Diagnostic research

Diagnostic studies as multivariable, prediction research

K G M Moons, D E Grobbee

Patient outcomes in diagnostic research

Dr Feinstein provides a topical overview of the history and current status of diagnostic research.[1] It addresses the various forms of bias that may occur in research aiming to evaluate the accuracy of diagnostic tests. Diagnostic accuracy is defined by the extent to which a test correctly indicates the ("true") presence or absence of the disease at issue as determined by a particular reference. We largely agree with this overview. However, we would like to discuss in more detail a certain issue raised by Dr Feinstein. Notably, the use of quantitative statistical models in diagnostic research and related to this the practitioners' judgement, and the role of test results in offering prognostic information.

MATHEMATICAL MODELS TO ESTIMATE THE TRUE (ADDED) DIAGNOSTIC ACCURACY OF A TEST

Dr Feinstein argues that diagnostic research is currently unfortunately dictated by mathematical, quantitative models, often ignoring the practitioners' judgement. We, however, believe that these quantitative models are *necessary* in order to estimate the true, independent (or added) value of a test, just as they are necessary in aetiological studies to estimate the independent association of a particular exposure and the occurrence of a particular outcome.

In medical practice, the diagnostic investigation starts with a patient presenting with a particular symptom or sign indicative for the presence of a particular disease, the so called target disease.[2] The diagnostic investigation is a consecutive (hierarchical) process always starting from patient history and physical examination, followed by more invasive, time consuming and costly tests such as imaging. It amounts to the estimation of the probability of presence of the target disease given all test results, further referred to as the diagnostic probability.[3] In practice no diagnosis is set by a single test result and each test result is judged together with other (previous) test results. Given this consecutive investigation, the fact that different tests provide to varying extents the same information and that each test may be more or less burdening for the patient, time consuming and costly, physicians need to know which tests additionally or independently contribute to the diagnostic probability estimation.[1–11] For example, to determine whether an exercise stress test in patients suspected of coronary artery disease has diagnostic value, one needs to quantify whether the test changes the diagnostic probability based on previous test results (for example, patient history and physical examination) that are commonly available anyway before such test is applied.

In our view, to quantify whether a particular test result additionally increases or decreases the diagnostic probability *requires* multivariable (prediction) modelling of the occurrence (prevalence) of the disease at issue as function of the different diagnostic test results. It requires a comparison of the probability (prevalence) of disease presence before and after the test, using, for example, stepwise logistic regression modelling according to the chronology in practice. The result of such analyses is the definition of one or more diagnostic prediction models including the relevant (contributing) tests. If needed, these models can be simplified to obtain easy applicable diagnostic decision rules for use in practice.[12–15]

Knowledge on the independent (added) diagnostic value can, as Dr Feinstein also suggests, not be inferred from the still widely used singular test parameters—that is, a test's sensitivity, specificity, and likelihood ratio.[1 11] Similarly to the variability of these parameters across patient characteristics, the independent value of a test can also not be validly quantified by the use of Bayes's theorem.[3] The need for multivariable models in diagnostic research is not different form other types of medical research such as aetiological and prognostic research. Not the singular association (odds ratio) between a particular exposure and the occurrence of the outcome is informative, but their association independent of other (co) factors.

PRACTITIONERS' JUDGEMENT TO ASSESS DIAGNOSTIC ACCURACY

Dr Feinstein argues that physicians are generally reluctant to use quantitative prediction models in their diagnostic practice but rather infer a test's accuracy from their own experience. In our view, this should be no reason to withdraw from the use of quantitative prediction models in diagnostic research.

For the evaluation of efficacy of preventive and therapeutic devices we have left the era of "experience or eminence based medicine" in favour of "evidence based medicine". This should also apply to diagnostic technologies.

Resistance to use prediction models in diagnostic practice is most likely because they are still seldomly used and reported in diagnostic studies, in contrast with aetiological and prognostic, including therapeutic, research. For the latter types of research a proper methodological framework encouraging the use of these methods can be found in all epidemiological textbooks. This is not the case for the few textbooks addressing methods of diagnostic research. Most diagnostic research still concentrates on the estimation of singular test parameters in the analyses. To enable diagnostic research to better serve practice the prevailing framework of methods for both design and analyses must be improved.

We agree, however, that the physician's judgement or experience is often an important "test" in the diagnostic investigation, which should not be ignored in research. However, we believe it is preferable to express this experience in more or less objective test results, or at least to quantify whether this "experience test" provides added information to the more objective test results (as described above).

THE VALUE OF TEST RESULTS IN TERMS OF PATIENT OUTCOME

To set a diagnosis is fundamental in medical care as it offers an indication of the patient's prognosis and directs therapeutic management. As Dr Feinstein suggests, a diagnostic technology may not contribute to the assessment of the final diagnosis, but rather provide information that could be relevant to the patient's prognosis or to therapeutic decisions.

We agree that diagnostic technologies should not only be evaluated on their diagnostic accuracy, for example, their ability to determine the presence or absence of the disease but sometimes also on their ability to change patient outcome.[16–19] These assessments, however, call for different types of outcomes and thus different study designs. Evaluation of a test on its (added) diagnostic accuracy requires a multivariable, cross sectional prediction study: of each subject the result of the test under study is measured as well as the simultaneous presence or absence of the disease as determined by a reference. Evaluating a test on patient outcome comprises the

evaluation of diagnostic tests plus all possible administered therapies combined. This enters the realm of aetiological and therapeutic research and requires follow up studies or trials rather than cross sectional, prediction studies.[19 20] We believe that evaluation of diagnostic tests on patient outcome is not always necessary. In general, we think that follow up studies are not necessary and the (beneficial) effect of a diagnostic test for patient outcome may be considered as established if (1) diagnostic (cross sectional) studies have shown the test's ability to detect a particular disease and (2) therapeutic studies provided evidence on efficacy of the management of this disease.[20] Follow up studies to quantify the effect of a diagnostic technology on patient outcome are necessary (1) if the disease at issue lacks a specific reference to determine its presence or absence (such as heart failure) or (2) if the (new) diagnostic technology provides other therapeutic information than the reference, potentially leading to other treatment choices or (3) if the technology itself may have therapeutic properties such as salphingography to determine patency of the uteral tubes.

J Epidemiol Community Health 2002;**56**:337–338

Authors' affiliations
K G M Moons, D E Grobbee, Julius Centre for Health Sciences and Primary Care, University Medical Centre, Utrecht, Netherlands

Correspondence to: Dr D E Grobbee, Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, PO Box 80035, 3508 TA Utrecht, Netherlands; www.juliuscenter.nl

REFERENCES

1 **Feinstein AR**. Misguided efforts and future challenges for research on "diagnostic tests". *J Epidemiol Community Health* 2002;**56**:330–2.
2 **Sackett DL**, Haynes RB, Tugwell P. *Clinical epidemiology; a basic science for clinical medicine*. Boston: Little, Brown, 1985.
3 **Moons KGM**, van Es GA, Deckers JW, Habbema JDF, *et al*. Limitations of sensitivity, specificity, likelihood ratio and Bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 1997;**8**:12–17.
4 **Feinstein AR**. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: WB Saunders, 1985.
5 **Harrell FE**, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 1988;**80**:1198–202.
6 **Begg CB**. Advances in statistical methodology for diagnostic medicine in the 1980s. *Stat Med* 1991;**10**:1887–95.
7 **Knottnerus JA**. Application of logistic regression to the analysis of diagnostic data: exact modeling of a probability tree of multiple binary variables. *Med Decis Making* 1992;**12**:93–108.
8 **Jaeschke R**, Guyatt GH, Sackett DL, *et al*. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients. *JAMA* 1994;**271**:703–7.
9 **Grobbee DE**, Miettinen OS. Clinical epidemiology: introduction to the discipline. *Neth J Med* 1995;**47**:2–5.
10 **Miettinen OS**, Henschke CI, Yankelevitz DF. Evaluation of diagnostic imaging tests: diagnostic probability estimation. *J Clin Epidemiol* 1998;**51**:1293–8.
11 **Moons KGM**, van Es GA, Michel BC, *et al*. Redundancy of single diagnostic test evaluation. *Epidemiology* 1999;**134**:276–81.
12 **Sox H Jr**. Clinical prediction rules in practice. *Hosp Pract* 1986;**21**:100–3.
13 **Wasson JH**, Sox HC, Neff RK, *et al*. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;**313**:793–9.
14 **Harrell FE**, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing erros. *Stat Med* 1996;**15**:361–87.
15 **Laupacis A**, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;**277**:488–94.
16 **Begg CB**. Experimental design of medical imaging trials. Issues and options. *Invest Radiol* 1989;**24**:934–6.
17 **Fryback D**, Thornbury J. The efficacy of diagnostic imaging. *Med Decis Making* 1991;**11**:88–94.
18 **Dixon AK**. Evidence-based diagnostic radiology. *Lancet* 1997;**350**:509–12.
19 **Bossuyt PPM**, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 2000;**356**:1844–7.
20 **Moons KGM**, Ackerstaff RGA, Moll FL, *et al*. Association of intraoperative transcranial doppler monitoring variables with stroke from carotid endarterectomy. *Stroke* 2001;**32**:813.

Diagnostic research

The need for expanding and re-focusing of statistical approaches in diagnostic research

H Brenner, T Stürmer, O Gefeller

Fitting statistical methodology to the need of diagnostic research

In his contribution "Misguided efforts and future challenges for research on diagnostic tests"[1] Dr Feinstein has identified major gaps and shortcomings in previous and current diagnostic research. While we fully agree with most of his criticisms, we would like to take issue with him over the role of statistics and mathematical formalisation in diagnostic research. In particular, we would like to emphasise the need and potential of expanding and re-focusing rather than abandoning statistical approaches in diagnostic research.

The traditional concepts of sensitivity and specificity as well as of the "posterior probabilities" have certainly been useful as a methodological framework for structuring efforts to quantify accuracy of diagnostic markers in well defined, very special settings in the past, and they may continue to be useful as such in the future. The major limitation of these concepts does not so much lie in their intrinsic properties, but in the uncritical adoption of these concepts to a wide range of different, usually more complex settings. This misapplication along with some misconceptions outlined below have often been severely misguiding indeed.

MISCONCEPTIONS AND LIMITATIONS IN DIAGNOSTIC MARKER EVALUATION

An important example is the dogma still found in most textbooks of clinical epidemiology and biostatistics, that the sensitivity and specificity as well as the likelihood ratios are constant benchmarks of test performance, which, in contrast to the posterior probabilities, are independent of disease prevalence in the population studied. As pointed out by Dr Feinstein, this dogma has repeatedly been challenged in various settings by empirical counter-evidence. Furthermore, it has been demonstrated by more general methodological work that in situations commonly encountered in practice, in which diagnostic tests are based on dichotomisation of inherently continuous traits rather than on inherently dichotomous traits, major departures from this dogma are expected to be the rule rather than the exception.[2] In particular, it has been shown that for tests based on dichotomisation of inherently continuous traits, variation with