heel belangrijke steun. In al die jaren hebben we zoveel ondernomen: skivakanties, vakanties in Barcelona, Montreal en Stockholm; weekendjes weg in Brussel, Dublin en Leipzig; zeilen, sporttoernooien, uitgaan, etc. etc. Het plezier en de ontspanning gaven de energie die nodig was om dit proefschrift te schrijven. Mijn familie en schoonfamilie bedank ik voor de onvoorwaardelijke steun en alle liefde en warmte. Pap en Mam, dankzij jullie voel ik me altijd sterk. Twee voeten op de grond, stabiel en gelukkig. Bedankt! Jurgen en Mark weten wat het is om te promoveren, Pauline wilt het nog ontdekken. Moniek, Pa en Ma, Astrud, Erik, Petra en Erwin, wat zijn jullie hartverwarmend!

Lieve Frank, je bent al die tijd mijn grootste en belangrijkste steun geweest. Niets is zo fijn als samen zijn, we gaan een mooie toekomst tegemoet. En nu het proefschrift klaar is moeten we maar eens snel in ons bootje stappen en gaan varen op de Amsterdamse grachten.

De huisartsen die meewerkten aan het onderzoek:

In de regio Eindhoven: Dhr. P Meulesteen, Dhr. PJM Bertholet, Dhr. CGJ Dekkers, Dhr. GRI Slock, Mw. HJC Becx, Dhr. PJ Wouda, Mw. PJ Dalinghaus-Nienhuys, Dhr. LWF Maartens, Mw. AAM Schoenmaker-Willemse, Dhr. JJ Dijkmans, Dhr. JGM Moors, Dhr. H Dierick, Mw. AEM Lucas, Mw. M Gilissen, Mw. F Wouters, Dhr. HJ van Steenbergen, Mw. M Mateijssen, Dhr. MLF Klomp, Mw. G van der Veen.

In de regio 's-Hertogenbosch: Dhr. D Engelen, Mw. PFJM Nieuwenhuis, Mw. M Langenberg-Ledeboer, Dhr. JHAH Ferrée, Dhr. JJCM Versteeg, Dhr. JA Houwen, Mw. MGM Peeters.

In de regio Helmond: Dhr. CA de Kock, Dhr. V Kaiser, Mw. A Breekveldt, Mw. MJ Geraedts, Dhr. WJ van Geldrop, Dhr. APAM Leijte, Dhr. FJM Raaymakers, Dhr. PAHM Verbruggen, Dhr. VM van Roessel, Mw. MWEJ van Arendonk, Dhr. CPWJ Nieuwenhof, Dhr. WFSM Suijs, Dhr. AM Engels.

In de regio Geldrop: Dhr. RH Veenstra, Dhr. A Mohamedhoesein, Dhr. FPM van Gerven, Mw. PLML Gruyters, Dhr. AP Nederstigt, Dhr. J Kroon, Dhr. GJW Geboers, Mw. M Zwartendijk, Dhr. JWM van den Boomen.

In de regio Weert: Dhr. DPBM Snel, Dhr. JJM Takken, Dhr. RCC Weijnen, Dhr. M Cornel, Dhr. W van Hapert, Dhr. E Heemstra, Mw. CM Husaarts, Mw. MEW Kessels, Dhr. JSM Koopmans, Dhr. PJ Meurs.

In de regio Sittard: Dhr. HWPM Bergmans, Dhr. HMJ Boesten, Mw. Van der Busse, Dhr. JPH Dolhain, Dhr. MAMF Eyck, Mw. I van der Heijden, Dhr. K Heijmans, Dhr. MJFM Klaassen, Mw. CME Liedekerken, Dhr. CJGM Pustjens, Mw. JMO Sikkes-Mekel, Dhr. ATJ Verijdt, Dhr. CJ Vincentie.

In de regio Heerlen: Dhr. P de Bruyckere, Mw. S de Vree, Dhr. D van der Wissel, Mw. G van Zanten, Dhr. F Donders, Dhr. F Guldmond, Dhr. G Habets, Mw. ML Hermans-Bloemen, Dhr. D Kuijpers, Dhr. W Niessen, Dhr. F Soomers, Dhr. D Kronjee, Dhr. H van der Ploeg, Mw. M van Vliet, Dhr. P Bots, Mw. B Bruls.

ALLEN BEDANKT!

Curriculum Vitae

Curriculum Vitae

Paul Houben was born in Horst, The Netherlands, on May 7, 1976. He graduated in 1994 from secondary school at Van Maerlant Lyceum in Eindhoven. That same year he started his medical studies at Maastricht University. He was actively involved in student life as well as student committees. He spent an elective in community medicine at the University of Belo Horizonte in Brazil. In 2000 he graduated as medical doctor.

In 2001 he started as a researcher at Maastricht University, department of general practice, School for Public Health and Primary Care (CAPHRI). He also worked as a resident for a chronic low back pain rehabilitation project at Revalidatiecentrum Hoensbroek. In 2002 he began vocational training in general practice in Eindhoven and he combined this with his research work. In 2004 he was elected 'best junior researcher' at the Wonca Europe congress in Amsterdam (Wonca: World Organization of Family Doctors). He has actively supervised numerous medical students' research electives and electives in developing countries.

In 2006 he started working as a general practitioner (GP) in Eindhoven. In 2009 he moved to Amsterdam. He now combines his work as a GP with his work as a teacher for medical students at the department of general practice, VU University Amsterdam. Paul happily lives together with his friend Frank Ebbink.