Ned Tijdschr Klin Chem 2001; 26: 236-242

Thema Besliskunde

Evidence Based Laboratory Medicine: linking tests to outcomes

C.P. PRICE

There are challenges surrounding the appropriate use of diagnostic tests. There is the fundamental issue regarding the type and quality of evidence required to demonstrate clinical and economic effectiveness. Central to this issue is ensuring that studies address the right question and that the study design is robust with measurement of appropriate outcomes for the clinical questions being asked. Several systematic reviews to date indicate that these requirements are not being met. However, there is also an important constraint in study design because the outcome that one wishes to measure is dependent not only on the use of the diagnostic test but also on the implementation of an action plan, eg therapeutic intervention that follows from receipt of the test result. These points are discussed, identifying some of the outcome measures that can be used to judge effectiveness of a diagnostic test.

Key-words: diagnostic test; decision making; clinical outcome; economic outcome; evidence; effectiveness

There are many challenges facing laboratory medicine today, which offer both opportunities and constraints. Many of the opportunities derive from the wealth of basic and applied research in clinical medicine which enhances our knowledge of the pathophysiology of diseases and begins to define health and wellbeing. This base of knowledge also leads to the development of new therapies and the recognition of new disease markers. The endeavours of the analytical and manufacturing scientists then produces innovative analytical methods and devices that enable the marker to be used for the screening, diagnosis and management of disease. The activities of the diagnostics and pharmaceutical industries ensure a constant supply of new and innovative technologies.

It might be argued that the changes seen in population demographics with an ageing population, with morbidity and mortality statistics affected by other factors such as changes in lifestyle, also provides an opportunity. However, in that these changes can also increase the demand on the laboratory services they may be considered as a constraint.

Similarly changes in clinical practice might be seen as an opportunity – in, for example, a greater use of point of care testing. On the other hand there is no doubt that changes in clinical practice with the advent of more rapid triage strategies, more one-stop clinics and the evolution of ambulatory care and diagnostic facilities, will place greater demands on the laboratory services, with significant reduction in result turnaround times required. Changes in clinical practice are also associated with a constant drive to improve the quality of services, and outcomes, for the patient.

There is no doubt that the overall increase in demand for laboratory services is a major constraint when viewed in the context of limited, and sometimes capped, resources. It is uncommon for service level agreements between purchasers and providers to have the level of sophistication to take account of changes in workload. Consequently, an increasing workload is often reflected in improved productivity of equipment and staff. This in turn drives down the cost per test – in some respects a laudable outcome, but inevitably leading to cost constraints and limiting the opportunity for investment in the development of new tests (1, 2).

However, of far greater concern are the general perceptions of laboratory testing. They can be broadly described in three categories – too much testing, too expensive and with a limited impact on outcomes. Whilst it is recognised that workloads in most laboratories have risen dramatically over the last two decades – in part with the increasing use of automated instrumentation, there is little evidence of a correlation with increasing clinical activity and improved outcomes. That said, there are specific examples where the introduction of a test has improved the clinical outcomes – albeit the evidence is not always systematic. There have been a number of studies published where investigators have looked at the level of testing and attempted to link requesting to outcomes, without much success (3-5). This is partly because of the absence of good experimental design and the lack of appropriate and measurable outcomes. Certainly the high proportion of "normal results" being an indication of inappropriate testing is totally unacceptable as a premis. The absence of good quality evidence

Department of Clinical Biochemistry, St Bartholomew's and The Royal London Hospitals, London

Address correspondence to: Prof. C.P. Price, Department of Clinical Biochemistry, Barts and the London NHS Trust, Turner Street, London E1 2AD, UK
e-mail: c.p.price@mds.qmw.ac.uk

Table 1. Examples where the diagnostic test and the intervention are inextricably linked with patient outcome

Test	disease/procedure	patient outcome
Blood glucose	diabetes mellitus	delay onset of complications
Serum phenytoin	epilepsy	reduced incidence of convulsions
Serum parathyroid hormone	intraoperative parathyroidectomy	reduced re-operation rate
Blood prothrombin time	warfarin therapy	reduced risk associated with poor anticoagulant status
Urine albumin excretion	hypertension	reduced risk of cardiovascular disease

has been recognised and the call for an outcomes agenda for diagnostic testing has been made (7-9).
The true cost of provision of diagnostic services is an extremely difficult issue to address and it depends on a number of factors. The cost comprises a number of elements, including the staff, the estate and the raw materials. All managers of a service are required to ensure that provision is cost effective and in so doing tend to resort to comparative measures – comparison with last years costs and comparison with a neighbours costs. These forms of benchmarking provide useful management tools but tend to provide a limited perspective – focusing on the resource allocation for only one element of healthcare provision rather than taking an holistic view of the completed patient episode.
The greatest concern however has to be the perception that the diagnostic test has a limited impact on the final patient outcome (10). A review of the "patient consultation", whether it be in the primary or acute care setting, begins with a problem – the patient presenting with symptoms, and a question (albeit in many cases not explicitly asked and not always understood). The clinician may use a diagnostic test to answer the question, to which the answer will determine the action that is taken. The action that is taken, which may involve an intervention, will determine the outcome; it will resolve the problem that sent the patient to the doctor in the first instance. In many respects it is therefore understandable that the intervention is seen as providing the key to a successful outcome (11). Yet there are many examples that demonstrate the importance of both the diagnostic test and the intervention in determining the final patient outcome; some are given in Table 1.

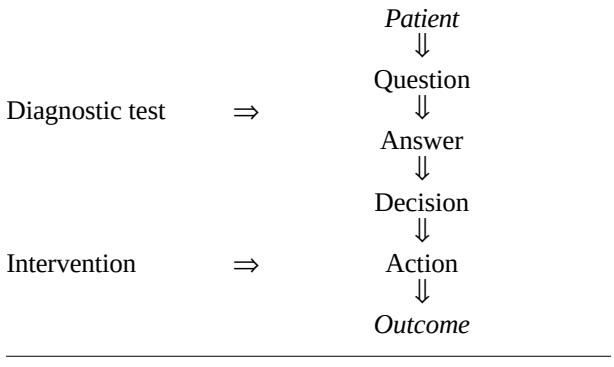

Figure 1. The "clinical episode"

The role of the diagnostic test
An evidence based approach to the practice of medicine has been described as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of patients" (12). The diagnostic test is an important tool for the clinician making a decision about a patient. The role of the diagnostic test depends on the relevance to the problem or question at hand and the success with which the result provides a solution to the problem or answer to the question.
Herein lie the clues to the debate on the appropriateness of diagnostic tests and ultimately the justification, or otherwise, to the increasing workload seen in many laboratories. It is then by auditing the use of the diagnostic services that it will be possible to identify whether utilisation is appropriate and then to instigate an education programme if that is found not to be the case.
The crux of the matter is therefore the identification of the question that the clinician poses when presented with a patients' situation. The appropriateness of a test is therefore defined as being when the result of that test provides an answer to the question. However, the question must be relevant; the relevance is determined by the fact that the answer to the question enables a decision to be made and an action taken. Only the complete process will generate the outcome (Figure 1).
The nature of the question will vary according to the patient, to the severity of the symptoms, to the clinical setting and to the action that may be taken. As a consequence the choice of test may depend on all of these circumstances and require a clear understanding of the question that is being asked. It is therefore likely that the performance of a test will depend on all of these issues – and vary from setting to setting. It is quite obvious that the diagnostic performance of a test will depend on the way in which the disease evolves and the value of the test will also change in relation to the time at which the questions are asked. Some examples of different questions being asked in a given clinical situation are illustrated in Table 2.
It becomes clear that a number of different questions may be asked – and they may not all be strictly "diagnostic" questions. However, in terms of a decision making process there are clearly questions that "rule in" and questions that "rule out" a particular decision pathway. It is therefore quite possible that a test will perform well as a "rule in" test and not as a "rule out"

Table 2. Examples of clinical questions in various clinical settings that might be answered with the aid of a diagnostic test

Test	setting	question
Troponin	primary care	is this chest pain?
	accident and emergency	what is the risk of a further cardiac event?
	coronary care unit	has reperfusion been successful?
Prostate cancer	primary care	have I got prostate cancer?
	primary care (screening)	should this patient be referred for further tests?
	outpatient clinic	what is the likelihood this patient has prostate cancer?
	oncology unit	has the cancer relapsed?
Chlamydia (molecular test)	primary care (screening)	has this patient a chlamydia infection?
	infertility clinic	is infertility due to chlamydia infection?

test. Some examples of clinical settings in which a "rule in" or "rule out" strategy can be used are given in Table 3. Empirically it can be argued that when a test demonstrates a high sensitivity for detection of disease a negative result may be good for a "rule out" strategy, whereas a positive result for a test with high specificity is required for a "rule in" strategy. If one takes the example of myoglobin in the case of a patient with chest pain, due to its rapid rise following myocardial ischaemia it can provide an early or sensitive indicator of disease, but due to the lack of specificity for myocardial muscle damage it cannot be used as a "rule in" test; however, a negative result may be used as a "rule out" test (13). In this situation the value of the test can be judged on the basis of there being a decision strategy (or pathway) that can take advantage of an effective "rule out" test; in this case it may influence the triage protocol and improve the overall cost or economic outcomes of managing patients with chest pain (14). In the case of patients who may suffer from a urinary tract infection and visit a primary care physician, it has been suggested that simple urine screening tests using the leucocyte esterase and nitrite levels may reduce the number of urine samples referred to the laboratory for further analysis (15).

Therefore, in determining the appropriateness of the use of a test and the effectiveness of the use of that test it is essential that the nature of the question is quite clearly stated and understood – and that there is an appropriate decision pathway leading to an outcome.

Evidence and its quality

There is a growing body of literature that discusses the issues surrounding the quality of evidence as well as reviewing the quality of the evidence currently available on the use of diagnostic tests (16-19). There are two fundamental issues – relevance and quality.

Relevance is critically important because it determines whether the evidence will provide answers to questions posed about the utility of a test. Typically the research literature on any diagnostic test will comprise: a) observations that underpin our knowledge on the basic pathophysiology of a disease; b) observations on a molecule, cell or structure that is uniquely related to a pathological change; c) description of the development and validation of an analytical method; and d) data on the diagnostic performance of the method. The latter may include information on the sensitivity, specificity and predictive value of the test. However, invariably there is little in the literature that directly shows the performance of the test in providing an answer to the question being asked and the impact of this on the final outcome of the process described in Figure 1. The outcomes can be considered in terms of clinical or economic outcomes, the combination having wider societal or health outcomes which may ultimately have implications in determine health policy.

The quality of evidence is clearly determined by relevance; however, quality is a broader issue than merely the correct identification of the question being asked. It is also important that the performance of a test is studied in the population of patients, and in the clinical setting, close to that for which the test is intended.

Experimental design is an important determinant of the quality of evidence, the goal being the minimisation of bias. Some of the key features of good experimental design are listed in Table 4. Reid et al (16) reviewed a large number of papers dealing with the performance of diagnostic tests against a set of design criteria and found that the standards were not met in more than 50 per cent of papers. Lijmer et al (20) in an analysis of 218 studies on the performance of diagnostic tests showed that the use of a case control approach had the greatest impact on bias. Other authors have shown that the use of an imperfect reference method – with the opportunity for greater interoperator variability can also bias the findings (16,21,22). Moore and Fingerova (23) reviewed the

Table 3. Examples of some tests used for "ruling in" and "ruling out" a diagnosis; some tests perform both functions equally well but the actions that follow are different

test	rule in/out	diagnosis
urine leucocyte esterase and nitrite	out	urinary tract infection
serum myoglobin	out	myocardial infarction
blood alcohol	in	confusion due to alcohol intoxication
serum PSA <2.5 µg/l at 50 years	out	prostate cancer
serum digoxin	out	digoxin toxicity

Table 4. Key features of good experimental design that maximise relevance and limits the potential for bias

Explicit and relevant question identified
Relevant patient cohort and clinical setting
Preferable avoidance of case control approach
Adequate number of patients studied
Use of reference standard on patients and controls
Use of a measurable outcome
Operator of test blinded to reference and outcome
Full description of methods used

impact of aspects of study design on the proportion of the effect, pointing out that most of the data came from studies on the efficacy of pharmaceutical products. In the case of a diagnostic test the reference procedure may be difficult to define and, when assessing the impact of a new test on the diagnostic strategy, one may be using a reference procedure or marker for which the pathophysiological basis of the test is identical to the new test being evaluated, eg the use of troponins compared to that of the creatine kinase MB isoenzyme.

It is also acknowledged that any review of the performance of a test may be influenced by publication bias. Several studies have shown that there is a tendency to publish data where statistical significance has been shown; this will positively bias the apparent effectiveness of the test. In addition, multiple publications of data will also result in a positive bias on the true picture; this may unwittingly be achieved by publishing "in stages" as the cohort of patients being studied increases (24-26).

Outcomes

The delineation of the efficacy of a diagnostic test into the diagnostic performance and the impact on the outcome of the diagnostic process was described by Fryback and Thornbury (27) who differentiated between a clinical and an economic outcome. When viewed in the context of a decision making process underpinned by a philosophy that seeks to deliver the highest quality of service to the patient – and whilst also ensuring value for money, their proposals provide a valuable framework for the development of a case for investing (or disinvesting) in a test or procedure (28).

Clinical outcomes

The impact of a diagnostic test on the clinical, or health, outcome can be on the diagnostic strategy or on the therapeutic strategy. The ultimate benefit in the health outcome will be measured in terms of objective statistics, such as morbidity and mortality. It can also be assessed in more subjective ways, such as quality of life, improved functional status and patient satisfaction.

Some examples of diagnostic tests that have been demonstrated to improve clinical outcomes are listed in Table 5. It is worth stressing again that the clinical benefit of a diagnostic test however, will only be apparent if the diagnostic test and therapeutic intervention are implemented in an integrated way; in the case of chronic diseases, such as diabetes mellitus, epilepsy and asthma, where a test is used to monitor therapy, this implies continuing compliance with the complete protocol – test and intervention.

It has become clear that the design of protocols for the evaluation of clinical outcomes for diagnostic tests is complex in two major ways: i) the reliance of the outcome on the intervention; and ii) the length of study required to measure changes in morbidity and mortality. In the latter case the use of surrogate markers has been suggested – it obviously being a requisite that there is a robust link between surrogate and final outcome. Thus bone mineral density measurement has now become the accepted means of defining osteoporosis and therefore a surrogate marker for the assessment of other tests and interventions. Similarly, HbA_{1C} is a surrogate marker for assessing means of improving glycaemic control.

Economic outcomes

The diagnostic test can have an impact on several aspects of the provision of healthcare embraced by the term economic. It can encompass an operational impact, eg an improvement in efficiency, which will in itself impact on cost effectiveness. There may be an impact on the health economy – how resources are dispersed. There is then a wider societal impact from the use of a diagnostic test.

The operational impact can be assessed in terms of the effect on the utilisation of other resources. Some examples of the impact of diagnostic tests on economic outcomes are shown in Table 6. One of the most common examples of an operational impact

Table 5. Some examples of tests where usage has been shown to improve clinical outcomes

Test	disease/procedure	outcome measured
Blood glucose	diabetes mellitus gestational diabetes	delay in onset of complications reduced neonatal complications
Serum phenytoin	epilepsy	reduced incidence of convulsions
HbA_{1C} at point of care	diabetes mellitus	improved glycaemic control and patient satisfaction
Paracetamol	drug overdose	reduced liver damage, improved morbidity and mortality
PT at point of care	warfarin therapy	reduced deviation from target INR
Intraoperative PTH	parathyroidectomy	reduced re-operation rate
Intraoperative ionised calcium	liver transplant	reduced risk of cardiac arrest

Table 6. Examples where the use of a test or testing modality yields an economic benefit

Test	economic benefit
Troponin I	reduced length of stay in chest pain evaluation unit
HbA_{1C} at point of care	reduced average annual visits to diabetes clinic
Phenytoin at point of care	reduced time to reach optimal dosage regime
Urine leucocyte esterase and nitrite	reduced number of urines referred to laboratory
Chlamydia by molecular technique	reduced prevalence of complications, clinic visits and hospital stay
Intraoperative PT and APTT	reduced usage of blood products during coronary artery bypass surgery

today is the attempt to reduce the length of hospital stay or number of clinic visits. Thus measurable means of resource utilisation will comprise the use of staff, estate and consumables. The operational impact of a test may be on the utilisation of intervention resources, eg operating facilities, blood products and drugs. Thus, therapeutic drug monitoring whilst seen as benefiting clinical outcomes can also have an impact on resource utilisation; as an example, Patsalos et al (29) showed that measuring drug levels at the point of care reduced the number of clinic visits and the time taken to optimise therapy in newly diagnosed patients with epilepsy.

Risk assessment (clinical and economic)

It is important when considering a change in practice that due consideration is given to all potential aspects of the change on clinical and economic outcomes. This will be particularly important when using a surrogate or intermediate marker to assess the efficacy of the procedure. Some examples of this in terms of the clinical impact of a test or procedure are: i) the risk of hypoglycaemic episodes when adopting a more aggressive approach to maintenance of normoglycaemia; ii) the risk of a subsequent cardiac event in a patient discharged within 24 hours after admission with chest pain using a rapid rule out strategy; iii) an inability to regularly monitor compliance with drug therapy when there is no readily monitored symptom, eg reduction in pain, improved mobility, etc. Anecdotal evidence suggests that poor compliance with therapy is a feature of many cohorts of patients with chronic diseases and may warrant greater attention where relevant diagnostic tests exist.

The economic risks associated with a change of practice should also be taken into account. One of the greatest risks in terms of economic assessment is due to the limited ability of most healthcare providers (and purchasers) to take an holistic view of the health economics of a disease when planning the provision of services. Typically the resource allocation for diagnostic services is determined on an historical basis, is not directly correlated to demand and is "ring fenced", ie it is allocated and managed in a degree of isolation from the clinical groups which it serves.

Examples of the extreme approaches that have been taken in undoubtedly difficult and complex areas of resource allocation are the comparisons between point of care and laboratory testing approaches to glucose measurement. Invariably, and not entirely unexpectedly, the point of care approach is more costly because it fails to take advantage of the potential for economies of scale available in a centralised laboratory testing approach (30). The benefits of self- and point of care testing have to be viewed in a more holistic context where the benefits of the immediacy of testing can be realised. In the context of diabetes mellitus this is measured in the longer term benefits of better glycaemic control, namely delay in the onset of complications (31). This has been attempted in relation to the impact of renal complications illustrating the complexity of the analysis – and the associated policy making that is required (32).

It is acknowledged that health economics is not an exact science but provides a valuable management tool for decision makers and operational managers (33). Difficulties arise in particular with the breadth of the services involved, the way in which these services are resourced, and the length of time it takes for the benefit to be achieved. The example of the economic impact of self-testing for glucose, HbA_{1C} and microalbumin monitoring is clearly a complex subject. The advent of molecular testing for bacterial and viral antigens – with the inherent improvement in the sensitivity of detection – appears to be less complex. Evidence and economic modelling in the case of molecular testing for tuberculosis and chlamydia infection have shown that there is a significant reduction in time to produce a result, earlier diagnosis is possible and the reduction in complication rates and associated healthcare costs is dramatic. The overall reduction in healthcare costs is significant – however, the cost of the laboratory testing is increased over conventional technology (34, 35).

How to use evidence

The last example given of the potential impact of molecular testing offers an example of the challenge that exists in laboratory medicine today and which calls for a change in culture both within the laboratory medicine community and amongst healthcare policy makers (36, 37). The change in culture required within the laboratory medicine community is for greater integration into multidisciplinary clinical teams in order that a more holistic view can be taken of the impact of the diagnostic services. This integration must be combined with a new direction for research and development to encompass analysis of the clinical and economic impact of diagnostic tests – including modalities of delivery.

Resource allocation for healthcare varies from country to country but invariably fails to respond quickly to changes in technology or to clinical need. Health technology assessment was developed in part to try and provide a rapid means of evaluating new technologies in order to speed up implementation when benefit could be demonstrated (38). At a local hospital or laboratory level the production of a business case offers the most valuable tool for securing investment, whilst clinical audit provides the tool for assuring that best practice is maintained. Clinical audit requires the explicit statement of a set of standards of practice against which current practice is then assessed – evidence, based on clinical outcomes, is used to define these standards (39, 40).

Concluding remarks

The discipline of evidence based practice which seeks to promote the use of the best evidence available to guide decision making in the best interests of patients is a powerful tool to guide the appropriate utilisation of the laboratory medicine services. It provides a means of ensuring that research is directed toward relevant clinical outcomes. At the present time the evidence on diagnostic tests is focused primarily on technical and diagnostic performance with little attention to outcomes. The link with health economics to assess the economic outcomes derived from diagnostic tests has the potential to ensure the appropriate allocation of resources which is of critical importance as more biomarkers and test delivery platforms are discovered.

Literature

1. Kricka LJ, Parsons D, Coolen RB. Healthcare in the United States and the practice of laboratory medicine. Clin Chim Acta 1997; 267: 5-32.
2. Price CP, Barnes IC. Laboratory medicine in the United Kingdom: 1948-1998 and beyond. Clin Chim Acta 2000; 290: 5-36
3. Walraven C van, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. JAMA 1998; 280: 550-558.
4. Perraro F, Rossi P, Liva C, Bulfoni A, Ganzini G, Giustinelli A, et al. Inappropriate emergency test ordering in a general hospital: preliminary reports. Qual Assur Health Care 1992; 4: 77-81.
5. Sandler G. Do emergency tests help in the management of acute medical admissions? BMJ 1984; 289: 973-977.
6. Bareford D, Hayling A. Inappropriate use of laboratory services: long term combined approach to modify request patterns. BMJ 1990; 301: 1305-1307.
7. Witte DL. Measuring outcomes: why now? Clin Chem 1995; 41: 775-580.
8. Lundberg GD. The need for an outcomes research agenda for clinical laboratory testing. JAMA 1998; 280: 565-566.
9. Lundberg GD. How clinicians should use the diagnostic laboratory in a changing medical world. Clin Chim Acta 1999; 280: 3-11.
10. WD. Outcomes research in diagnostics: not yet. In Vivo 1995; Jan. 11-7.
11. Wong ET. Improving laboratory testing: can we get physicians to focus on outcome? Clin Chem 1995; 41: 1241-1247.
12. Sackett DL, Straw SE, Richardson WS, Rosenberg W, Haynes RB. Evidence based medicine. How to practice and teach EBM. 2nd edn. Churchill Livingstone, Edinburgh, 2000.
13. Winter RJ de, Koster RW, Sturk A, Sanders GT. Value of myoglobin, troponin T, and CK-MB mass in ruling out an acute myocardial infarction in the emergency room. Circulation 1995; 92: 3401-3407.
14. Anderson FP, Jesse RL, Nicholson CS, Miller WG. The costs and effectiveness of a rapid diagnostic and treatment protocol for myocardial infarction. In: Assessing Clinical Outcomes: utilizing appropriate laboratory testing to decrease healthcare costs and improve patient outcomes. AACC Leadership Series 1995; 20-24.
15. Rink E, Hilton S, Szczepura A, Fletcher A, Fletcher J, Sibbald B, et al. Impact of introducing near-patient testing for standard investigations in general practice. Brit Med J 1993; 307: 775-778
16. Reid CK, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA 1995; 274: 645-651.
17. Nierenberg AA, Feinstein AR. How to evaluate a diagnostic marker test: lessons from the rise and fall of dexamethasone suppression test. JAMA 1988; 259: 1699-1702.
18. Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120: 667-676.
19. Greenhalgh T. How to read a paper: papers that report diagnostic or screening tests. BMJ 1997; 315: 540-543.
20. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282: 1061-1066.
21. Valenstein PN. Evaluating diagnostic tests with imperfect standards. Am J Clin Pathol 1990; 93: 252-258.
22. Begg CB. Biases in the assessment of diagnostic tests. Stats in Med 1987; 6: 411-423.
23. Moore RA, Fingerova H. Evidence-based laboratory medicine: using current best evidence to treat individual patients. In: Point-of-Care Testing. (Eds CP Price, JM Hicks). AACC Press, Washington, USA. 1999: 265-288.
24. Easterbrook PJ, Berline JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991; 337: 867-872.
25. Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. JAMA 1992; 267: 374-378.
26. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. JAMA 1990; 46: 197-207.
27. Fryback FG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11: 88-94.
28. Price CP. Evidence-based laboratory medicine: supporting decision making. Clin Chem 2000; 46: 1041-1050.
29. Patsalos PN, Sander JWAS, Oxley J. Immediate anticonvulsive drug monitoring in management of epilepsy. Lancet 1987; ii: 39.
30. Hicks JM, Haeckel R, Price CP, Lewandrowski K, Wu AHB. Recommendations and opinions for the use of point-of-care testing for hospitals and primary care: summary of a 1999 symposium. Clin Chim Acta 2001; 303: 1-17.
31. Gray A, Raikou M, McGuire A, Fenn P, Stevens R, Cull C, et al. Cost effectiveness of an intensive blood glucose control policy in patients with type 2 diabetes: economic analysis alongside randomised controlled trial (UKPDS 41). Brit Med J 2000; 320: 1373-1378.
32. Borch-Johnsen K, Wenzel H, Viberti GC, Mogensen CE. Is screening and intervention for microalbuminuria worthwhile in patients with insulin dependent diabetes. Brit Med J 1993; 306: 1722-1725.
33. Williams CJ. Is health economics useful in cancer? Crit Rev Oncology Hematology 1999; 30: 25-33.
34. Howell MR, Quinn TC, Brathwaite W, Gaydos CA. Screening women for chlamydia trachomatis in family planning clinics. Sexually Transm Dis 1998; 25: 108-117.

35. Genç M, Mårdh P-A. A cost-effectiveness analysis of screening and treatment for *chlamdya trachomatis* infection in asymptomatic women. Ann Intern Med 1996; 124: 1-7.
36. Muir Gray JA (ed). Evidence-based healthcare. How to make health policy and management decisions. Churchill Livingstone, Edinburgh; 1997.
37. Haines A, Donald A. Getting research findings into practice. BMJ Publishing Group, 1998.
38. Price CP. Health technology assessment. RCPath Bull 2000; 110: 26-28.
39. Williams O. What is clinical audit? Ann R Coll Surg Engl 1996; 78: 406-411.
40. Lord J, Littlejohns P. Evaluating healthcare policies: the case of clinical audit. BMJ 1997; 315: 668-671.

Ned Tijdschr Klin Chem 2001; 26: 242-245

Pijlers van precisie: maten en getallen voor het onderbouwen van testgebruik

P. M.M. BOSSUYT

"*In God we trust, the others must provide sound data*". Dit is een tijd waarin weinigen nog op hun woord alleen worden geloofd. Alom klinkt de roep om onderbouwing van stellingen en de vraag naar rekenschap en verantwoording. Het liefst ziet men in de antwoorden maat en getal opduiken. Ook de professionals in de zorg ontkomen niet aan deze tendens. Bewust van alle vertekenende invloeden gaan zij op zoek naar gegevens die zich lenen voor het onderbouwen van besluiten over het te voeren beleid. Zie daar de basis voor 'Evidence based Medicine': een professioneel antwoord op een vraag naar onderbouwing, in een atmosfeer van toegenomen rekenschap en verantwoording (1). Worden patiënten hier beter van? Zo ja, in welke mate? Staat die verbetering in een redelijke verhouding tot wat die patiënten zelf, hun artsen en de maatschappij aan middelen moet investeren in die gezondheidswinst c.q. behoud van gezondheid? Dergelijke vragen zijn niet ongewoon bij nieuwe geneesmiddelen. Medische tests ontsnappen echter niet aan vergelijkbare verzoeken. Of het nu gaat om beeldvormend onderzoek of laboratoriumtesten, de vraag naar het waarom en waarvoor zal ook daar worden gesteld. In welke mate is voor een test een antwoord beschikbaar op deze vragen naar onderbouwing? Een verkenning hiervan kan enkel maar aanleiding geven tot gepaste bescheidenheid. Hieronder volgt een korte inleiding.

Sensitiviteit en specificiteit

De eerste vraag die wordt gesteld is: kan ik wel varen op de uitslagen van deze test? We gaan voor het gemak uit van de discussie dat evidente vragen over de veiligheid, de ijking en de betrouwbaarheid al naar tevredenheid zijn beantwoord. Spreekt de test de waarheid? Voor het antwoord hierop worden de uitslagen van de test die wordt geëvalueerd – laten we die de indextest noemen – vergeleken met die van een referentiestandaard. De mate van overeenkomst tussen de uitslagen kan op verschillende manieren worden uitgedrukt. Laten we als voorbeeld de evaluatie van D-dimer nemen, een test voor het aantonen c.q. uitsluiten van longembolie. Kline en collegae rapporteerden over deze test in JAMA (2). Ze hadden de test afgenomen bij 380 patiënten die met verdenking van longembolie op het "emergency department" van een van de deelnemende academisch ziekenhuizen waren gezien. De test was positief bij 164 van hen. Hiervan kon de diagnose bij 60 worden bevestigd: deze 60 hadden ook een positieve uitslag met de referentiestandaard. (Tabel 1). Omgekeerd hadden van de 216 patiënten met een negatieve uitslag op de D-dimer er 4 uiteindelijk toch een longembolie. De beste manier (op pathologie na) voor het aantonen c.q. uitsluiten van longembolie is longangiografie. Deze werd echter lang niet bij iedereen uitgevoerd. In het onderzoek werd een zogenaamde gemengde referentiestandaard toegepast. Dat betekent dat de diagnose longembolie kon worden geverifieerd door een "high-probability" V/Q scan, een afwijkende spiraal CT, een non-high V/Q scan of door overlijden tijdens de follow-up als niet kon worden uitgesloten dat dit het gevolg was van een veneuze tromboëmbolie. Een eerste blik op tabel 1 leert ons dat er een redelijke, maar verre van perfecte overeenkomst is tussen indextest en referentie standaard. Wie enkel zou varen op de uitslagen van deze indextest maakt een aantal fouten. Het is gebruikelijk om het percentage correcte uitslagen conditioneel op de geverifieerde ziektestatus uit te drukken.

Tabel 1. De uitslagen van een onderzoek naar de diagnostische waarde van D-dimer bij het uitsluiten van longembolie

	Referentiestandaard		
Indextest	Positief	Negatief	Totaal
Positief	60	104	164
Negatief	4	212	216
	64	316	380

Data uit Kline et al. JAMA 2001; 285: 761-768.

Afdeling Klinische Epidemiologie en Biostatistiek, Academisch Medisch Centrum, Universiteit van Amsterdam

Correspondentie: Prof. Dr. P.M.M. Bossuyt, Afdeling Klinische Epidemiologie en Biostatistiek, Academisch Medisch Centrum, Postbus 22660, 1100 DD Amsterdam
e-mail: p.m.bossuyt@amc.uva.nl