

Editorial

Tommaso Trenti and Mario Plebani

Diagnostic test accuracy: a valuable tool for promoting quality and patient safety

<https://doi.org/10.1515/dx-2018-0091>

The evidence-based recommendations for the appraisal of new tests to be introduced in clinical practice or at present used test are a key issue to developing diagnostic clinical pathways inducing the most effective care. As a general point of view a sound assessment in healthcare should be the intervention effect on patient health and translated in Laboratory Medicine how the evaluated diagnostic test improves the patient's health, however, this is very difficult to appraise because confounding intermediate networking issues at stake.

Decision models are proposed to structure the interpretation of the analyzed findings and such models or frameworks would include essential elements as the disease prevalence, probable outcomes, and the existing diagnostic and therapeutic interventions that may follow the laboratory assay [1–4]. Developing precision or personalized medicine demands novel and pioneering biomarkers for molecularly targeted therapies, possibly tailored for the individual patient's condition. Behind the conventional analytical requirements that should guarantee the appropriate clinical diagnostic performances in response to a specific clinical question, the outcomes of a new test should be clearly outlined and evaluated. Systematic reviews and meta-analysis of diagnostic test accuracy studies can be employed to achieve more precise estimates when studies focusing on the same test and patients in the same setting are available, however, systematic reviews of diagnostic test are still methodologically challenging nevertheless increasingly popular and published. Recently published papers emphasize the challenges inherent in linking laboratory testing to clinical decision making and downstream effects on diagnosis, treatment, patient outcomes and costs [5, 6]. Most research papers in this field deal with an evaluation of analytical performances alone, while other papers report on measures such as diagnostic sensitivity and specificity that are only surrogate end points. In fact, the ability of a diagnostic (particularly laboratory) test to identify a patient with a truly positive or truly negative result is defined by a test's sensitivity and specificity. However,

these terms are difficult to be applied in clinical practice. First, it is difficult for a practicing physician to remember what these terms represent and even more difficult to use them in the diagnostic process for individual patients. Second, the prevalence of disease in the population strongly affects the information requested by a physician regarding test accuracy. If the prevalence of disease in the population is very low, there is a large risk that the positive test is a false positive even if the test is very specific. The predictive positive value, in fact, is strongly related to pre-test probability and disease prevalence. It is analogous for a highly sensitive test and the negative predictive value (NPV). Third, the main question in diagnostic accuracy research is not what the diagnostic accuracy of a particular test is, but rather “whether it improves the diagnostic accuracy of the existing workup” [7] or better and even better “if it provides benefit to patient health” [8]. The weak reporting of primary diagnostic test accuracy research, the complexities with the interpretation of the results of diagnostic test accuracy research and their clear and useful report are important challenges.

Harmonization and risk management policies are key issues in laboratory medicine as they direct on a patient-centered release of laboratory information based on the perception of the value of the total testing process for ensuring healthcare quality and patient safety [9, 10]. The new lines to quality and patient safety in the healthcare system highlight that diagnostic improvements are founded on the assurance of the requested outcomes rather than on the exclusive documentation of the errors. The outcome-based approach suggested by Epner et al. [11] on testing-related diagnostic errors demands for a more effective demand and interpretation of effective biomarkers in order to prevent poor effects or events, mistake to diagnose and to deliver the appropriate treatment. Patient safety is compromised by incorrectly ordered tests or by misunderstanding of the results. It should be essential to look at the test-treatment pathway as a whole, and not in isolation, to evaluate the clinical impact of the test on patients. Therefore, the predictive values, the sensitivity, specificity, likelihood ratio and receiver operating curve area for a single test are not constant but will vary

across disease severity, patient characteristics and other variables [12].

The clinical-laboratory interface is fundamental to assure an effective clinical decision making process and successful patient care based on outcome assessment if the scope of the so-called “harmonization policy” in Laboratory Medicine goes beyond only method and analytical results appraisal to encourage the value of all other features of laboratory testing, facing strategies for test demand and criteria for result interpretation [13]. Analytical and diagnostic performances such as sensitivity, specificity, imprecision, positive, and NPVs are traditionally recognized measures but the clinical impact and the healthcare outcomes, to which these accuracy measures are related, are complex to be measured and rated. The extent of the improvement of the patients’ health due to a diagnostic test remains a “holy grail” notwithstanding it should be the ultimate goal [14]. The test accuracy in diagnostic test evaluation explains how a test recognizes the diseased patients (sensitivity) against healthy patients (specificity) but these two factors are not able to outline fully the true clinical value of the test.

The improvement in test accuracy may not benefit the patients if it does not affect patient management and the resulting health outcomes, highlighting the value of the processes that can drive changes to patient health from testing downstream consequences [15]. If harmonization and risk management increasingly understands the primary necessity to take into account patient outcomes in the evaluation of tests and test strategies, the promotion and development of high quality recommendations may be the framework to disseminate this view in the evaluation appraisal of proposed innovative diagnostic tests to be implemented in clinical practice from research [16, 17].

Diagnostic test accuracy systematic reviews are suitable tools for developing recommendations on the use of a test where all published studies are synthesized to obtain pooled and consistent DTA measures. However, such results are not the only the arbiter of judgments about the role of a test to induce quality and safety and the published DTA SRs often do not support a complete guidance on the new tests for those making test implementation decisions [18, 19].

Rubinstein and colleagues [20] publish in this issue of the journal a new effective approach to systematic review on diagnostic test accuracy as a potentially more useful tool for diagnostic quality and safety. The assumption is that DTA SRs provide insight into a test’s – or test combination – ability to contribute to quality and safety within diagnostic pathways by estimating a test’s clinical validity. In particular, DTA SRs should play an

essential role in the pre-analytical phase for improving appropriateness in test request, in the intra-analytical phase for setting (and monitoring) reliable performance specifications, and in the post-analytic for improving result interpretation. Although reporting standards for DTA studies have been developed and included in the Standards for Reporting of Diagnostic Accuracy, new efforts should be made to allow laboratorians and clinicians to better appreciate the importance of reliable accuracy measures and their utilization in clinical practice.

This new methodological approach is based on the graphical location of a diagnostic accuracy study within a four quadrant likelihood ratio scatter matrix together a matrix quadrant demarcation derived from established likelihood ratio thresholds. The plotted position is suggesting the DTA SRs clinical validity to resolve limitations related to the uncommon use of an analytic framework which provides a complete scope and context for DTA measures.

The authors summarize the positive and negative likelihood ratios definitions and interpretations as a general depiction of DTA measures to rule-in or rule out as increasing capability to assess if a patient’s probability of disease changes with a positive or negative test result. This is described and plotted in a four-quadrant likelihood ratio scanner matrix with a suggestion about the effect size ranking in terms of substantial, moderate or minimal. The proposed approach is likely to categorize the overall index of sample size as correlates to the tradeoff between sensitivity and specificity to evaluate the related clinical value. In this way, as reported by the authors, the resulting table of findings may provide counts for the highest assessed pairings of quality-to-effect for each practice category. The strength of DTA SRs body of evidence ratings may be used to inform derived practice recommendations. Other methods for grading the strength of evidence as GRADE may benefit from this effect size rating approach as DTA SRs may be used as surrogate or intermediate patient outcomes and the extent that rates TP, FP, TN and FN can be linked to the patient management or patient health assay consequence. So the LR scatterplot matrix may be useful to move beyond summary measures and identify how a new diagnostic test reclassifies patients [21] by supporting assessments of the DTA SRs effect associated with clinical practice.

The proposed +LR/–LR scatterplot matrix pairing may inform the “diagnostic process” and may provide an additional means to express whether diagnostic tests have adequate analytical and clinical validity to preventing diagnostic errors in agreement with the quality and safety

measures described in the 2017 National Quality Forum report “Improving Diagnostic Quality and Safety”.

This decision model could be used to structure the interpretation of the DTA SRs findings and such a model would incorporate important factors such as the disease prevalence, likely outcomes, and the available diagnostic and therapeutic interventions however, as reported by the authors, it needs a further evaluation of the indirectness of evidence to patient outcomes, such as costs, resource utilization, equity, acceptability, feasibility, etc.

The interpretation of the results offered in the systematic review should help readers to recognize the inferences of the given DTA SRs data for clinical practice and specifically if the evidence resulting from the review suitably addressed the objectives of the review. This may require attention about whether the study sample was correctly representative, whether the included studies investigated the proposed current or future role of the test under evaluation, and whether the results are suspected to be biased. The GRADE methodology, increasingly popular in the area of diagnostic testing, is aimed at including the main health outcomes in the evaluation process in light of the resulting downstream clinical actions. As direct studies assessing the impact of diagnostic tests or strategies on patient important outcomes are rarely available, the GRADE process requires two main steps. The first is the assessments in judging the directness between test accuracy and the evaluated health outcomes and the second is the transparent criteria used to move from evidence to a recommendation [22].

The GRADE rating the certainty of the evidence about the effects of a new test and the subsequent management decisions on patient-important outcomes may be used to complete the decision process about the evaluation of a new proposed test based on the application of the DTA SRs evaluations by the GRADE evidence to decision (EtD) frameworks. The GRADE EtD frameworks involves five criteria for making judgments and assessment of the evidence a certainty: (1) test accuracy, (2) any critical or important direct benefits, adverse effects or burden of the test, (3) effects of natural history or the management that is guided by the test results, (4) the link between the test results and the management decisions and (5) the evidence about the effects of the test. Of importance to complete the DTA SRs data assessment is the evaluation of the resource use as judgments about the extent of costs, certainty of evidence of resource requests and the cost-effectiveness of interventions. This should include the evaluation of the resource use and cost impact both within the laboratory and the downstream consequence. The great challenge is to recognize the total health care

cost and not only the plan cost of the laboratory assay itself [23]. Assessments of equity, acceptability, and feasibility cover both the test and the following interventions. The use or misuse of tests for a given clinical presentation in distinct professional settings influences equity of access to clinical care.

This overall broad-spectrum approach is even more important in the current scenario which highlights the role of clinical laboratory stewardship in modifying and improving the process of ordering, performing, and reporting laboratory tests to improve patient care and safety.

As diagnostic error is a major health care concern and worthy of much more attention, the paper by Rubinstein and colleagues should be welcomed and recognized as an important source of information and knowledge to improve the quality of diagnostic testing processes as well as quality of care.

The proposed high quality DTA SRs data assessment may support evidence based recommendations to promote a methodologically proper clinical governance policy where the introduction of new laboratory test based on assessed outcome for patients is an added value to a healthcare system allowing the laboratory personnel to agree to the management and the allocation of resources in terms of health priorities [24].

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

References

1. Monaghan PJ, Lord SJ, St John A, Sandberg S, Cobbaerth CM, Lennartz L, et al. Test evaluation Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine biomarker development targeting unmet clinical needs. *Clin Chim Acta* 2016;420:211–9.
2. Horvath AR, Lord SJ, St John A, Sandberg S, Cobbaert CM, Lorenz S, et al. From biomarkers to medical tests: the changing landscape of test evaluation. *Clin Chim Acta* 2014;427:49–57.
3. Monaghan PJ, Robinson S, Rajdl D, Bossuyt PM, Sandberg S, St John A, et al. Practical guide for identifying unmet clinical needs for biomarkers. *eJIFCC* 2018;29:129–37. Accessed: 12 Jul 2018.
4. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, et al. What is “quality of evidence” and why is it important to clinicians? *Br Med J* 2008;336:995–8.
5. Plebani M. Quality and future of clinical laboratories: the Vico’s whole cyclical theory of the recurring cycles. *Clin Chem Lab Med* 2018;56:901–8.

6. Rubinstein M, Hirsch R, Bandyopadhyay K, Madison B, Taylor T, Ranne A, et al. Effectiveness of practices to support appropriate laboratory test utilization: a laboratory medicine best practices systematic review and meta-analysis. *Am J Clin Pathol* 2018;149:197–221.
7. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. *Clin Chem* 2012;58:1408–17.
8. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ* 2012;344:e686.
9. Plebani M, Panteghini M. Promoting clinical and laboratory interaction by harmonization. *Clin Chim Acta* 2014;432:15–21.
10. Tate JR, Johnson R, Barth J, Panteghini M. Harmonization of laboratory testing – current achievements and future strategies. *Clin Chim Acta* 2014;432:4–7.
11. Epner PL, Janet EG, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. *BMJ Qual Saf* 2013;0:1–5.
12. Linnet K, Bossuyt PM, Moons KG, Reitsma JB. Quantifying the accuracy of a diagnostic test or marker. *Clin Chem* 2012;58:1292–301.
13. Panteghini M. Implementation of standardization in clinical practice: not always an easy task. *Clin Chem Lab Med* 2012;50:1237–41.
14. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149:889–97.
15. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. HJ GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
16. McLawhon RW. Patient safety and clinical effectiveness as imperatives for achieving harmonization inside and outside the clinical laboratory. *Clin Chem* 2011;57:936–83.
17. 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* 2016;54:535–43.
18. Singh S, Chang SM, Matchar DB, Bass EB. Grading a body of evidence on diagnostic tests. In: Chang SM, Matchar DB, Smetana GW, Umscheid CA, editors. *Methods guide for medical test reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US), 2012.
19. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. *Syst Rev* 2013;2:82.
20. Rubinstein ML, Kraft CS, Parrott JS. Determining qualitative effect size ratings using a likelihood ratio scatter matrix in diagnostic test accuracy systematic reviews. *Diagnosis (Berl)* 2018;5:205–14.
21. Cornell J, Mulrow CD, Localio AR. Diagnostic test accuracy and clinical decision making. *Ann Intern Med* 2008;149:904–6.
22. Gopalakrishn G, Mustaf RA, Davenport C, Scholten RJ, Hyde C, Brozek J, et al. Applying grading of recommendations assessment, development and evaluation (GRADE) to diagnostic tests was challenging but doable. *J Clin Epidemiol* 2014;67:760–8.
23. Brunetti M, Pregno S, Schünemann H, Plebani M, Trenti T. Economic evidence in decision-making process in laboratory medicine. *Clin Chem Lab Med* 2011;49:617–21.
24. Trenti T, Canali C, Scognamiglio AM. Clinical Governance and evidence-based laboratory medicine. *Clin Chem Lab Med* 2006;44:724–32.

Corresponding author: Tommaso Trenti, Laboratory Medicine and Pathology Department, Ospedale Sant'Agostino Estense, via Giardini 1355, 41126 Modena, Italy, E-mail: t.trenti@ausl.mo.it
Mario Plebani: Department of Laboratory Medicine, University-Hospital, Padova, Italy. <http://orcid.org/0000-0002-0270-1711>