

Education and debate

Evidence based diagnostics

Christian Gluud, Lise Lotte Gluud

Diagnostic tests are often much less rigorously evaluated than new drugs. It is time to ensure that the harms and benefits of new tests are fully understood

Cochrane
Hepato-Biliary
Group,
Copenhagen Trial
Unit, Centre for
Clinical
Intervention
Research, H:S
Rigshospitalet,
Copenhagen
University Hospital,
DK-2100
Copenhagen,
Denmark
Christian Gluud
head of department
Lise Lotte Gluud
specialist registrar

Correspondence to:
C Gluud
cgluud@ctu.rh.dk

BMJ 2005;330:724-6

No international consensus exists on the methods for assessing diagnostic tests. Previous recommendations stress that studies of diagnostic tests should match the type of diagnostic question.^{1,2} Once the specificity and sensitivity of a test have been established, the final question is whether tested patients fare better than similar untested patients. This usually requires a randomised trial. Few tests are currently evaluated in this way. In this paper, we propose an architecture for research into diagnostic tests that parallels the established phases in drug research.

Stages of research

We have divided studies of diagnostic tests into four phases (box). We use research on brain natriuretic peptide for diagnosing heart failure as an illustrative example.² However, the architecture is applicable to a wide range of tests including laboratory techniques, diagnostic imaging, pathology, evaluation of disability, electrodiagnostic tests, and endoscopy.

Establishing the normal range

In drug research, phase I studies deal with pharmacokinetics, pharmacodynamics, and safe doses.³ Phase I diagnostic studies are done to determine the range of results obtained with a newly developed test in healthy people. For example, after development of a test to

measure brain natriuretic peptide in human plasma, phase I studies were done to establish the normal range of values in healthy participants.^{4,5}

Diagnostic phase I studies must be large enough to examine the potential influence of characteristics such as sex, age, time of day, physical activity, and exposure to drugs. The studies are relatively quick, cheap, and easy to conduct, but they may occasionally raise ethical problems—for example, finding abnormal results in an apparently healthy person.⁶

Diagnostic accuracy

In phase II, studies explore the diagnostic accuracy of a test in participants with both known and suspected relevant disease. Phase IIa studies compare test results in participants with disease diagnosed by a standard method with those in healthy participants (from diagnosis to test result). For example, a phase IIa study found significantly raised concentrations of brain natriuretic peptide in participants with left ventricular dysfunction diagnosed by echocardiography (median 493.5 (range 248.9-909.0) pg/ml) compared with healthy participants (129.4 (53.6-159.7) pg/ml).⁷ Subsequently, brain natriuretic peptide was recommended as a useful diagnostic aid for left ventricular dysfunction.⁷

After an association has been found between test results and a certain disease, phase IIb studies may be done to examine whether test results are related to the severity of a disease. For example, in a phase IIb study, brain natriuretic peptide concentrations were measured in healthy participants and participants with congestive heart failure.⁸ The study found a linear relation between test values and the degree of ventricular dysfunction. The authors concluded that the concentration of brain natriuretic peptide is a good indicator of the severity of chronic heart failure.⁸ However, the design only allows inferences about how a test works under ideal conditions.

Phase IIc studies examine the predictive value of a test among people with suspected disease (from test results to diagnosis). For example, a phase IIc study measured brain natriuretic peptide concentrations in participants with suspected heart disease.⁹ All participants had transthoracic echocardiography. The results showed raised concentrations of brain natriuretic peptide in participants with left ventricular systolic



GLUSTO/SPL

The harms and benefits of diagnostic tests needs evaluating—just as drugs do

Four phases in architecture of diagnostic research

Phase I—Determining the normal range of values for a diagnostic test through observational studies in healthy people

Phase II—Determining the diagnostic accuracy through case-control studies, including healthy people and (a) people with known disease assessed by diagnostic standard and (b) people with suspected disease

Phase III—Determining the clinical consequences of introducing a diagnostic test through randomised trials

Phase IV—Determining the effects of introducing a new diagnostic test into clinical practice by surveillance in large cohort studies

dysfunction (median 79.4 (interquartile range 35.9–151.0) pg/ml) compared with those with normal ventricular systolic function (26.7 (12.2–54.3) pg/ml).⁹ A concentration > 17.9 pg/ml had a sensitivity of 88% and specificity of 34%. Choosing different cut-off points did not improve the predictive characteristics.

The authors concluded that measuring brain natriuretic peptide in addition to routine investigations provides a small diagnostic advantage.⁹ However, the characteristics of the test may be different in other settings. A narrative review summarised several phase II studies on brain natriuretic peptides for diagnosing left ventricular systolic dysfunction.¹⁰ The studies found that sensitivity ranges from 26% to 92% and specificity from 34% to 89%. The predictive ability seemed to depend on sex, and the test performed less well in community based studies than in referral series.

Several concerns surround the validity and applicability of phase II studies. Two of the most important concerns are blinded evaluations of test results and selection of cut-off values or limits for normal values.² To improve the quality of reporting of studies of diagnostic tests, the Standards for Reporting of Diagnostic Accuracy (STARD) Initiative was launched.¹¹ Checklists and flowcharts were developed to aid authors of phase II studies. Future studies are planned to evaluate the effect of the initiative.

Clinical effects

In some cases, the value of a diagnostic test is self evident—for example, in genetic testing. However, for most diagnostic tests, phase III studies are necessary to evaluate the beneficial and harmful effects of implementing a new test. The potential effects depend on how the information is used in subsequent clinical decisions. In phase III diagnostic studies, randomisation determines whether participants have the test or not. In some randomised trials, the result of the test may be used to determine a specific clinical course, including treatment. Alternatively, knowledge of a test result may be incorporated into standard clinical practice and treatment strategies remain unchanged.

A phase III study compared the effect of using brain natriuretic peptide concentrations or clinical assessment to guide treatment.¹² The study included 69 participants with impaired systolic function and symptomatic heart failure. Participants were randomised to receive treatment guided by brain natriuretic peptide

concentrations or by a clinical score of symptoms and signs of heart failure. Fewer deaths, hospital admissions, and cases of decompensation of heart failure occurred among participants whose treatment was guided by brain natriuretic peptide values than among those whose treatment was guided by clinical score.

The study shows the way for diagnostic research. However, the interpretation of the results is not simple. Larger trials with the most recently developed drugs are necessary before the test is implemented in clinical practice. The benefits and harms of the test in other settings—for example, in screening for asymptomatic left ventricular dysfunction—also seem relevant.

Methodological issues also arise. Estimation of required sample size is difficult in diagnostic trials.¹³ In randomised trials comparing two binary diagnostic tests, patients in the two arms with concordant results will not contribute to the final difference. Sample size estimations in such trials therefore include discordance rates. Other methodological aspects are similar to those in randomised drug trials. In both trial types, methods for adequate generation of the allocation sequence, allocation concealment, and blinding deserve attention.¹⁴ When several randomised trials on diagnostic tests are completed, systematic reviews and possibly meta-analyses are warranted.¹⁵

Long term consequences

Logistical problems such as storage, freezing, and thawing of samples or poor calibration of equipment may affect the accuracy of a diagnostic test after it is introduced into routine clinical practice. Several factors, such as a change in diagnostic indications, may influence the circumstances under which a test is used. Phase IV studies are therefore needed to determine whether the diagnostic accuracy of a test in practice corresponds to predictions from systematic reviews of phase III trials.

Phase IV studies include large cohorts of consecutive participants. Regular reports on regional, national, and international quality and bench markings may also help improve quality of testing in clinical practice. Phase IV diagnostic studies are an important aid in quality assurance and quality development and are necessary to identify rare adverse events.¹⁶

Conclusion

Few will argue that valid evidence is necessary before we introduce new drugs in clinical practice. The randomised trial is the best method for comparing interventions. Randomised trials are also necessary to evaluate the potential effects of introducing a diagnostic test. Unfortunately, few randomised trials deal with diagnostic tests. We searched the Cochrane Central Register of Controlled Trials (Issue 1, 2005) and found that only 4.2% (18 366 of 435 786 records) dealt with diagnostic tests or screening. Awareness of the need for evidence based diagnostic testing must be increased. Organisations such as the Cochrane Collaboration can help by improving facilities for and methodological quality of systematic reviews of diagnostic tests.

The demand for diagnostic phase III and phase IV studies is increasing with the continuous development of new diagnostic methods. Although defensive use of

Summary points

The harms and benefits of diagnostic tests should be fully evaluated before they are used in clinical practice

A four phase process of assessment is suggested, mirroring that used for new drugs

The first phase focuses on establishing the normal range

The second phase focuses on establishing sensitivity and specificity and other measures of diagnostic accuracy

Randomised trials are then needed to determine whether patients benefit from the testing

The final phase is large continuous surveillance studies to identify consequences of testing in clinical practice

diagnostic tests improves clinical outcomes for some patients, it worsens clinical outcomes for others.¹⁷ The four temporal phases of research provide a logical, stepwise procedure for development of diagnostic tests. However, the four phases do not apply to all diagnostic tests or provide an adequate basis for all types of diagnostic studies. Furthermore, one type of study may occur in several phases. The phase concept is meant as a guide that may be adjusted according to individual circumstances.

Contributors and sources: CG directs The Copenhagen Trial Unit, a non-specialty oriented centre for clinical intervention research and studies random and systematic errors in clinical research. LLG studies random and systematic errors in clinical research. CG and LLG are physicians and editors of the Cochrane Hepato-Biliary Group. The literature came from unsystematic and systematic searches of PubMed, The Cochrane

Library, and personal files. CG drafted and LLG revised the paper. CG is the guarantor.

Competing interests: None declared.

- 1 Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: WB Saunders, 1985.
- 2 Sackett D, Haynes RB. The architecture of diagnostic research. *BMJ* 2002;324:539-41.
- 3 International Conference on Harmonisation Steering Committee. *ICH harmonised tripartite guideline. General considerations for clinical trials*. http://www.ich.org/MediaServer.jserv?@_ID=484&@_MODE=GLB (accessed 29 Jan 2005).
- 4 Ationu A, Carter ND. Brain and atrial natriuretic peptide plasma concentrations in normal healthy children. *Br J Biomed Sci* 1993;50:92-5.
- 5 Jensen KT, Carstens J, Ivarsen P, Pedersen EB. A new, fast and reliable radioimmunoassay of brain natriuretic peptide in human plasma. Reference values in healthy subjects and in patients with different diseases. *Scand J Clin Lab Invest* 1997;57:529-40.
- 6 Illes J, Desmond JE, Huang LF, Raffin TA, Atlas SW. Ethical and practical considerations in managing incidental findings in functional magnetic resonance imaging. *Brain Cognition* 2002;50:358-65.
- 7 Talwar S, Sieberhofer A, Williams B, Ng L. Influence of hypertension, left ventricular hypertrophy, and left ventricular systolic dysfunction on plasma N terminal pre-BNP. *Heart* 2000;83:278-82.
- 8 Selvais PL, Donckier JE, Robert A, Laloux O, van Linden F, Ahn S, et al. Cardiac natriuretic peptides for diagnosis and risk stratification in heart failure. *Eur J Clin Invest* 1998;28:636-42.
- 9 Landray MJ, Lehman R, Arnold I. Measuring brain natriuretic peptide in suspected left ventricular systolic dysfunction in general practice: cross-sectional study. *BMJ* 2000;320:985-6.
- 10 Wang TJ, Levy D, Benjamin EJ, Vasani RS. The epidemiology of "asymptomatic" left ventricular systolic dysfunction: implications for screening. *Ann Intern Med* 2003;138:907-16.
- 11 Bossuyt PM, Retisma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. *Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative*. www.consort-statement.org/stardstatement.htm (accessed 7 Jun 2004).
- 12 Thoughton RW, Frampton CM, Yandle TG, Espiner EA, Nicholls MG, Richards AM. Treatment of heart failure guided by plasma aminoterminal brain natriuretic peptide (N-BNP) concentrations. *Lancet* 2000;355:1126-30.
- 13 Lijmer JG, Bossuyt PM. Diagnostic testing and prognosis: the randomised controlled trial in diagnostic research. In: Knottnerus JA, ed. *The evidence base of clinical diagnosis. How to do diagnostic research*. London: BMJ Books, 2002: 61-80.
- 14 Kjaergard LL, Villumsen J, Gluud C. Reported methodological quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135:982-9.
- 15 Cochrane Collaboration. Cochrane Screening and Diagnostic Tests Methods Group. *Cochrane Library*. Issue 2. Oxford: Update Software, 2003.
- 16 Knottnerus JA. Epilogue: overview of evaluation strategy and challenges. In: Knottnerus JA, ed. *The evidence base of clinical diagnosis. How to do diagnostic research*. London: BMJ Books, 2002:209-15.
- 17 DeKay ML, Asch DA. Is the defensive use of diagnostic tests good for patients, or bad? *Med Decis Making* 1998;18:19-28.

(Accepted 24 January 2005)

Submitting articles to the *BMJ*

We are now inviting all authors who want to submit a paper to the *BMJ* to do so via the web (<http://submit.bmj.com>).

We have introduced Benchpress, our web based manuscript tracking system, with the aim of streamlining our processes and providing better, quicker information for authors, reviewers, and editors.

Benchpress is a website where authors deposit their manuscripts and editors go to read them and record their decisions. Reviewers' details are also held on the system, and when asked to review a paper reviewers will be invited to access the site to see the relevant paper. The system is secure, protected by passwords, so that authors see only their own papers and reviewers see only those they are meant to. The system is run by Highwire Press, who host bmj.com, and is already being used by 30 journals, including most of the BMJ Publishing Group's specialist journals.

For authors in particular the system offers several benefits. The system provides all our guidance and forms and allows authors to

suggest reviewers for their paper—something we'd like to encourage. Authors get an immediate acknowledgement that their submission has been received, and they can watch the progress of their manuscript. The record of their submission, including editors' and reviewers' reports, remains on the system for future reference.

Anyone with an internet connection and a web browser can use the system.

The system itself offers extensive help, and the BMJ Online Submission Team is geared up to help authors and reviewers if they get stuck. We see Benchpress as part of our endeavour to improve our service to authors and reviewers and, as always, we'd welcome feedback.

Benchpress is accessed via <http://submit.bmj.com> or via a link from bmj.com