The winding road towards evidence based diagnoses

I Hernández-Aguado

Matching diagnostic research with the knowledge needed for an evidence based diagnosis

Why are quality standards for therapeutic research so high in comparison with diagnostic standards? This difference has negative effects on the quality of diagnostic research and on its application to medical care. Various reviews dealing with the quality of research on clinical diagnosis have repeatedly reported that a large number of studies have serious flaws, with just a small proportion of studies fulfilling a high number of methodological standards. Improvements have certainly been made in recent years, but they fall short of what has been achieved in other areas of clinicoepidemiological research.¹⁻³ The results of therapeutic research seem to be more rapidly applied and to readily form the basis for new recommendations in clinical practice; whereas the findings of diagnostic research are incorporated to a much slower extent in clinical practice and in the formulation of clinical practice recommendations. While the body of evidence based medicine is rich in evaluations and recommendations concerning therapeutic interventions, the number of reviews or recommendations regarding diagnostic procedures is scant. Another striking, though not surprising, aspect is the dissociation between academic proposals concerning how clinical diagnoses should be made and what doctors actually do when weighting a diagnosis. In particular, the actual use of the indices and quantitative procedures proposed by academic circles is the exception rather than the rule.⁴

The great schism that exists between research and clinical practice, suggests that significant changes should be made in diagnostic research. The *Journal of Epidemiology and Community Health* decided it would be a good idea to publish an exchange of views on the future of diagnostic research, which might point to possible paths to be taken by colleagues conducting research into clinical epidemiology. With this aim in mind, Dr Alvan R Feinstein was invited to write a manuscript on the future challenges for diagnostic research.⁵ We also invited prestigious investigators in this field to comment on the subject, based on Dr

Feinstein's article.⁶⁻¹¹ We know that Dr. Feinstein would have been delighted to respond to the various comments, as planned. Unfortunately, this has not been possible, because Dr Feinstein suddenly died while the debate that we publish in this issue of *JECH* was in press. Obviously, in this editorial, it would be improper to adopt the position that Dr Feinstein might hypothetically have taken. Rather, my intent is to summarise the different opinions that this issue hosts, and to attempt to underline certain questions that I deem of utmost significance to improve diagnostic research.

The mathematical complexity of the analysis and presentation of the results of diagnostic research is a subject of recurring debate, on which no agreement exists. Undoubtedly, an excess of statistical approaches to the evaluation of diagnostic procedures constitutes a barrier to clinical application of research findings and for the adoption of explicit quantitative procedures in the evaluation of the diagnosis of specific patients. However, I believe it is necessary to make a clear distinction between the practical application of diagnostic research findings, and the objective of making the explicit quantification of diagnostic probabilities a more frequent occurrence in medical practice. In the two instances, it may be necessary to use complex mathematical approaches, but there is little question that to achieve the second objective it will be essential to simplify the mathematics involved. If the mathematics are not made plain, practising clinicians will continue to be unwilling to implement the different indices and probabilities in their daily tasks. Simplification could be achieved either as Choi indicates—using a simple model-user interface⁶ or by other means.

Even though the analytical procedures may often remain complex, the practical implications of the findings need not necessarily be so. The articles by Moons and Grobbee and by Brenner *et al*⁷⁻⁹ point out areas where application of multivariate analysis procedures is essential, such as the evaluation of the effect of various covariables on the sensitivity and specificity of a specific test, or the analysis of

the value of a diagnostic test according to relevant clinical strata. The results of these analyses may imply, for example, that a test is recommended in the diagnosis of a certain target disease in specific strata of age or gender; or, conversely, that its application is unwarranted in the presence of specific comorbidity. Therefore, the use of complex analytical methods does not necessarily entail that the implications of the research findings are difficult to turn into practical recommendations. It is quite another thing to expect clinicians to use multivariate models to evaluate specific patients when they are reluctant to apply much simpler indices.

Controversies on the dominance of statistical approaches may have overlooked what in my view is a key point. In this field, much research and scientific literature have been based on false assumptions, such as the alleged immutability of the sensitivity and specificity of diagnostic tests. As Knottnerus very rightly points out,¹⁰ the most important factor is the definition of the research question and the study design. Surprisingly as it may seem, in the diagnostic arena, it is rare for the research question to be clearly defined, and for the design to correspond with the objective. By contrast, in therapeutic research, it is customary for the research question to address unobtrusively a highly topical and relevant clinical question, that is, to tackle a need for specific knowledge in order to make clinical practice more effective. This is much less often so in the field of diagnostic research, where the research objective seldom fits a real need for knowledge for an evidence based diagnosis. Perhaps this reflects the fact that the therapeutic decision is generally the sole responsibility of the doctor in charge of the patient; whereas diagnostic tests are usually performed by a wide range of professionals, most of whom are not directly in charge of the patient and, consequently, do not follow up the patient's clinical course. This distinction, performance of diagnostic tests versus comprehensive patient care, is mirrored in various aspects of the research.

In my opinion, in most studies on diagnosis the key limitation is the composition of the population included: the subjects studied rarely constitute a group of patients with a specific diagnostic problem, the so called "indicated" population.¹² By this I mean a consecutive series of patients with a specific complaint to whom, after recording their clinical history and carrying out a clinical examination, a specific diagnostic question is asked; a problem that must be resolved, either the confirmation of a suspicion or ruling out of an improbable diagnostic hypothesis, etc—in other words, a stage IV study, according to Feinstein's terminology.¹³

Diagnostic laboratories, services of diagnosis by imaging techniques, nuclear medicine, and the like rarely have access to consecutive series of patients who are homogeneous as far as the clinical problem is concerned. Therefore, in order to study relevant diagnostic problems, they should work in perfect coordination with the doctors who are in charge of the patients.

The scarcity of stage IV studies has several undesirable consequences, some of which are pointed out in the articles published in this issue.⁵⁻¹¹ In stage IV studies, one or more diagnostic tests are evaluated in consecutive patients in whom there is a specific diagnostic uncertainty. This is the type of research that can really provide information applicable to clinical practice. Any other more selected population studied just provides preliminary information that is rarely applicable to clinical decision making. Therefore, when a diagnostic test is evaluated by comparing a sample of patients—with one or more specific diseases—with a group of healthy volunteers, the results can only give information about the future potential of the test and the indication of whether or not to proceed with additional investigations. At this stage, no matter how sophisticated the mathematical analysis may be, no matter how detailed the analysis of covariables, it is impossible for the results to be applicable to an actual clinical population. Furthermore, their inclusion in meta-analyses is usually unadvisable. Colditz rightly reminds us that useful methodologies have been developed to overcome some of the problems usually encountered in diagnostic research.¹¹ However, no amount of methodological inventiveness can enable the findings of diagnostic research to be applied to specific individuals when the population studied is not an actual clinical population.

Though related, another problem is that much diagnostic research remains in these preliminary stages, where the starting point is the disease and not the diagnostic problem that arises in clinical practice. As much research is not based on the clinical problems arising in doctor-patient encounters, the diagnostic procedure and the part the diagnostic tests are supposed to play are not considered. Consequently, the specific role attributed to each test in each clinical context is often forgotten, and no research is done into what Knottnerus calls the “doctor’s black box” of diagnostic decision making.¹⁰ Therefore, as Feinstein emphasises,⁹ qualitative research and the active participation of clinicians are essential. Such research may help to determine, with greater precision, the different functions that clinicians attribute to or expect from each diagnostic test, and to improve the definition of the research questions.

Some research questions may require that we bring on board new study designs. Moreover, as Moons and Grobbee indicate, it will be necessary to find innovative ways to measure and include the doctor’s perceptions as additional tests in diagnostic practice.⁹ We should also bear in mind that experimental designs are seldom used in diagnostic research. Despite some very attractive examples,¹⁴ investigators are reluctant to use this type of designs. Only in the evaluation of screening tests have clinical trials been used with any frequency. This may explain why, in the literature on clinical evidence, screening is the only section on diagnosis of any length.

Moons and Grobbee provide some useful indications to determine whether it is necessary or not to evaluate diagnostic tests by means of follow up studies or clinical trials, instead of by using cross sectional studies. Some points raised by Feinstein might also be added here.⁵ The part played by diagnostic tests is not limited to the consideration of a single disease. Diagnostic tests, in particular imaging techniques, often provide a great deal of information. The impact of this additional information on the management of the patient and on the outcome of the process in terms of health has seldom been evaluated. The introduction of a specific diagnostic test, because it diagnoses a treatable disease better than the usual test, may occur along with a series of cointerventions that could have different effects, including iatrogenia, in terms of health. Consequently, the need for clinical trials should be carefully assessed.

Certainly, sometimes it may be unnecessary to resort to complex designs to evaluate diagnostic questions. The follow up of patients who attend primary care centres for specific complaints may be sufficient to provide information of great use in their management. It is essentially a matter of forming cohorts of patients who are homogeneous in terms of demand for medical help; for example, patients who present to the doctor because they have dizzy spells. Once the follow up is completed and the relevant information collected, we may find that we have simple rules on which to base decisions in an important percentage of patients characterised by different variables (such as age and sex), for instance, in the groups in which the problem was found to resolve spontaneously in a short time. The proper, comprehensive follow up of these cohorts until the final diagnosis is made or the condition resolves spontaneously also provides the diagnostic probabilities and pre-test prognoses for decision making.

As if the diagnostic field was not already complex enough, we are now

faced with the added complexity of genetic testing, which entails challenges and difficulties of enormous significance, well covered by Coughlin.⁸ The methodological defects shown in a series of papers on DNA research,¹⁵ led Feinstein to warn of the danger of iatrogenia. This is a time when we may witness spectacular advances in the diagnostic field, perhaps even more impressive than those in the therapeutic field. Spectacular at least from a technological and mechanistic viewpoint. Whether they will also prove to be useful to diagnose “common” patients and to improve their outcomes, that remains to be seen. The need to evaluate the actual clinical impact of the new genomic technologies is hence enormous. Professionals involved in clinical epidemiology have a particular responsibility to promote such evaluation, so that efforts invested in diagnostic research are not misguided. If creative, rigorous, and clinically meaningful research methodologies are applied to assess the diagnostic usefulness of genomic discoveries, the legacy of Alvan Feinstein will continue to thrive.

ACKNOWLEDGEMENTS

Miquel Porta Serra has contributed to this editorial with useful comments. I am also grateful to Carlos Alvarez-Dardet for his help in the revision of the manuscript.

J Epidemiol Community Health
2002;**56**:323–325

Author's affiliation

I Hernández-Aguado, Department of Public Health, Facultad de Medicina, Universidad Miguel Hernández, Spain

Correspondence to: Professor I Hernández-Aguado, Department of Public Health, Facultad de Medicina, Campus de San Juan, Universidad Miguel Hernández, 03551 San Juan de Alicante, Spain; ihernandez@umh.es

REFERENCES

- 1 Sheps SB, Schechter MT. The assessment of diagnostic tests. A survey of current medical research. *JAMA* 1984;**252**:2318–22.
- 2 Reid ML, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic research. Getting better but still not good. *JAMA* 1995;**274**:645–51.
- 3 Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.
- 4 Reid CM, 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.
- 5 Feinstein AR. Misguided efforts and future challenges for research on “diagnostic tests”. *J Epidemiol Community Health* 2002;**56**:330–2.
- 6 Choi BCK. Future challenges for diagnostic research: striking a balance between simplicity and complexity. *J Epidemiol Community Health* 2002;**56**:334–5.

- 7 **Brenner H**, Stürmer T, Gefeller O. The need for expanding and re-focusing of statistical approaches in diagnostic research. *J Epidemiol Community Health* 2002;**56**:338–9.
- 8 **Coughlin SS**. Future challenges for research on diagnostic tests: genetic tests and disease prevention. *J Epidemiol Community Health* 2002;**56**:335–6.
- 9 **Moons KGM**, Grobbee DE. Diagnostic studies as multivariable, prediction research. *J Epidemiol Community Health* 2002;**56**:337–8.
- 10 **Knottnerus JA**. Challenges in diagnostic research. *J Epidemiol Community Health* 2002;**56**:340–1.
- 11 **Colditz GA**. Improving standards of medical and public health research. *J Epidemiol Community Health* 2002;**56**:333–4.
- 12 **Van der Schouw YT**, Van Dijk R, Verbeek ALM. Problems in selecting the adequate patient population from existing data files for assessment studies of new diagnostic tests. *J Clin Epidemiol* 1995;**48**:417–22.
- 13 **Feinstein AR**. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: WB Saunders, 1985.
- 14 **Douglas CD**, Macpherson NE, Davidson PM, *et al*. Randomised controlled trial of ultrasonography in diagnosis of acute appendicitis, incorporating the Alvarado score. *BMJ* 2000;**321**:1–7.
- 15 **Reid CM**, 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.

APHORISMS OF THE MONTH

SOCIAL MEDICINE

This month we publish two aphorisms.

"The House of Commons is the pharmacy of social medicine."

"Man is the greatest pathogen known to man."

F A E Crew, cited by John Pemberton (see page 342).

John R Ashton, CBE