Advances in diagnostics for microbial agents: can clinical validation keep pace with the technical promises?

Giuseppe Giocoli(a), Cornelis J. Biesheuvel(b), Heather F. Gidding(c) and David Andresen(d)
(a) Associazione Microbiologi Clinici Italiani, Milan, Italy
(b) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
(c) The National Centre in HIV Epidemiology and Clinical Research, The University of New South Wales, Sydney, Australia
(d) Department of Microbiology and Infectious Diseases, and Centre for Kidney Research The Children’s Hospital at Westmead, Westmead, Australia

Summary. New technologies are revolutionizing diagnostic microbiology, but implementation of methodological advances in test evaluation has been extraordinarily slow. Published reports frequently fail to clearly distinguish between studies of analytical accuracy and those of clinical diagnostic performance. We argue that the evaluation of sensitivity and specificity, while necessary and occasionally problematic, is often inadequate to define the appropriate role of a new diagnostic test. To determine whether a test adds additional (incremental) value to the diagnostic information already available to the clinician, evaluation studies with a multivariable approach may also be needed. The impact of a test on patient outcome is best measured by a randomized controlled trial, but this may be unnecessary in certain well-defined situations. To improve the quality of published test evaluations close collaboration between clinicians, clinical microbiologists, and epidemiologists, as well as insistence by journal editors on the use of established reporting standards are essential.

Key words: laboratory techniques and procedures, diagnostic tests, sensitivity and specificity, microbiology, translational medicine.

Riassunto (Progressi in diagnostica microbiologica: riusciremo ad adeguare la validazione clinica dei test alle promesse della tecnologia?). Nuove tecnologie stanno cambiando il volto della diagnostica microbiologica, ma la metodologia degli studi di valutazione dei test non si è adeguata con tempestività ai progressi statistico-epidemiologici. Ne consegue, ad esempio, che nelle riviste di microbiologia clinica è spesso difficile distinguere gli studi di accuratezza analitica da quelli di accuratezza diagnostica. Riteniamo che sensibilità e specificità, misure indispensabili ma talvolta infide, non siano spesso sufficienti per definire il ruolo clinico di un test; per individuarne il valore aggiunto all’informazione diagnostica disponibile al clinico possono allora essere utili studi basati sull’analisi multivariata; o studi randomizzati, quando conviene apprezzare le ripercussioni del test sull’esito della malattia, anche se ciò non è sempre indispensabile. Per migliorare la qualità degli studi di valutazione dei test crediamo essenziali una stretta collaborazione tra clinici, microbiologi ed epidemiologi, ma anche l’esplicita richiesta da parte delle riviste di microbiologia clinica di un corretto e completo reporting degli studi eseguiti

Parole chiave: tecniche e procedure di laboratorio, test diagnostici, sensibilità e specificità, microbiologia, medicina translazionale.

INTRODUCTION

Modern technologies such as nucleic acid amplification (NAA) techniques are revolutionizing microbiology, and scientific journals devoted to clinical microbiology and infectious diseases frequently report evaluations of the performance of these new, as well as existing, diagnostic tests. However, we believe that the implementation of methodological advances in diagnostic test accuracy studies in clinical microbiology has been slower than in other diagnostic fields. A staged model has been proposed for the evaluation of medical tests [1, 2]. First, studies of technical/analytical accuracy (the test’s ability to accurately measure the target of interest) need to be conducted. Then, once analytical accuracy is established, studies of clinical accuracy (the ability to discriminate between patients who have the target associated disease and those who have not) are carried out. In microbiology journals these two stages are often difficult to distinguish at least in part because of poor reporting. The use of the terms “sensitivity” and “spe-
specificity” when reporting both technical/analytical and clinical evaluations is a major source of confusion in this regard. We would suggest that these terms be reserved for “epidemiological” performance characteristics as assessed by clinical studies. “Lower limit of detection” is a more appropriate term for technical/analytical “sensitivity” (the smallest quantity of an organism or analyte that may be detected in a specimen), and “non-cross-reactivity” is a better term for technical/analytical “specificity” (the lack of ability of an assay to detect other, usually non-pathogenic, organisms which may also be present in a sample).

The standards for reporting of diagnostic accuracy (STARD) framework provides authors and reviewers with a checklist of 25 recommended items for reporting test accuracy studies and encourages the use of a flow diagram to represent the design of the study and the flow of the patients through the study [3, 4]. It was first published in 2003 and subsequently endorsed by the International Committee of Medical Journal Editors in February 2006.

Despite their endorsement, these guidelines have rarely been implemented by clinical microbiology journals. The Journal of Clinical Microbiology (JCM) is maybe the most consulted microbiology journal to have formally adopted the STARD framework, and while it does refer to it in its instructions for authors, these instructions communicate no expectations regarding its use [5]. We could identify only three references to STARD in papers or letters published in the JCM since 2003, compared with 39 such references in Clinical chemistry and 42 in Radiology.

As a consequence of inadequately reported or methodologically suboptimal diagnostic test evaluations, clinicians and laboratorians alike are hamstringed in their attempts to identify appropriate diagnostic evaluations on which to base evidence-based guidelines for use in the microbiology laboratory [6]. The introduction of diagnostic techniques requires robust and valid diagnostic evaluation studies. We aim to provide some guidance on how and when to conduct such studies.

ROLE OF A TEST

New diagnostic tests may be used in one or more of three different ways [7] they may:
- replace an existing test;
- be used for triage (an initial test to decide who warrants further investigation); or
- be used as an add-on test subsequent to the performance of existing tests.

The role of the test in the clinical diagnostic pathway determines which characteristics the new test needs to have [8]. For example, when the test is intended to be used as an add-on performed after a highly sensitive screening (triage) test, specificity is paramount in order to decrease the number of patients with false positive test results.

EVALUATION OF CLINICAL DIAGNOSTIC TEST ACCURACY: UNIVARIABLE AND MULTIVARIABLE APPROACHES

Once analytical accuracy has been established by studies which show that the new test can identify the target of interest, formal assessment of clinical diagnostic accuracy is the next step in the evaluation process [9]. In a clinical diagnostic accuracy study, typically with a cross-sectional design, the test under evaluation (index test) is applied to a consecutive series of subjects suspected of having the disease of interest. Clinical test accuracy is assessed by comparing the results of the index test with the results of a reference standard (see next section) in the same subjects [10]. In the most common forms of diagnostic test evaluation, performance parameters such as sensitivity, specificity, likelihood ratios, and predictive values are estimated. A receiver operator characteristic (ROC) curve is also frequently derived, which represents the trade-off between sensitivity and specificity achieved by manipulation of the threshold for test positivity [3]. This process can be regarded as a univariable approach since it aims to quantify the performance characteristics of a single test without considering the diagnostic information already available from the patient’s history, physical examination, and possibly other more simple laboratory test results.

To determine whether the test adds additional or incremental value to the diagnostic information already available to the clinician, evaluation studies with a multivariable approach in a clinical setting may be needed [11]. Multivariable analyses account for the mutual dependencies between different sources and types of diagnostic information and thus indicate to what degree a diagnostic test independently contributes to the estimation of disease probability [11]. A new test might conceivably be highly sensitive and specific, but if the diagnosis could already be reliably predicted on clinical grounds then the incremental diagnostic value of the test would still be low. Clearly these more demanding multivariable studies will usually only be performed where a test appears promising in the initial univariable evaluation.

In clinical microbiology, a multivariable approach is important when evaluating tests for micro-organisms which are not exclusively pathogenic determinants of an illness. For example, to determine the presence of peptic ulcers in dyspeptic patients in primary care, Weijnen et al. [12] used multivariable logistic regression analyses to evaluate whether a non-invasive Helicobacter pylori test provides additional diagnostic information to history taking. Multivariable analyses are often used by microbiologists to identify independent predictors of an infection, but Weijnen’s is the only published study we are aware of that reports the incremental diagnostic value of a microbiological test.

The incremental value of rapid tests for life-threatening infections (eg. sepsis, meningitis) can and
should also be evaluated by the same multivariable approach. For example, Swiss researchers assessed the additional diagnostic value of serum procalcitonin measurement, over and above conventional markers of sepsis in intensive care unit (ICU) patients newly admitted with features of a systemic inflammatory response syndrome [13].

On the other hand, detection of an unequivocally pathogenic microbiological agent may be of such extraordinary diagnostic value in itself that a simple univariable comparison of the new test with traditional testing methods may be sufficient. This would be true for the detection of Neisseria gonorrhoeae in a genital swab, Shigella species in a watery stool sample and Bacillus anthracis in a cutaneous lesion.

SOLUTIONS TO OVERCOME THE ISSUE OF AN IMPERFECT REFERENCE STANDARD

Since clearly a gold (error-free) standard does not exist for many conditions, the terminology has shifted to the more neutral term reference standard, which is the best available method to determine the presence or absence of the disease of interest. Several solutions to the problems posed by an imperfect reference standard have been proposed. These include construction of a multifactorial reference standard using predefined rules, use of a consensus panel, or statistical modelling methods [14]. In situations where an acceptable reference standard does not exist, a direct assessment of the ability of the index test to predict future events of interest is a more radical “validation” approach, distinct from the classical diagnostic accuracy paradigm [14].

In the microbiology setting, a statistical modelling technique known as “latent class analysis” has been successfully applied to the evaluation of a NAA test for Chlamydia trachomatis where a single reference standard was inappropriate [15]. Another example is where a combined reference standard including culture and either clinical diagnosis or lung-bronchial biopsy has been used to study the diagnostic accuracy of PCR in smear-negative pulmonary tuberculosis [16]. In the absence of a reliable reference standard for latent tuberculosis, a “validation” model may come from observational studies evaluating whether interferon-gamma assay results are predictive of the development of active tuberculosis [17].

IMPACT ON PATIENT OUTCOME

One should only perform a test if it helps to establish a correct diagnosis and therefore benefits the patient’s well being. This may be the case if a positive test result allows the institution of effective treatment, or if a negative result avoids the need for potentially harmful, onerous or costly therapy. The impact of a new test on patient outcome is usually best determined by a randomised controlled trial (RCT) [10]. However under certain conditions, such laborious and time-consuming studies are not needed. For instance, if a diagnostic accuracy study has proven the test’s ability to detect a particular disease and there is a known effective treatment for that disease, it is reasonable to assume that the diagnostic test will improve patient outcomes [18]. This would be the case for most tests for aetiological agents of common acute infections.

Lord et al. [19] expand on these conditions, and show that a clinical accuracy study is sufficient if a new diagnostic test is safer or more specific than, and as sensitive as, the existing test which it is intended to replace. However if the new test is more sensitive, it may detect extra, and potentially less severe or less infectious cases. Results from studies of therapeutic or infection control interventions used for patients detected by the old test may not apply to these extra cases. For instance, in situations where the epidemiological or infection control value of early diagnosis has still to be clarified, such as testing for MRSA (meticillin-resistant Staphylococcus aureus) carriage [20], evaluation of the effects of a novel diagnostic test on patient outcome as well as on organism transmission would be required.

By contrast, if it can reasonably be assumed on biological grounds that the new test detects the same spectrum of disease, or if the potential harm from a missed diagnosis is great, then use of the new test without formal evidence of improved patient outcome may be justifiable. For instance, when a new, more sensitive, test is used to identify when to implement an infection control intervention, such as isolation procedures or contact prophylaxis against a pathogen that may cause a severe epidemic (H5N1 influenza virus, SARS-associated corona virus, Ebola virus), accurate pathogen detection is likely to be of such substantial public health benefit that its effect on individual patient outcomes would not need to be formally evaluated.

PROOF OF CAUSALITY

An issue of particular importance to clinical microbiology is proof of causality. The clinical relevance of many micro-organisms detected by the new, often highly sensitive detection tools such as NAA in biological samples is often unclear. We believe laboratories should not routinely offer diagnostic tests for micro-organisms before their pathogenic role has been confirmed by appropriate studies which use methods derived from Koch’s postulates or go beyond them [21]. For example, the association between Chlamydia pneumoniae and multiple sclerosis is unconfirmed [22] and therefore detection of this micro-organism has no clinical relevance for patients with multiple sclerosis.

THE NEED FOR CHANGE

In this era of evidence-based medicine and outcome-orientated research, we are puzzled by the paucity of methodologically appropriate and advanced
evaluations of microbiological tests on which to base clinical practice guidelines and diagnostic test implementation decisions.

We can understand why seminal papers on the methodology of diagnostic test evaluation, even where infection scenarios are described, are missing from the microbiological literature. It is however regrettable that, with very few exceptions, microbiology journals are barren when it comes to providing guidance about appropriate design, analysis, and reporting methods for diagnostic test evaluations. Paradoxically, high quality evaluations of tests for agents or biomarkers of infection may be more likely to be located in a clinical chemistry rather than a microbiology journal. As the methodology for meta-analyses of diagnostic test performance is developed [23] and such analyses are performed, reviewers will be reliant on adequate reported primary studies to produce valid and meaningful secondary analyses.

We believe that the attitudes and expectations of journal editors must be partly responsible for the current situation. Is it possible that microbiology journal editors feel that requiring a STARD checklist and patient flow chart from the authors is a pretension rather than the solution to improving the quality of diagnostic studies? It is arguable that a clear statement in each journal’s “instructions for authors” communicating an expectation that STARD guidelines be followed would go a long way towards improving the conduct and reporting of future diagnostic test evaluations.

The mission of clinical microbiologists, as for other laboratory medicine specialists, is ultimately to improve patient care by optimising laboratory testing in order to provide additional guidance to treating clinicians [24]. We have tried to indicate some of the barriers which need to be removed in order to achieve such ambitious goals, particularly the long-established separation between the epidemiology of diagnostic test evaluation and its application to microbiological diagnostics. We believe that the remedies must include close collaboration between clinicians, clinical microbiologists, and epidemiologists when planning, conducting and reporting diagnostic test evaluations, and insistence by the editors of relevant journals that submitted evaluations adhere to established reporting standards.

Acknowledgements

We thank Les Irwig of the Screening and Test Evaluation Programme, School of Public Health, Australia for his comments on drafts of this paper. At the time of writing this review, C. Biesheuvel was employed by the Screening and Test Evaluation Programme, School of Public Health, Australia and the Children’s Hospital at Westmead, Australia. The National Centre in HIV Epidemiology and Clinical Research is funded by the Australian Government Department of Health and Aging and is affiliated with the Faculty of Medicine, The University of New South Wales.

Conflict of interest: All authors declare that they do not have an association that might pose a conflict of interest. The work of C. Biesheuvel was partly funded under a program grant (number 402764) from the National Health and Medical Research Council of Australia. The funding source had no involvement.

Received on 27 October 2008.
Accepted on 24 March 2009.

References


