Improving accuracy of diagnostic studies in a world with limited resources: a road ahead

Giuseppe Lippi[1], Mario Plebani[2]

[1]Section of Clinical Biochemistry, University of Verona, Verona, Italy; [2]Department of Laboratory Medicine, University Hospital of Padova, Padova, Italy
Correspondence to: Prof. Giuseppe Lippi. Section of Clinical Biochemistry, University Hospital of Verona, P.le LA Scuro 10, 37134 Verona, Italy. Email: giuseppe.lippi@univr.it; ulippi@tin.it.

Diagnostic testing

The worldwide macroeconomics is only now recovering from an unprecedented economic crisis, which has profoundly changed the organization of several national healthcare systems around the globe (1). In most industrialized and emerging countries, health care expenditure represents an increasing part of the gross domestic product (GDP), exhibiting a growth rate that has largely exceeded other sectors of economy, and ultimately constitutes one of the largest components (up to 10–12%) of national budgets (2,3). Due to remarkable technological advances in diagnostics and targeted therapeutics, most healthcare systems are now squeezed between costs that continue to increase, especially for hospital care and medications, and funding from national governments that progressively decreases. The issue of how to redesign the entire system around a more efficient provision of high quality care is hence crucial for both private and public payers.

There is now firm evidence that laboratory diagnostics, also known as *in vitro* diagnostic (IVD) testing, generates a kaleidoscope of clinically useful information for the screening, diagnosis, prognostication and therapeutic monitoring of most (if not all) human disorders (4). Despite being unquestionable that the huge amount of valuable data that can be generated by this branch of science and medicine is changing the way we diagnose and treat our patients, and that the relative cost of diagnostic testing only modestly impacts on the overall health care expenditure (typically between 1.5% and 2.0%) (5), laboratory managers and scientists are subjected to a huge pressure to cut down costs for reagents and personnel, and streamline their laboratories around a paradigm of enhanced efficiency, which often does not go hand in hand with improved clinical efficacy (6). At variance with other branches of medicine, diagnostic testing is a continuously evolving scenario, characterized by undefined landmarks and boundaries. The traditional model of clinical laboratory has dramatically evolved in the past few years towards a new environment increasingly pervaded by high-throughput genomic, transcriptomic, and proteomic technologies. Cancer diagnostics is a paradigmatic example on how recent technological advancements are impacting clinical medicine. With cancer increasingly being recognized as a highly heterogeneous and virtually individual disease, the shift towards personalized medicine will progressively need that the tumor of every patient be treated uniquely, according to peculiar genetic traits (that is, pharmacogenomics) (7-10). In breast cancer, for example, the HER2/neu has become the target of the monoclonal antibody trastuzumab, from which women with HER2/neu overexpressing malignancies may obtain considerable benefits in both metastatic and adjuvant settings (11). Paradoxically, genetic studies have also revealed the existence of molecular hallmarks and biological pathways that are shared between cancers that are apparently unrelated, so that treatments which are efficiently used for one type of cancer may be translated to other malignancies. However, the intriguing and valuable perspectives offered by IVD testing do require that a rigorous approach is followed to generate reliable scientific evidence from basic and validation studies.

Page 2 of 4

Lippi and Plebani. Improving accuracy of diagnostic studies

The Standards for Reporting of Diagnostic Accuracy Studies (STARD)

Inaccurate reporting is currently regarded as a leading source of avoidable waste in biomedical studies. As a rule of thumb, necessary data is seldom unavailable, thus making critical appraise and replication studies virtually impossible. The STARD guidelines have hence been developed in 2003, with the aim of improving accuracy of diagnostic studies to be published as scientific articles, conference abstracts, or be included in trial registries (12). The original STARD statement consisted of 25 specific items organized within a prototypical flow diagram, aimed to provide comprehensive information about ideal number of subjects enrolled, methods of patient recruitment, order of test, reference and benchmark techniques. Some limitations of the STARD statements have been highlighted, particularly in the field of non-invasive liver fibrosis biomarkers, and a proposal for an extension of those statements has been reported (13). Nearly a decade after the former STARD statement, the consortium has released a revised and updated version of their guidelines (14). The new STARD 2015 statement follows the same organization as the previous, but the list has now been expanded to 30 items, grouped in homogenous classes. More specifically, the STARD 2015 replaces the original version and provides a clear and useful guidance for planning reports of diagnostic accuracy studies, from the title to the discussion of the report. Special focus is placed on the section "methods", entailing characteristics of study design, participants (e.g., eligibility criteria), test methods (e.g., adequate description of index and reference methods, selection of diagnostic thresholds) and data analysis (e.g., the approach used for assessing diagnostic accuracy). Notably, the STARD 2015 also provides clear definitions of key terminology (e.g., definitions of medical and index tests, target condition, clinical reference standard, sensitivity, specificity, intended use and role of the test), along with a tentative diagram for reporting participants flow throughout the study. Among the new items available in STARD 2015, two deserve special attention. Specifically, the full study protocol should be made available to the readers, in order to allow possible reproduction and critical appraisal. Particular emphasis is also placed on funding declaration, so that potential conflicts of interest are fully disclosed to editors and readers.

What emerges clearly from the publication of the STARD 2015 revised statement is that a more rational and standardized strategy would be possible in the very next future for reporting data of diagnostic accuracy studies, provided that this approach is endorsed by national and international scientific societies, and widely accepted by journals publishing articles in this field. A more rigorous approach would also enable to increase comparability among different investigations (and thus reducing the heterogeneity of meta-analytic reviews based on pooled data), generate a more credible and solid scientific evidence for introducing (or maintaining) diagnostic tests within healthcare pathways, support the process for validating innovative biomarkers or technology and their translation into routine practice, thus providing the best possible care to the patients, and ultimately allowing a more efficient use of healthcare resources. Indeed, the STARD 2015 statement is also intended to complement the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2), a widely used tool for systematic reviews of diagnostic accuracy studies (15). Nevertheless, although the revised STARD 2015 guidelines should hence be considered an unquestionable advancement, some questions remain.

How accuracy of diagnostic testing can be defined?

Diagnostic accuracy (or efficiency) is conventionally defined as the ability of a test to identify or exclude a given disease, whereas diagnostic efficacy (or effectiveness) defines whether the same test generates a significant change in managed care and ultimately produces an improvement of clinical outcomes (16). Pragmatically, the diagnostic efficiency only expresses the accuracy to correctly identify a given condition, with the awareness that excellent diagnostic performance does not necessarily translate into improved outcomes (e.g., an highly accurate but delayed diagnosis of metastatic cancer has a little impact on mortality rates). Conversely, clinical efficacy implies that a specific healthcare intervention actually helps achieving a primary or secondary endpoint in the clinical setting (17).

Is the STARD universally applicable?

Generalization is often a risky inclination, especially in science and medicine. Some rational doubts emerges as to whether the STARD criteria can be indifferently applied to assess the diagnostic accuracy of both innovative and consolidated biomarkers, or else used for evaluating either prototype methods or commercial assays. This is not an

Table 1 Statistical criteria for assessing diagnostic accuracy

General
Value distribution (normal, non-normal, skewed)
Value of statistical significance (e.g., P<0.05; P<0.01, etc.)
Comparison against the reference method
Correlation, linear regression analysis, Deming fit
Bias estimation (e.g., by means of Bland and Altman plots)
Diagnostic agreement (e.g., kappa statistics)
Evaluation of diagnostic performance against clinical outcomes
Sensitivity and specificity
Negative predictive value (NPV) and positive predictive value (PPV)
Negative likelihood ratio (LH$^-$) and positive likelihood ratio (LH$^+$)
Diagnostic odd ratio (DOR)
Graphical representation of receiver operating characteristics (ROC) curve
Number need to test (NNT)

ancillary issue, wherein basic or applied research often recognize different bases and frequently pursue different targets. Putative biomarkers are occasionally developed for unraveling complex biological pathways and increasing our understanding of health and disease. In such circumstance, the application of stringent criteria may be unnecessary. Conversely, the validation of diagnostic (commercial) kits has a rather different scope, that is to establish whether or not they are aligned with the reference method and, preferably, if the results of testing do generate a significant change in the clinical decision making.

Can we add more to the STARD criteria?

The 30-item list contained in the 2015 STARD statement is indeed thorough and accurate, since it targets the most critical steps throughout data reporting of diagnostic accuracy studies. Nevertheless, no precise definition of the most suitable approach that should be used for evaluating diagnostic performance has been put forward, thus allowing a relative degree of freedom and arbitrary methodology for analyzing data. A tentative guidance, such as that described in *Table 1*, might be an advisable appendix to the 2015 STARD guidelines.

The history of diagnostic testing has been categorically built around the rather narrow concept of diagnostic

accuracy (or efficiency). To put it simple, we have been answering for decades to highly focused questions such as "does this test help me reach a final diagnosis?" or "does this test help me rule out a given disease?", or even "does this new test adequately compare with the reference method?". Under a genuine clinical perspective, the landscape and the characterizing paradigms of diagnostic testing are different, however, and we have often (and guiltily) ignored the existence of the moon beyond the finger. The add value of a clinical information, either gathered from physical examination, diagnostic imaging or laboratory testing, lies in the opportunity to generate a favorable impact on the clinical course of disease, which means improving quantity or quality of life. Therefore, the new paradigm for evaluating diagnostic test implies moving from the narrow concept of diagnostic accuracy to the broader, more appropriate and useful notion of clinical efficacy. Some tools have been recently developed to assist this paradigm shift. The grading of recommendations assessment, development and evaluation (GRADE) evidence to decision (EtD) frameworks is aimed to supply a systematic and transparent approach for translation of clinical evidence into healthcare behaviors and recommendations. The application of GRADE to diagnostic testing should hence be regarded as a precious resource for assessing accuracy of diagnostic testing, but also for evaluating the clinical impact on key endpoints such as mortality, morbidity and quality of life (18). Cost-benefit analyses, using for example the well-established health technology assessment (HTA) approach, are also essential to substantiate that a high diagnostic accuracy of the index test does really translate into clinical and economical benefits.

Conclusions

Squeezed between reduced funding and increasing volumes and complexity (19,20), diagnostic resources should be increasingly offered according to the principles of evidence-based (laboratory) medicine. The future of the field is still unwritten, and many potential scenarios are emerging, including multi-analyte testing at low prices in commercial outlets (21) or rapid diagnostics of infectious diseases using smartphones (22). Indeed, we all agree that a widespread adoption of STARD 2015 criteria may be seen as a road ahead for improving accuracy of diagnostic studies in a world with limited resources. Yet, STARD 2015 is not the panacea and many other problems remain almost unanswered. Bridging the gap between the bench

Page 4 of 4

Lippi and Plebani. Improving accuracy of diagnostic studies

and the bedside, promoting a cultural sea change from a consolidated concept of diagnostic accuracy to that more pervasive of clinical efficacy, and placing major focus on personalized care are promising opportunities for improving the value of diagnostic testing.

References

1. Appleby J, Helderman JK, Gregory S. The global financial crisis, health and health care. Health Econ Policy Law 2015;10:1-6.
2. Budhdeo S, Watkins J, Atun R, et al. Changes in government spending on healthcare and population mortality in the European union, 1995-2010: a cross-sectional ecological study. J R Soc Med 2015;108:490-8.
3. Zhang W. The other side of the Chinese economic miracle. Int J Health Serv 2012;42:9-27.
4. Lippi G, Plebani M. Laboratory medicine does matter in science (and medicine)… yet many seem to ignore it. Clin Chem Lab Med 2015;53:1655-6.
5. Lippi G, Mattiuzzi C. Testing volume is not synonymous of cost, value and efficacy in laboratory diagnostics. Clin Chem Lab Med 2013;51:243-5.
6. Plebani M. Clinical laboratories: production industry or medical services? Clin Chem Lab Med 2015;53:995-1004.
7. Alix-Panabières C, Pantel K. Real-time liquid biopsy: circulating tumor cells versus circulating tumor DNA. Ann Transl Med 2013;1:18.
8. Ryu JS, Memon A, Lee SK. ERCC1 and personalized medicine in lung cancer. Ann Transl Med 2014;2:32.
9. Zhai H, Zhong W, Wu Y. Research, evidence, and ethics: new technology or grey medicine. Ann Transl Med 2015;3:15.
10. Lippi G, Plebani M. Personalized medicine: moving from simple theory to daily practice. Clin Chem Lab Med 2015;53:959-60.
11. Cho SH, Jeon J, Kim SI. Personalized medicine in breast cancer: a systematic review. J Breast Cancer 2012;15:265-72.
12. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Clin Chem Lab Med 2003;41:68-73.
13. Boursier J, de Ledinghen V, Poynard T, et al. An extension of STARD statements for reporting diagnostic accuracy studies on liver fibrosis tests: the Liver-FibroSTARD standards. J Hepatol 2015;62:807-15.
14. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. Clin Chem 2015;61:1446-52.
15. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36.
16. Lippi G, Mattiuzzi C. The biomarker paradigm: between diagnostic efficiency and clinical efficacy. Pol Arch Med Wewn 2015;125:282-8.
17. Lippi G, Mattiuzzi C, Cervellin G. Biomarker validation in the emergency department. General criteria and clinical implications. Emerg Care J 2014;10;1860.
18. Trenti T, Schünemann HJ, Plebani M. Developing GRADE outcome-based recommendations about diagnostic tests: a key role in laboratory medicine policies. Clin Chem Lab Med 2015. [Epub ahead of print].
19. Lippi G, Plebani M. Biomarker research and leading causes of death worldwide: a rather feeble relationship. Clin Chem Lab Med 2013;51:1691-3.
20. Lippi G, Di Somma S, Plebani M. Biomarkers in the emergency department. Handle with care. Clin Chem Lab Med 2014;52:1387-9.
21. Diamandis EP. Theranos phenomenon: promises and fallacies. Clin Chem Lab Med 2015;53:989-93.
22. Bates M, Zumla A. Rapid infectious diseases diagnostics using Smartphones. Ann Transl Med 2015;3:215.