WHAT EVIDENCE IS THERE FOR BIOCHEMICAL TESTING?

Andrea Rita Horvath

Department of Clinical Chemistry, University of Szeged,
POB 482, H-6701 Szeged, Hungary

INTRODUCTION

Evidence based medicine has become accepted practice in the provision of modern healthcare. However, the approach has had limited impact on laboratory medicine compared with that in other clinical disciplines. The appropriate use of diagnostic tools in clinical decision-making is of crucial importance, as further management of patients depends on it. But many diagnostic investigations have never been subjected to systematic evaluation using modern standards of clinical epidemiology, especially when compared with the rigorous approval of therapeutic drugs. Furthermore, despite the remarkable achievements in the analytical performance of tests, less attention has been paid to outcomes research to evaluate the diagnostic impact and clinical utility of laboratory investigations (1). The lack of good quality research in the field contributes not only to inappropriate utilisation of laboratory services, which might harm patients, but also to wasting significant resources. Clinical laboratories may therefore be seen by some managers as cost centres, rather than – what they should be – resource centres.

Patients and society expect physicians to base their approach to any type of clinical problem on informed diagnostic reasoning. Informed diagnostics means that clinicians understand and readily apply the principles of diagnostic decision making, which include an estimate of the pre-test probability/prevalence of diseases and information about the performance characteristics and discriminatory power of the applied investigations. Interpretation of results is also concerned with supporting informed clinical decisions by synthesizing and converting analytical and research information to knowledge that, together with clinical experience and patients' preferences, can be transformed to the wisdom needed in making individual choices for patients (2).

If laboratory medicine professionals wish to offer high quality, efficacious and effective services, it is important that the pre-analytical, analytical, and post-analytical phases of the diagnostic process are based, as far as possible, on the best available scientific evidence.

In this paper, I would like to discuss

- the definition and aims of evidence-based laboratory medicine (EBLM),
- what kind of evidence we need,
- what kind of evidence we have in laboratory medicine,
- and what we can do about improving the current situation.

DEFINITION AND AIMS OF EVIDENCE-BASED LABORATORY MEDICINE

Evidence-Based Medicine (EBM) needs to be constantly promoted because medicine based on tradition, false conviction, superstition, or unjustified authority, has not yet disappeared, and will perhaps exist as long as

medicine itself remains what it is: a combination of science and humanities (3).

But what is EBLM, and how can it be used in facing the modern challenges of diagnostic services in the provision of health care? According to Sackett et al. EBM is about improving decisions on the diagnosis and treatment of individual patients, by adding as much scientific reasoning as possible to the art of medical practice (4). In other words, like the Chain Bridge of Budapest linking the old historical town of Buda to the new town of Pest, EBM forms a bridge between old knowledge or experience and new knowledge coming from systematic research. Rephrasing this definition:

Evidence-based laboratory medicine integrates into clinical decision-making the best research evidence for the use of laboratory tests with the clinical expertise of the physician and the needs, expectations and concerns of the patient.

The aims of EBLM in the pre-analytical phase are:

- to eliminate poor or useless tests before they become widely available (stop starting),

- to remove old tests with no proven benefit from the laboratory's repertoire (start stopping) (5), and

- to introduce new tests, if evidence proves their effectiveness (start starting or stop stopping).

- In the post-analytical phase to:

- improve the quality and clinical impact of diagnostic test information – diagnostic accuracy,

- improve patient outcomes – clinical effectiveness

- reduce health care costs – cost effectiveness

But is it also, what clinicians want from us? When the editor of Bandolier asked David Sackett, a respected authority in the field of EBM, what he needed most from our profession, he expressed three desires, all related to the pre- and post-analytical activities of laboratories:

1. "to be able to discuss a patient's illness with a colleague;"

2. "to be able to abandon reporting of normal ranges" in favour of decisions limits;

3. "to have evidence available to support the validity, importance and clinical usefulness of biochemical tests" (6).

In my view, out of these "three wishes" the third point is perhaps the most important one in terms of EBLM.

WHAT KIND OF EVIDENCE DO WE NEED IN LABORATORY MEDICINE?

In order to achieve the aims of both the clinical and laboratory professions, what kind of evidence do we actually need in laboratory medicine? We need:

- high quality, reliable evidence;

- evidence that supports the pre-analytical, analytical and post-analytical activities of laboratories;

- evidence that can be easily interpreted, accessed and used at the point of service delivery and for making clinical decisions.

But what do we call evidence? Sackett et al. termed the best available external evidence in medicine as clinically relevant research, often from the basic sciences of medicine, but especially from patient-centred clinical research, into the accuracy and precision of diagnostic tests, the power of prognostic markers and the efficacy and safety of therapeutic, rehabilitative and preventive regimens (4). According to the definition of the Committee on Evidence-Based Laboratory Medicine (C-EBLM) of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), evidence in laboratory medicine is (and the words in italics have particular importance): systematically compiled and critically appraised information, preferably coming from well-designed primary research studies, to answer a specific question on diagnosis, differential diagnosis, screening, monitoring and prognosis, which provides an explicit framework for making informed medical decisions.

WHAT KIND OF EVIDENCE DO WE HAVE IN LABORATORY MEDICINE?

The criteria of what kind of evidence is needed have been defined. Now, let's see whether we have systematically compiled, critically-appraised, high-quality and reliable information, from well-designed primary research studies in laboratory medicine?

May I ask the reader a question? If I told you, that trials on a new treatment for multiple sclerosis were shown to have massively biased results, would you decide to prescribe that drug? My personal view is that the majority of the audience would hold the opinion: "We would hesitate to base major decisions on trials of treatment that were known to have massively biased results, and yet" - quoting the editor of Bandolier (2) – "for diagnostic testing that's usually all we have". "The evidence-base for effective...diagnosis is rather thin... and if one thinks for a moment that effective treatment depends on effective diagnosis, it makes one a bit concerned about the efficiency of our health services."

Why do we not have high quality evidence in laboratory medicine?

There are several reasons for not having high quality and reliable evidence in laboratory medicine.

The gold standard problem

One of the major problems is the lack or the inappropriate application of the gold standard (7-8). The accuracy of a diagnostic or screening test should be evaluated by comparing its results with a reference criterion. The reference test may be a single test, a combination of different tests, or the clinical follow-up of patients. Ideally, the same reference test should be applied to all study participants and its result should be interpreted without the knowledge of the examined test and vice versa, in order to avoid different forms of verification biases. The problem is that we often do not have any gold standard at all, or even if so, it is not a "true" gold standard, and has its own uncertainty of estimations (e.g. a histological finding of appendicitis is clinically insignificant, if the patient has no clinical symptoms and the condition goes silent). Often the new test is more advanced than the reference test (e.g. due to change in technology), or the reference test is too expensive, or invasive which may limit its use, either because it cannot be assessed independently or blindly, or can cause harm to patients, and thus its use is unethical.

Problems related to inappropriate design of primary studies

In addition to the gold standard problem, there are several methodological traps related to study design, which should be avoided if primary researchers wish to produce high quality and reliable evidence. It has been shown that the diagnostic accuracy of many laboratory tests are seriously overestimated due to different forms of biases related to inappropriate study design, which affect both the internal and external validity of results (9). For instance, the optimal design for assessing the diagnostic accuracy of a test is considered to be a prospective blind comparison of the index and reference test in a consecutive series of randomly selected patients, from a relevant clinical population, suspected of having the disease. For example, a group of patients with a clinical suspicion of prostate cancer go through PSA testing, and groups of patients with both negative and positive results go through the reference tests, which are histology of biopsies and clinical follow-up to assess true and false positive and negative rates.

Spectrum bias is another common problem, resulting from the inappropriate selection of study patients, which threatens both the internal and external validity of diagnostic studies (7). In test evaluation studies the spectrum of pathological and clinical features should reflect the spectrum of setting where the test is meant to be used. Diagnostic accuracy, in terms of sensitivity and specificity can be overestimated, if the test is evaluated in a group of patients already known to have fairly advanced disease (CASE), compared to a separate group of perfectly healthy individuals (CONTROL), rather than in a clinically more relevant population. This typical case-control design results in spectrum bias and overestimates the diagnostic performance of a test (9).

The spectrum of patients influences not only the internal, but also the external validity, i.e. the transferability of test results. For example, a test may appear more sensitive but less specific (i.e. true negative rate falls and false positive rate increases) if evaluated in the tertiary care setting with more advanced disease and co-morbidities. The estimates of diagnostic accuracy vary considerably along the referral pathway and the results of small studies are applicable only to the setting where the study was performed. It is therefore essential to provide accurate and detailed information about the setting of care, spectrum of disease and patient characteristics, but once again this information is often not reported adequately in current medical literature. In addition, much larger studies with patient populations covering the whole spectrum of disease are needed

to ensure that estimates of test accuracy travel (10).

Verification bias is another common problem. Verification bias looms when the decision to perform the reference test is based on the result of the experimental test. In many diagnostic studies with an invasive reference test, mostly those with positive test results go through the reference test, while those with a negative result either do not get the reference test at all (partial verification, or workup selection bias) or get different, less thorough reference tests, e.g. follow up, than the positive cases (differential verification or workup detection bias). This latter case will lead to misclassification of false negatives as true negatives and will bias both sensitivity and specificity upwards (11).

Review bias may occur if the reference test is interpreted with knowledge of the results of the experimental test or vice versa. This may lead to overestimation of both sensitivity and specificity, especially if the interpretation of test results is subjective.

Lijmer et al. have provided empirical evidence and have quantified the effects of all these design related biases as relative DORs (9). In this important study, they demonstrated that:

- using a case-control design tends to overestimate the DOR 3-fold compared with studies with a clinical cohort.

- differential verification bias results in a 2-fold overestimation of accuracy compared to studies that used one reference test.

- no blinding resulted in an approx. 30% overestimation of results compared to studies with proper blinding.

- in studies where the examined test or the study population were not sufficiently described, there was an overestimation of accuracy by 70% and 40%, respectively.

- studies that did not report the cut-off values of the reference test underestimated the accuracy of the examined test by 30%.

It has also been demonstrated, however, that the quality of diagnostic studies is improving, but still many suffer from methodological flaws, such as poor reporting of the spectrum and subgroups of study patients, lack of avoidance of reviewer bias and poor reproducibility of studies (12).

Lack of outcome studies in laboratory medicine

The ultimate purpose of laboratory medicine is improving clinical outcomes and prognosis. If so, then the efficacy of diagnostic interventions and laboratory monitoring should ideally be assessed in randomised trials. Unfortunately it is rarely feasible to assess the effect of diagnostic tests in randomised trials (13); a recent search in MEDLINE in April 2003 returned just 28 citations. Studying laboratory-related outcomes is difficult due to the methodological problems of defining and measuring hard and soft measures of outcomes (14). Technology often moves too fast, or patients receive alternative diagnostic and therapeutic interventions that make the assessment of the correlation between testing and outcome rather difficult.

Problems of systematic reviewing in laboratory medicine

Systematic reviews and meta-analyses are considered the highest level of evidence; however, due to numerous methodological problems and shortcomings of the primary literature we lack such high quality evidence in laboratory medicine. Oosterhuis et al. showed that out of 23 systemic reviews in laboratory medicine none met six basic quality criteria, and only 48% met half of the criteria (15). Similarly, we lack good quality technology assessments, and systematically collected prevalence/pre-test probability data, or good evidence on diagnostic thresholds. All these problems are related to the heterogeneity of primary research, mainly due to the lack of internationally agreed methodological standards for designing, conducting and reporting of primary studies, and systematic reviews in diagnostics.

In laboratory medicine, like elsewhere, it has become simply impossible for the individual to read, critically evaluate and synthesise the current medical literature. There has been increasing focus on formal methods of systematically reviewing studies to produce explicitly formulated, reproducible and up-to-date summaries of the effects of health care interventions. However, these efforts have so far been largely confined to the evaluation of efficacy and cost-effectiveness of therapeutic and preventive interventions. This is illustrated in the Figure 1, which shows the yearly number of systematic reviews and meta-analyses of randomised

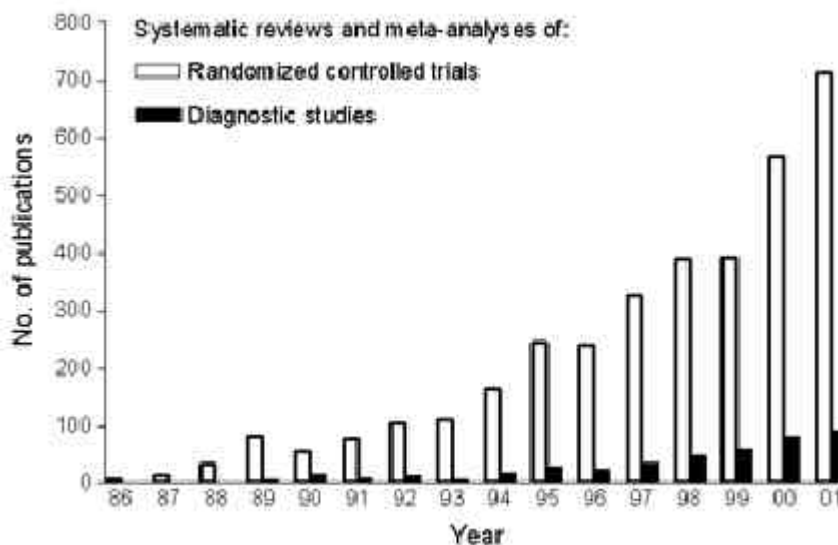controlled trials and diagnostic test evaluation studies in 1986 to 2001 (16).

Fig. 1 Number of systematic reviews and meta-analyses

Several systematic review databases have been established for diagnostics, such as the MEDION (http://www.hag.unimaas.nl/Internationalisering/onderzoek/Cochrane/2002/medion/ medionSearch1.asp) and our Committee's database (http://www.ckchl-mb.nl/ifcc). The DARE database also contains a number of critically appraised diagnostic overviews (http://agatha.york.ac.uk/darehp.htm). A new initiative, called the Bayes Library, will be discussed shortly (11). However, Bandolier Extra has recently expressed a very pessimistic view on systematic reviewing of diagnostic tests and called it "a complete waste of time" (2). This view is not generally shared, as the aim of doing systematic reviews is not only to produce high quality evidence on a given topic, but also to identify gaps in our knowledge and to promote well-designed research in the area. Also, systematic reviewing in laboratory medicine is educational in developing critical appraisal skills and thus represents a learning curve towards informed decision making and practicing EBLM.

It is acknowledged, however, that systematic reviewing and thus the production of the highest quality evidence in laboratory medicine are not only limited by the poor quality of primary studies, but also by different forms of publication and related reporting biases (17). It is well known that only a proportion of all studies conducted is ever published and reviewed (Figure 2).
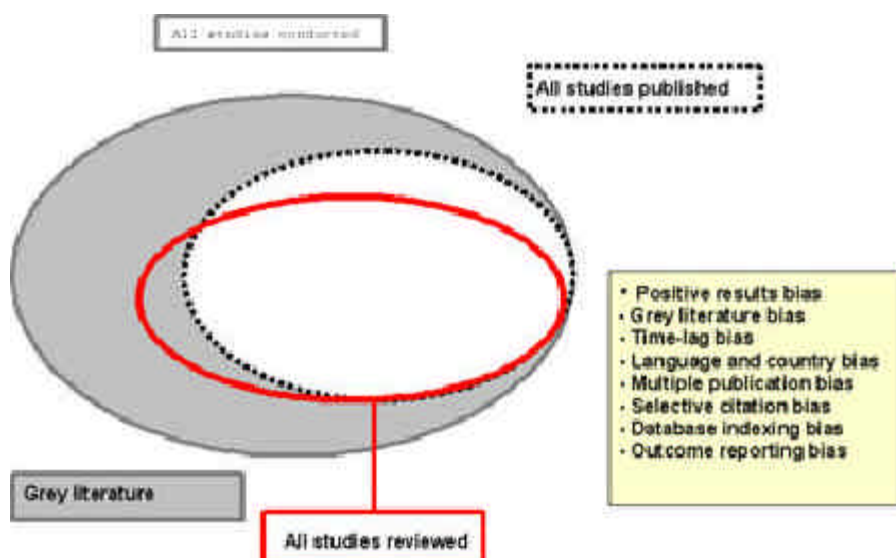

Fig 2. Publication and other reporting biases (17)

There are several forms of publication biases (17):

- positive results bias is when research with positive results is more likely to be published than with negative results;

- grey literature bias is when many studies with significant, but negative results remain unpublished;

- time lag bias, when publication is delayed by several months or years;

- language and country bias refers to when significant results are more likely to be published in English than in other languages;

- duplicate or multiple publications occur when original findings or parts of their results are published more than once, which could distort the conclusions of a systematic review;

- selective citation of references means that positive results are more likely to be cited than negative ones;

- database indexing bias, similar to selective citation, means that positive results are indexed on databases more often than negative results;

- selective reporting of measured outcomes occurs when studies investigating multiple outcomes tend to report only those which show the most favourable results.

How big is the problem? If the published studies represent a biased sample of all studies that have been conducted, the results of the literature review, by definition, will also be misleading. In the field of diagnostics the rate of publication bias is unknown, but we assume that it is an even bigger problem than in the area of therapeutic research, due to the lack of registers of ongoing or accomplished unpublished diagnostic studies and the difficulties of accessing the "grey literature". Publishers, researchers, pressure from academic institutions and from industry are all responsible for this hardly controllable situation.

Another problem of systematic reviews is whether the data from different primary studies can be synthesized and pooled into a summary estimate of diagnostic accuracy (18). Statistical pooling of likelihood ratios in the form of a meta-analysis should be carried out with great care and only if the patient populations and tests used across studies are homogeneous (19). However, such likelihood:ratio plots, if done properly, allow the rapid assessment of the power of a test in ruling in or out the diagnosis of a disease.

HOW CAN WE IMPROVE THE CURRENT SITUATION?

Having seen the difficulties of producing the evidence, how can we provide better quality and more reliable evidence, and what should we do to avoid the numerous pitfalls mentioned:

- Firstly, and most importantly, we need methodological standards for designing and reporting primary studies of diagnostic accuracy (20), and based on these, better quality primary research in the future.

- Similarly, we need methodological standards for systematic reviewing in laboratory medicine and thus better systematic reviews/meta-analyses of data on diagnostic accuracy.

- In addition, we need high quality outcomes research in laboratory medicine.

- Furthermore, we need to make research evidence easily understood and accessible at the point of clinical decisions.

The STARD initiative (Standards for Reporting of Diagnostic Accuracy)

Referring to the first point above, the STARD group has recently published the recommended procedures for designing reliable studies of diagnostic accuracy (21). Their checklist sets high standards for describing both the methodology of the study and also its results (22). It is expected that the STARD checklist will be widely adapted by medical journal editors and researchers, and thus will contribute to better quality primary studies in the future.

New initiatives for better systematic reviews in laboratory medicine

The principles of systematic reviewing in diagnostics have been described by several authors and also by members of C-EBLM, and provide useful tools to many methodological issues that need to be addressed if one starts the hard work of critical appraisal and systematic reviewing of the literature (1, 16, 18, 23). One new approach is the so-called Bayes Library, an international collaborative project initiated by a Swiss group of

primary physicians and epidemiologists, in association with C-EBLM, and the Cochrane Collaboration. This initiative has elaborated and currently pilot tests the "Bayes Handbook", including a detailed critical appraisal checklist, for systematically reviewing the primary literature of diagnostic studies (11). The Bayes Library intends to be a systematically compiled database of standardised and critically appraised information on the characteristics and discriminatory power (sensitivity, specificity, likelihood ratios, etc.) of tests used in health care. In addition, the database will contain information on the prevalence of diseases and conditions in different settings and patient groups, and provide a user-friendly interface and search engine. A similar new initiative is being developed by the Screening and Diagnosis Methods Group of the Cochrane Collaboration, and it is foreseen that the Cochrane Library will publish diagnostic reviews in the future (24).

The evidence should support the whole laboratory process in order to improve laboratory-related outcomes

Once we have the systems to produce high quality evidence in place, what should the evidence be for in laboratory practice? Ultimately, every single step in the whole laboratory process aims at improving laboratory-related and patient outcomes. Thus the evidence produced should support the whole laboratory process, including pre-analytical, analytical and post-analytical activities. Let's see how and what are the key targets for EBLM?

The role of EBLM in the pre-analytical phase

One of the most important aspects of the pre-analytical phase is the selection of the right test(s) for the right patient and at the right time. Several factors influence our decision on whether a test should be ordered or not. One of the most important of these is the type of the clinical question asked and the prevalence of the condition in different care settings in relation to that question. In the case of diagnosis, it is a general rule that, if the prevalence:pre-test probability of the condition is either too low or too high, it is unnecessary to request a battery of laboratory tests, as they will not add much to the post-test probability, i.e. the diagnosis of the disease. Other factors that influence test selection are related to events in both the analytical and post-analytical phases, such as the technical performance, the diagnostic accuracy of laboratory investigations, the clinical outcome and the organisational impact of testing, the costs, and burden to the patient (25). Insufficient evidence to support the role of testing can result in early diffusion of the technology by enthusiasts. Even if the utility of the test proves to be less significant by future research, the test which became routine practice will be difficult to eliminate from the laboratory's repertoire.

From the above it follows that EBLM can support the pre-analytical phase by providing evidence for rational test selection and ordering. To that end, data should be collected on disease prevalence/pre-test probability (e.g. the Bayes Library intends to offer such information). Technology appraisal of diagnostic interventions together with economic evaluation of the impact of testing and evidence-based diagnostic guideline recommendations on test selection, are also useful tools in this respect. The activities of the C-EBLM in the field of guidelines will be discussed later.

The role of EBLM in the analytical phase

At first sight, analytical quality itself does not seem to have very much to do with EBLM because analytical performance of tests is supposed to be based on basic sciences which, in theory, are evidence- or research-based by definition. However, method performance goals are either established on the basis of biological variation or on medical decision limits, and data gathered for achieving these goals may not always be evidence-based (26). This is why EBLM in the analytical phase can contribute to the scientific establishment of method performance specifications, and thus to the better evaluation of diagnostic accuracy of laboratory investigations.

The role of EBLM in the post-analytical phase

The aims of EBLM in the post-analytical phase are to improve the quality of diagnostic test information. EBLM can assist clinicians both in the interpretation and clinical utilization of laboratory results. There are several factors that influence the use of evidence in the post-analytical phase. First, it is the very nature of the evidence on the diagnostic utility of laboratory investigations, the pitfalls of which we discussed before. Secondly, how can we deliver this information to clinicians in a meaningful way? This post-analytical activity should clearly be concerned with teaching both clinical and laboratory staff on how to use the evidence when interpreting data and making informed diagnostic decisions.

Suppose we are dealing with a 70-year old patient who has a 50% pre-test probability of iron-deficiency

anaemia. A systematic review shows you the probability of anaemia expressed as likelihood ratios (LR) at different ferritin concentrations. Our patient, with an otherwise 'normal' ferritin of 30 ug/L, has a positive LR of 4.8, i.e. an intermediate high probability of being iron deficient (27). Using the Fagan's nomogram and the LR of 4.8 as a probability modifier, one can quickly estimate the post-test probability of anaemia, which becomes nearly 85%, in a patient with otherwise normal ferritin (28). This example perhaps explains better, why David Sackett asked for abandoning the reference ranges!

(Figure 3 here)

To build up a database of such diagnostic accuracy data, we not only need more and better research to be carried out, but it is also important that members of the profession collect and synthesize the results and experience, accumulated over decades, and present them in a meaningful way of LRs, or ROC curves. Such a database, in our example, could not only show that ferritin is a useful test in diagnosing anaemia, but also that it is, for example, a much better test than transferrin saturation in diagnosing iron deficiency.

It is the laboratory's responsibility to interpret the evidence and make it accessible at the point of clinical decisions

While we, laboratorians, are "excited" about data such as sensitivity, specificity, LRs and ROC curves, are our clinical colleagues just as well...? Can they easily interpret and use this information in practice? An interesting study in 1998 asked groups of about 50 physicians and surgeons how they used diagnostic tests. The results showed that very few (1-3%) knew or used Bayesian methods or ROC curves or LRs (29). If asked, what most doctors want is not LRs or sensitivity and specificity or predictive values...but simple answers, preferably at the bedside, of whether the patient with a given test result has or does not have a condition in question.

Interpreting the evidence

A recent example quoted in the British Medical Journal has just confirmed this need very convincingly (30). General practitioners in a regular continuing medical education session were given the following case: "Prevalence of uterine cancer in all women with abnormal uterine bleeding is 10%. What is the probability of uterine cancer in a 65 year old woman with abnormal uterine bleeding with the following result of a transvaginal ultrasound scan?"

- The first set of results was given as: Transvaginal ultrasound showed a pathological result compatible with cancer. Based on this information, may I ask the reader, how high do you think the probability of cancer is?

- In the second set, the sensitivity and specificity of the test were also given as 80% and 60%, respectively. In view of this information, how high do you think the probability of cancer is?

- In the third set, it was explained in plain language that a positive result is obtained twice as frequently in women with cancer than in women without the disease which, in other words, means that the positive LR is 2. How high is the probability now? Would you change your previous estimate(s)?

When the responses of doctors were converted to likelihood of disease, the group that was given no information on the diagnostic accuracy of transvaginal ultrasound gave a high probability of cancer, with a LR of 9. When they were told the values of sensitivity and specificity of the test, the probability of cancer dropped to 6, and when the diagnostic accuracy of the test was explained in simple non-technical terms, it dropped significantly to 3, which was very close to the true estimate and to guideline recommendations (30). So, well-trained physicians, when presented with a positive test result alone, grossly overestimate the diagnostic power of tests.

Therefore, it is the responsibility of laboratory professionals to express and interpret the diagnostic accuracy of test results in clinically meaningful ways. However, this requires that laboratory staff has access to relevant patient data and knows the exact reason for testing. Unfortunately, it is difficult to meet these requirements for most of the tests performed in a routine department of clinical biochemistry, with many thousands of test results per day, therefore computerised decision support tools directly linked to clinical data and evidence-based information could assist in this task.

Evidence-based guideline recommendations in laboratory medicine

It is also the laboratory's responsibility, together with clinicians and patients, to transform evidence-based

information to knowledge, and to make easily understood recommendations (2). To deal with the ever-expanding body of medical information, there has been a widespread move towards developing clinical practice guidelines which are increasingly viewed as a mechanism of distributing knowledge to practitioners. However, the approaches used for therapeutic guidelines are diverse, and neither the methods nor the grading systems for recommendations can be fully adapted to laboratory medicine.

Therefore, there is a need for methodological standards and toolboxes for the development of evidence based recommendations for the use of laboratory investigations in screening, diagnosis and monitoring. The C-EBLM has recently prepared a document which outlines the general principles, methods and processes of the development and critical appraisal of evidence based diagnostic guidelines, which will support guideline development teams, and hopefully contribute to better quality guidelines in laboratory medicine in the future (31).

Putting research evidence into practice

Developing evidence-based guideline recommendations is a hard task, especially for the lack of high quality systematic reviews in laboratory medicine. However, the role of EBLM in the post-analytical phase does not end here. It has been demonstrated that even when sufficient evidence and authorative guidelines are available, changing behaviour of physicians is difficult. Empirical evidence suggests that passive dissemination of guidelines is not enough; it needs to be combined with a multifaceted and individualized dissemination and implementation strategy (32). It has been shown that education, outreach visits, individually tailored academic detailing, electronic reminder systems, feedback on performance, and participation of doctors in post-analytical quality assessment or case interpretation programmes, and clinical audit schemes are much more efficient in getting research evidence into practice.

As it was already said by Wilson: "It is much harder to 'run' a constitution than to frame one". Or more bluntly by Schumpeter: "It was not enough to produce satisfactory soap, it was also necessary to induce people to wash." So how can we ascertain that they wash...? If we want people to do something, it has to be made simple, easy, and the users must see the advantage of doing it. To help physicians utilize laboratory services efficiently, the supporting evidence or guidelines should be made easily accessible at the point of clinical decisions, preferably directly linked to patient data. Automation/information technology can provide means to integrate decision support into patient care. Patient data can directly be linked to evidence-based guidelines, which could provide graded recommendations on the prevalence, aetiology, symptoms and management of, for example, hyperkalaemia. Such systems have been developed, for example, by Jonathan Kay in Oxford and in close collaboration with the Oxford Clinical Intranet project, in Hungary (http://oxmedinfo.jr2.ox.ac.uk/; **http://tudor.szote.u-szeged.hu**) (33).

But is it only clinicians who should learn how to use the "soap"? Utilization of laboratory services is driven by many factors, not just doctors and their professional attitudes. Internal drivers are the demands of patients', nurses, and other health care staff practising defensive medicine, for example. External drivers are scientists enthused by new technology, industry willing to sell, insurance companies willing to buy services, and even the media which picks up new and sensational "discoveries". Due to the broad scope, the diversity and complexity of methodological, technical, organisational and economical issues, and the numerous tasks associated with EBLM, improvements in this field can only be achieved by a multidisciplinary and world-wide collaboration. To achieve EBLM and thus higher efficiency of laboratory services, more effective collaboration between laboratory professionals, clinicians, epidemiologists, biostatisticians, industry, quality-, technology assessment- and government agencies, and an international harmonization of these approaches are needed.

ACKNOWLEDGEMENTS

REFERENCES

1. Sandberg S, Oosterhuis W, Freedman D, Kawai T. Systematic Reviewing in Laboratory Medicine. Position Paper from the IFCC Committee on Systematic Reviewing in Laboratory Medicine. JIFCC 1997; 9(4): 154-155.

2. Anonymous. Evidence and Diagnostics. Bandolier Extra, 2002; February, 1-9. http://www.ebandolier.com

3. Watine J, Borgstein J. Evidence-based illiteracy or illiterate evidence. Lancet 2000; 356(9230): 684 (downloadable at: http:/www.thelancet.com).

4. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine. How to Practice and Teach EBM. Churchill Livingston, London. 1997. pp250.

5. Gray, MJA. In: Evidence-based Healthcare. How to Make Health Policy and Management Decisions. Churchill Livingstone, London, 1997. pp23.

6. Moore RA. Evidence-based clinical biochemistry. Ann Clin Biochem 1997; 34: 3-7.

7. Haynes RB. Keeping up to date with the best evidence concerning diagnostic tests. In: Black ER, Bordley DR, Tape TG, Panzer RJ, eds. Diagnostic Strategies for common medical problems. 2nd ed. Philadelphia: American College of Physicians, 1999; pp37-45.

8. Knottnerus JA, vanWeel C. General Introduction: evaluation of diagnostic procedures. In: Knottnerus JA, ed. The Evidence Base of Clinical Diagnosis, London:BMJ Books, 2002; pp1-17.

9. Lijmer, J.G., Mol, B.W., Heisterkamp, S., et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282(11): 1061-1066.

10. Irwig L, Bossuyt P, Glasziou P, et al. Designing studies to ensure that estimates of test accuracy are transferable. BMJ 2002; 324: 669-71.

11. Battaglia M, Bucher H, Egger M, et al. (writing committee). The Bayes Library of Diagnostic Studies and Reviews. 2nd Edition Basel: Division of Clinical Epidemiology and Biostatistics, Insitute of Social and Preventive Medicine, University of Berne and Basel Insitute for Clinical Epidemiology, University of Basel, Switzerland; 2002; pp1-60. (available at daniel.pewsner@bluewin.ch)

12. Reid, M.C. et al. Use of methodological standards in diagnostic test research: getting better but still not good. JAMA 1995; 274(8): 645-651.

13. Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. Lancet 2000; 356: 1844-7.

14. Bruns DE. Laboratory-related outcomes in healthcare. Clin Chem 2001; 47: 1547-52.

15. Oosterhuis WP, Niessen RWLM, Bossuyt PMM. The science of systematic reviewing studies of diagnostic tests. Clin Chem Lab Med 2000; 38: 577-88.

16. Horvath AR, Pewsner D, Egger M. Systematic reviews in laboratory medicine:

Potentials, principles and pitfalls. In: Evidence-based Laboratory Medicine: From Principles to Practice. Price CP, Christenson RH (eds) AACC Press, Washington. 2003; pp137-158.

17. Song F, Eastwood A, Gilbody S, et al. Publication and related biases. Health Technol Assess 2000; 4: 1-105.

18. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Smith GD, Altman DG, eds. Systematic Reviews in Health Care. Meta-analysis in context. 2nd ed. London: BMJ Books, 2001; pp248-82.

19. Irwig L, Macaskill P, Glasziou P, et al. Meta-analytic methods for diagnostic tests accuracy. J Clin Epidemiol 1995; 48: 119-30.

20. Bruns DE, Huth EJ, Magid E, Young DS. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. Clin Chem 2000; 46: 893-895 (full text available free of charge, together with ensuing e-responses, on the web on: http://www.clinchem.org)

21. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Clin. Chem. 2003a; 49(1): 1-6.

22. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and Elaboration. Clin. Chem. 2003b; 49(1): 7-18.

23. Cochrane Methods Group on Screening and Diagnostic Tests. 1996.
(http://www.hiru.mcmaster.ca/COCHRANE/DEFAULT.htm; accessed August 2003)

24. Deeks J, Gatsonis C, Clarke M, Neilson J. Cochrane systematic reviews of diagnostic test accuracy. The Cochrane Collaboration Methods Groups Newsletter 2003; 7: 8-9.

25. Price CP. Evidence-based Laboratory Medicine: Supporting Decision-Making. Clin. Chem. 2000; 46 (8): 1041-1050.

26. Westgard JO. Why not evidence-based method specifications? 2002.
(http://www.westgard.com/essay44.htm; accessed August 2003)

27. Guyatt GH, Oxman AD, Ali M, et al. Laboratory diagnosis of iron deficiency anaemia: an overview. J. Gen. Intern. Med. 1992; 7: 145-53.

28. Sackett DL, Straus S. On some clinically useful measures of the accuracy of diagnostic tests. Evidence-Based Medicine 1998; 3: 68-70.

29. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgements: practicing physicians' use of quantitative measures of test accuracy. Am J Med 1998; 104: 374-380.

30. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. BMJ 2002; 324: 824-826

31. Oosterhuis WP, Bruns DE, Watine J, Sandberg S, Horvath AR. Recommendations for evidence-based guidelines in laboratory medicine. (submitted for publication)

32. Anonymous. Getting Evidence into Practice. Effective Health Care Bulletin, 1999; 5(1): 1-16.

33. Kay JD. Communicating with clinicians Ann.Clin. Biochem. 2001; 38(2): 103-110.