

Rational use of diagnostic tests for clinical decision making

 Anna Maria Buehler¹
 Bruna de Oliveira Ascef¹
 Haliton Alves de Oliveira Júnior¹
 Cleusa Pinheiro Ferr¹
 Jefferson Gomes Fernandes¹

¹. Health Technology Assessment Unit, Hospital Alemão Oswaldo Cruz – São Paulo – SP – Brasil

<http://dx.doi.org/10.1590/1806-9282.65.3.452>

SUMMARY

OBJECTIVE: To assist clinicians to make adequate interpretation of scientific evidence from studies that evaluate diagnostic tests in order to allow their rational use in clinical practice.

METHODS: This is a narrative review focused on the main concepts, study designs, the adequate interpretation of the diagnostic accuracy data, and making inferences about the impact of diagnostic testing in clinical practice.

RESULTS: Most of the literature that evaluates the performance of diagnostic tests uses cross-sectional design. Randomized clinical trials, in which diagnostic strategies are compared, are scarce. Cross-sectional studies measure diagnostic accuracy outcomes that are considered indirect and insufficient to define the real benefit for patients. Among the accuracy outcomes, the positive and negative likelihood ratios are the most useful for clinical management. Variations in the study's cross-sectional design, which may add bias to the results, as well as other domains that contribute to decreasing the reliability of the findings, are discussed, as well as how to extrapolate such accuracy findings on impact and consequences considered important for the patient. Aspects of costs, time to obtain results, patients' preferences and values should preferably be considered in decision making.

CONCLUSION: Knowing the methodology of diagnostic accuracy studies is fundamental, but not sufficient, for the rational use of diagnostic tests. There is a need to balance the desirable and undesirable consequences of tests results for the patients in order to favor a rational decision-making approach about which tests should be recommended in clinical practice.

KEYWORDS: Clinical Decision-Making. Diagnostic Tests, Routine. Evidence-Based Practice. Sensitivity and specificity. Predictive value of tests. Diagnostic equipment.

INTRODUCTION

Diagnostic tests are used in clinical practice to identify the presence, absence or characteristics of a condition of interest in a patient, aiming to develop a plan to the appropriate treatment¹. Additionally, they can be used to identify physiological changes, establish a prognosis, monitor the progress of a disease or response to treatment, and assist in clinical management.

A new test can play one of three fundamental

roles: 1) act as a screening test (to minimize the use of invasive or expensive tests, as, for example, the use of magnetic resonance imaging before a biopsy in patients with prostate cancer); 2) replace a current test (by presenting a higher precision, being less invasive, having a shorter execution time or lower cost - for example, patients with pneumonia associated with mechanical ventilation who received a quick test of antimicrobial susceptibility had their results, on av-

DATE OF SUBMISSION: 15-May-2018

DATE OF ACCEPTANCE: 26-May-2018

CORRESPONDING AUTHOR: Anna Buehler

Rua João Julião, 245, 1o andar – São Paulo – CEP 01323-903 – Brasil

E-mail: abuehler@haoc.com.br; annabuehler@yahoo.com.br

erage, 2.8 days before the standard tests for susceptibility and were able to receive early treatment, with better prognosis²; or 3) can be added to an existing test (improve the diagnostic accuracy of an existing test - for example, myocardial perfusion study and electrocardiogram in coronary artery disease)³.

To recommend a diagnostic test in clinical practice, you must consider its impact on the improvement of outcomes that are important to the patient, such as those related to the disease (for example, reduction of mortality), reduction of the psychological consequences of tests to patients, reduction of risks associated to the test and also evaluate the outcomes related to the use of resources⁴. However, in clinical practice, decision making is often based only on parameters of test accuracy (sensitivity and specificity, among others), which is not the most appropriate approach, considering that the parameters of accuracy are considered surrogate endpoints capable of providing indirect evidence about the effectiveness of the diagnostic strategy adopted. It is necessary to bear in mind that merely establishing a diagnosis does not provide information about whether a patient or group of patients will benefit from the diagnosis⁵.

New diagnostic tests are constantly developed, driven by demands for improvements in diagnostic speed, cost, ease of application, patient safety, and accuracy⁶. However, the efforts to detect diseases early can always be accompanied by unintentional damage. These include false-positive results and indeterminate discoveries that may worry patients, lead to more tests, increase the load of clinical work and distract doctors from the most important work. Also, excessive diagnoses may lead to unnecessary treatments⁶. Therefore, in clinical practice, it is common to have more than one option of test available for diagnosing a certain condition.

Thus, this review aims to discuss topics that help the decision/recommendation of a diagnostic test in clinical practice, reviewing the key concepts, study design, the proper interpretation of data accuracy, and inferring the clinical impact of the test in the diagnosis and the implications of choosing it.

IDENTIFYING AND ASSESSING THE RISK OF BIAS IN STUDIES OF DIAGNOSTIC ACCURACY

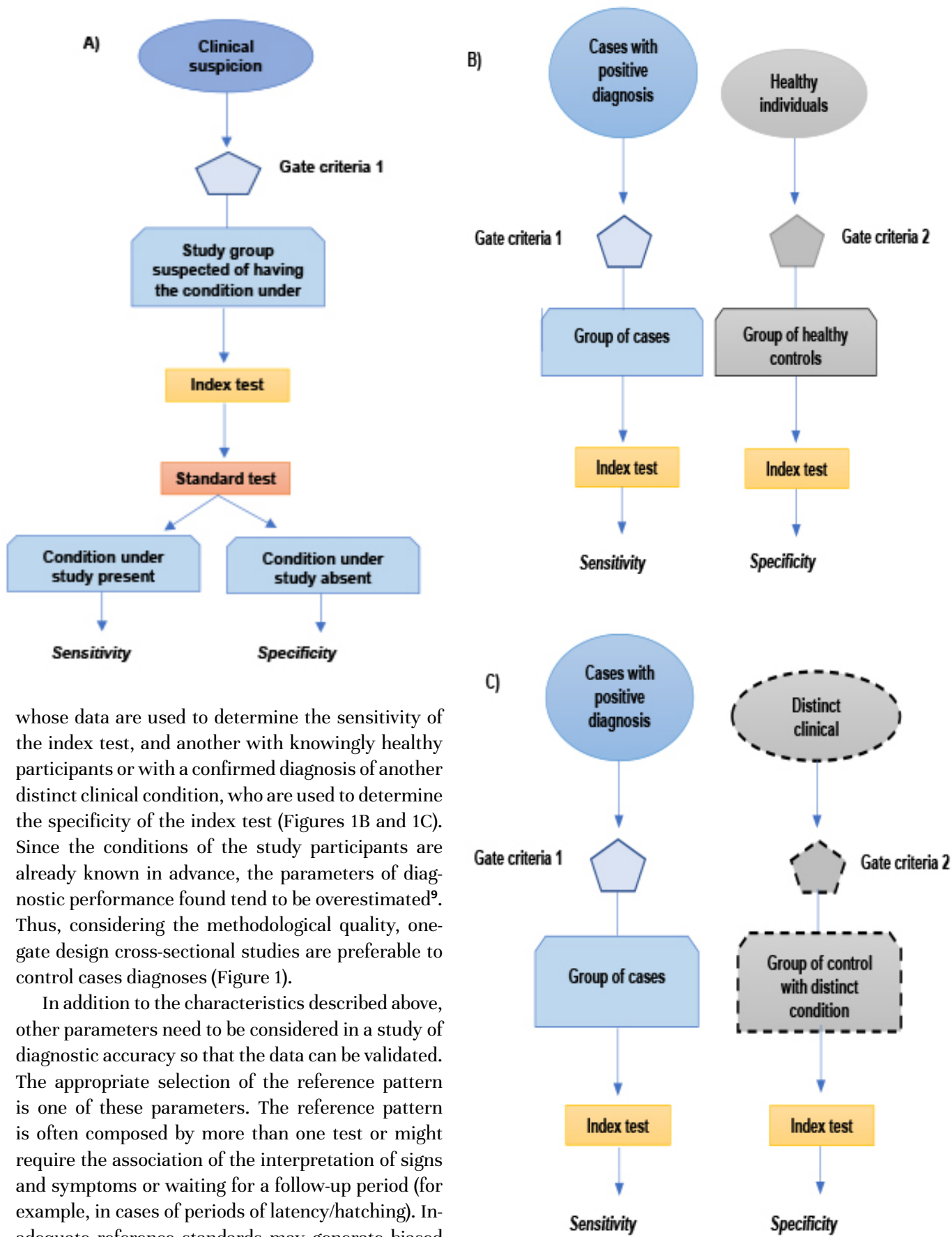
Considering results in clinical practice, the best way to evaluate any diagnostic strategy, especially new ones with supposedly superior accuracy, is

through a randomized clinical trial in which patients are grouped randomly into experimental diagnostic approaches and control, and the clinically relevant outcomes are measured^{7,8}. For example, a randomized study on the diagnosis of tuberculosis compared the applicability and diagnostic accuracy between the Xpert MTB/RIF and the sputum smear microscopy, the latter being the most commonly used⁸. The cell culture was used as the gold standard. The TBscore for positive cultures was similar for both groups, and both showed high values of sensitivity and specificity. However, the Xpert test had more patients with a positive culture who started treatment on the same day and required a shorter treatment time⁸.

As previously mentioned, studies on diagnostic tests rarely focus on this type of outcome, but on those that reflect the diagnostic performance of the test. Most of the literature available is based on specific randomized studies of diagnostic accuracy, which, by definition, is a cross-sectional study. The new test is referred to as the index test in the study, while the test that is already used in clinical practice, formerly the gold standard, is referred to as the reference standard. Since this study design allows variations and may be subject to bias in various stages, the first major challenge encountered by clinicians is the critical evaluation of the quality of the body of scientific evidence. Like any scientific literature, this is vast and needs to be interpreted using criteria as a basis for the decision on estimates that can be considered reliable.

Two main variations of studies of diagnostic accuracy are the most commonly used and they differ, essentially, on the source of inclusion of study participants: one-gate design cross-sectional studies of diagnostic accuracy (Figure 1A) and case-control diagnosis (*two-gate design*) (Figures 1B and 1C). One-gate design studies have a single source of study participants, all suspected of having the disease, but it is not known beforehand if the diagnosis is positive or negative. Thus, all participants in the study are submitted to both tests (index test and reference standard), and the accuracy outcomes of the index test are determined based on the diagnostic findings obtained from the standard reference. Whereas in case-control diagnosis studies (which differ from the concept of epidemiological case-control), two distinct sources of study participants are used: one that includes participants with a positive diagnosis of the disease, obtained using the reference standard,

FIGURE 1. TYPES OF STUDIES OF DIAGNOSTIC ACCURACY. A) CLASSIC DESIGN OF A CROSS-SECTIONAL DIAGNOSIS STUDY; B) CASE-CONTROL DIAGNOSIS USING HEALTHY CONTROLS; C) CASE-CONTROL DIAGNOSIS USING CONTROLS WITH A DISTINCT CLINICAL CONDITION. ADAPTED FROM RUTJES ET AL.¹⁰



whose data are used to determine the sensitivity of the index test, and another with knowingly healthy participants or with a confirmed diagnosis of another distinct clinical condition, who are used to determine the specificity of the index test (Figures 1B and 1C). Since the conditions of the study participants are already known in advance, the parameters of diagnostic performance found tend to be overestimated⁹. Thus, considering the methodological quality, one-gate design cross-sectional studies are preferable to control cases diagnoses (Figure 1).

In addition to the characteristics described above, other parameters need to be considered in a study of diagnostic accuracy so that the data can be validated. The appropriate selection of the reference pattern is one of these parameters. The reference pattern is often composed by more than one test or might require the association of the interpretation of signs and symptoms or waiting for a follow-up period (for example, in cases of periods of latency/hatching). Inadequate reference standards may generate biased

estimates of index test accuracy. In addition, it is extremely important that the results of the tests, both the index test and the reference standard, are interpreted without the prior knowledge of the results of the test previously applied (index or reference), so as not to influence the results, as in more subjective tests, such as imaging. The order in which the tests are conducted can be randomized or known, provided that researchers do not know, in advance, the results of the index and standard tests¹¹. The time between conducting both tests must be sufficiently short so that the severity of the condition does not change. If the period between the reference standard test and index test allows a change in the spectrum of the disease, the performance of the index test may be overestimated.

One way to check the adequacy of the items mentioned above is to critically assess the risk of bias in studies of diagnostic accuracy. The Quadas-2 tool can be used to analyze the risk of bias through four domains: selection of patients, index test, reference standard, and flow and time¹². This tool gives a more transparent analysis of bias, including guiding questions per domain, adding reliability to the findings of the study so that it can be used with more security in clinical practice¹².

ESTIMATING AND INTERPRETING OUTCOMES IN STUDIES OF DIAGNOSTIC ACCURACY

In primary studies of diagnostic accuracy, the experimental test (index test) is compared to the existing diagnostic test (reference standard), known as the best test currently available to identify, with precision, the presence or absence of the condition of interest. The results of both tests are then compared to assess the performance of the index test concerning the reference standard based on its sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios (Figure 2).

The sensitivity of a clinical trial is the proportion of individuals with the condition of diagnostic interest that is correctly identified by the index test concerning the proportion of individuals with a positive diagnosis of the disease obtained using the reference standard (true positives + false negatives by the test index)^{13,14}. Specificity is the proportion of individuals without the condition that is correctly identified by the test index, concerning the proportion of individu-

FIGURE 2. CALCULATION OF SENSITIVITY, SPECIFICITY, POSITIVE PREDICTIVE VALUE AND NEGATIVE PREDICTIVE VALUE, POSITIVE LIKELIHOOD RATIO AND NEGATIVE LIKELIHOOD RATIO OF A CLINICAL EXAMINATION FROM A CONTINGENCY TABLE 2×2 .

		Condition	
		Present	Absent
Test	Positive	True positive (a)	False positives (b)
	Negative	False negatives (c)	True negatives (d)

Sensitivity = $a / (a+c)$
 Specificity = $d / (b+c)$
 Positive predictive value = $a / (a+b)$
 Negative predictive value = $d / (c+d)$
 Positive likelihood ratio: $a / (a+c) / b (b+d)$ or (Sensitivity/ 1- Specificity)
 Negative likelihood ratio: $c / (a+c) / d (b+d)$ or (1 - Sensitivity/Specificity)

als with a negative diagnosis of the disease obtained using the reference standard (true negatives + false positive by the test index)^{13,14}.

The positive predictive value (PPV) of a test is defined as the probability of an individual having the disease, given that they had a positive result in the index test (true positives + false positive). The negative predictive value (NPV) of a test is defined as the probability of an individual not having the disease, given that they had a negative result in index test (true negative + false negative)¹³.

Although the predictive values are more valuable to clinicians, since they provide a direct assessment of the practical usefulness of the test, they are influenced by the prevalence of the condition in the population to whom the test is applied^{13,14}. Sensitivity and specificity, on the other hand, are not affected by the prevalence of the condition, but need to be interpreted in light of the population studied in the study, particularly regarding the spectrum of the disease, presenting limitations to the applicability of the test. For example, when assessing a particular condition in a general asymptomatic population (screening) or in a symptomatic one: a study that evaluated the sensitivity and specificity of the clinical test of breasts in a population of 752,081 women¹⁵ identified a sensitivity of 85.2% and a specificity of 72.9% in symptomatic women *versus* 36.1% and 96.2%, respectively, in asymptomatic women.

The positive and negative likelihood ratios are the best indicators of the usefulness of a clinical trial. These proportions compare the probability of obtaining a positive or negative test result if the individual really has or does not have the condition, respectively. So, the greater the likelihood ratio of a positive test, the more certain it is that a positive test result indicates that the individual has the condi-

tion. A value of 10 or more is considered an indicator that a positive test result is very good at identifying a condition. In the same way, the lower the likelihood ratio of a negative test, the more certain it is that a negative test result indicates that the individual does not have the condition. A value of 0.1 or lower is considered an indicator that a negative test result is very good to dismiss the condition. If the likelihood ratio is close to 1.0, the test result is not a good indicator that the subject has (for a positive test result) or does not have (for a negative test result) the condition of interest. Since the calculation of these ratios uses the four parameters previously described and is not influenced by the prevalence of the condition, they are considered more valuable for clinical decision making¹⁴.

MAKING INFERENCES ABOUT THE IMPACT AND CONSEQUENCES OF A DIAGNOSTIC TEST

Knowing how to interpret studies of diagnostic accuracy is an essential step, but that is not enough for recommending a test for clinical practice. As previously discussed, the preference should be for randomized clinical trials that evaluate the diagnostic strategies in outcomes that are relevant for the patient.

In the absence of those in the literature, the search must be for studies of diagnostic accuracy, preferably through systematic reviews and meta-analyses of performance estimates. In this sense, the synthesis of evidence in diagnostic tests is particularly challenging because the statistical methods used to aggregate data of diagnostic accuracy are conceptually complex (binomial and hierarchical models), often leading to difficulties in the interpretation of the results¹⁶.

Additionally, evidence from studies of diagnostic accuracy is indirect. The clinician will need to make inferences about the probable impact and consequences of the test in outcomes that are important to the patient¹⁷. This situation also becomes a challenge, since many times there is no direct link between the results of diagnostic accuracy and a relevant outcome, especially when the test is part of an algorithm.

Generally, the inference requires the availability of an effective treatment⁷. However, even without an effective treatment, accurate testing can be benefi-

cial to improve the well-being of patients with prognostic information, to reduce their anxiety, or even prevent them from being subjected to more unnecessary testing.

There are clinical situations in which it is more important to exclude a certain disease that has serious consequences if not identified. In these cases, the use of screening tests with high sensitivity and high negative predictive value may be of more value than the actual diagnosis of the disease, which becomes less important. This is the case, for example, with natriuretic peptides in congestive heart failure, given the consequences of the disease and the challenge of confirming it in patients with suspicion of failure, which occurs in only 40%-50% of cases¹⁸.

When the treatment for a disease has high associated risks, in conditions of a difficult diagnosis, or when higher diagnostic precision is needed, multiple tests are often used. Thus, clinicians may be interested in additional benefits of a new adding-on test. An example of that is the use of computed tomography angiography tests additional to computed tomography without contrast in acute ischemic strokes¹⁹.

Therefore, it is necessary to have a clear understanding of where this new test will be inserted along the diagnostic process, to achieve a balance between the desirable and undesirable consequences of the test. The inferences about the probable impact and consequences of the test in outcomes that are important to the patient may be carried out by simulating data in a hypothetical scenario of 1,000 patients, for example, using the prevalence data of the condition of interest or the pre-test probability of the patient having the disease based on signs and symptoms²⁰. It is a simplified approach that classifies the test results to produce true positives and true negatives, false positives and false negatives cases, as in the following example.

Consider the consequences of replacing invasive angiography by multi-faceted spiral computed tomography (MCT) for the diagnosis of coronary artery disease (CAD). Suppose that the clinician wants to consider replacing it in patients in their clinical practice, which has 30% of patients with CAD.

The estimates of diagnostic accuracy of the computed tomography compared to angiography for CAD were obtained by identifying the meta-analysis of Hamon et al.²¹, which included data from 27 studies (2,024 patients) corresponding to a cumulative number of 22,798 segments evaluated. This study

evidenced the summary estimates of 0.96 sensitivity (CI 95% 0.94-0.98), 0.74 specificity (0.65 to 0.84), 5.36 positive likelihood ratio (3.45 to 8.33), and 0.05 negative likelihood ratio (0.03 to 0.09).

Considering the context of service, it would be expected that the conventional coronary angiography would diagnosis 300 cases of CAD for every 1,000 patients treated and 700 patients would be diagnosed with the absence of CAD using angiography. Based on the estimates from the meta-analysis, a MCT sensitivity of 96% represents the correct diagnosis of 288 out of 300 patients (true positives). In the same way, a specificity of 0.74 would represent the identification of 518 out of 700 non-cases of CAD (true negatives). Consequently, we would have 12 false negative and 182 false positive results (Figure 3).

FIGURE 3. DISTRIBUTION OF CASES OF TRUE POSITIVES, TRUE NEGATIVES, FALSE POSITIVES, AND FALSE NEGATIVES IN THE HYPOTHETICAL SCENARIO OF 1,000 PATIENTS AND SERVICE POPULATION WITH CAD PREVALENCE OF 30%. CAD: CORONARY ARTERY DISEASE; MCT: MULTIFACETED COMPUTED TOMOGRAPHY

		CAD (angiography)		
		Present	Absent	Total
MCT	Present	288	182	470
	Absent	12	518	530
	Total	300	700	1000

Table 1, below, shows the approach of the GRADE group (Grading of Recommendations, Assessment, Development and Evaluation) to assess the consequences of the diagnostic strategies adopted in the above example. In this table, each possible outcome of the tests assumes a clinical/practical consequence. Thus, for the true positive cases, the patients would receive the treatments of known effectiveness (drugs, angioplasty, and stents, revascularization surgery) as clinical conduct, and the true negative cases would be spared the treatments and unnecessary use of resources. On the other hand, the false positive cases (26% of patients who would have a negative diagnosis using angiography) would be subjected to the unnecessary risks of the adverse events associated with the treatments (use of medication, segmental angioplasty, etc.), in addition to more anxiety and unnecessary loss of quality of life, while the false-negative cases would not receive adequate treatment that could help reduce the risk of subsequent coronary events, representing a loss of diagnosis in 4.0% of the patients with CAD.

Thus, it is relatively certain that minimizing false positives and false negatives will benefit patients. However, the complications of invasive angiography - cardiac catheterization, myocardial infarction, death -, although rare, are undoubtedly important.

TABLE 1. EXAMPLE OF THE CONSEQUENCES FOR THE PATIENT OF BEING CLASSIFIED ACCORDING TO THE PERFORMANCE OF THE TEST. SOURCE: HSU ET AL.⁴

Research question 1: What are the clinical implications of replacing invasive angiography with multifaceted computed tomography in the diagnosis of coronary artery disease?	
Population	Patients with suspected coronary artery disease (CAD)
Intervention	Multifaceted computed tomography (MCT)
Compared to	Conventional coronary angiography
Outcomes	
True positive	The patient will be submitted to the MCT, and the result will be positive; the patient will receive the treatments of known effectiveness (drugs, angioplasty, and stents, revascularization surgery) as the clinical conduct; there will be savings in expenses with the test
True negative	The patient will be submitted to the MCT, and the result will be negative; they will not have to go through other tests or unnecessary treatments, and there will be savings in expenses with the test
False positive	Anxiety, time and unnecessary expenses for the patient and family, unnecessary confirmatory tests; patients may be subjected to unnecessary risks of adverse events associated with treatments and loss of quality of life
False negative	The patient may have serious consequences from the disease, such as acute myocardial infarction, which may result in death; the true cause of the symptoms (i.e., CAD) will not be detected, leading to investigations and unnecessary treatments; an impediment to receiving the treatment
Inconclusive results	Time and resources spent; patients subjected to unnecessary risks, likewise the consequence of false negative and false positive
Tests complications	Both tests require the use of contrast in the patient, which can cause allergies, for example. Angioplasty is an invasive procedure, with the risk for cardiac catheterization
Use of resources (costs)	Angiography costs are higher than those of MCT, and it demands great resource utilization

CAD: coronary artery disease; MCT: multifaceted computed tomography.

Therefore, experts need to consider the desirable and undesirable consequences of diagnostic tests for patients, for making rational decisions about which test to recommend in clinical practice. The link between accuracy of results and clinical outcomes depends on this weighting, as well as the cognitive or emotional impact resulting from the knowledge of the diagnosis²².

Other aspects should be part of the rational use of diagnostic tests, some of which have been already explored throughout the article, such as: the quality of the scientific evidence of studies available, the impact of the imprecision of the findings between studies, if the populations included in the studies correspond to the population that is the focus of the clinical decision making, the costs associated with the strategies and values and preferences of the patient⁵. Based on the example used, MCT could reduce the chance of adverse events of invasive angiography, which includes serious events, although rare, in addition to being less costly than angiography. On the other hand, the large number of false positives and the risk of losing patients with coronary artery disease treated effectively could lead to the patient's preference for the angiographic approach, even though it is more invasive.

CONCLUSIONS

Diagnostic tests are used to identify a condition of interest with the intention of providing adequate treatment and can be used as a diagnostic screening, replacing a diagnostic test, or be added to an existing test.

The appropriate decision regarding the adoption or recommendation of a diagnostic test should be based on evidence from randomized clinical trials, in which the diagnostic strategies are compared using

measurements of clinical outcomes that are relevant to the patients.

However, such literature is scarce, and most of the commonly found evidence refers to estimates of the diagnostic performance of tests, measured using studies of diagnostic accuracy. Cross-sectional studies involving patients suspected of having the condition, without previous knowledge of the diagnosis status, provide the evidence of higher methodological quality when compared to diagnosis case-control designs.

Considering the accuracy parameters that are measured by this type of study, the positive and negative likelihood ratios are the best indicators of the usefulness of a clinical trial, because they can estimate a post-test probability of disease.

Clinical decision-making must, therefore, use indirect evidence from studies of diagnostic accuracy to make inferences regarding the impact and consequences of important results for the patient. The connection between accuracy outcomes, such true positives and false positives, and clinical results depends on the balance between benefits and damages of the tests available, as well as on the cognitive or emotional outcomes resulting from the knowledge of the diagnosis.

In addition, it is crucial that the clinician has knowledge in epidemiology and evidence-based health care, to properly evaluate the quality of evidence available, identify similar characteristics of the populations included in the studies in comparison with the population of the clinical practice and apply the estimates for the rational use of diagnostic tests.

Finally, considerations on resources should also be explored, as well as the preferences and values of the patient, who should participate in clinical decision making whenever possible.

RESUMO

OBJETIVO: Auxiliar os clínicos na interpretação adequada das evidências científicas de estudos que avaliam testes diagnósticos, de modo a permitir seu uso racional na prática clínica.

MÉTODOS: Revisão narrativa da literatura dos principais conceitos, desenhos de estudo, interpretação adequada dos dados de acurácia diagnóstica e realização de inferências sobre o impacto do teste diagnóstico na prática clínica.

RESULTADOS: A maioria da literatura que avalia o desempenho de testes diagnósticos utiliza como delineamento os estudos transversais. Ensaios clínicos randomizados, avaliando desfechos clínicos, que seriam considerados ideais, são escassos. Os estudos transversais mensuram desfechos de acurácia diagnóstica que são considerados indiretos e insuficientes para definir o real benefício para os pacientes. Dentre os desfechos, as razões de verossimilhança positiva e negativa são as mais úteis para a decisão da conduta clínica. Variações no delineamento transversal do estudo, que podem acrescentar vieses aos resultados, bem como outros domínios que contribuem para diminuir a confiabilidade dos achados, são discutidos, além de como extrapolar tais achados de acurácia em impacto e

consequências consideradas importantes para o paciente. Aspectos sobre custos, tempo para a obtenção do resultado, preferências e valores dos pacientes devem, preferencialmente, participar da tomada de decisão.

CONCLUSÃO: Conhecer a metodologia dos estudos de acurácia diagnóstica é fundamental, porém não suficiente, para o uso racional de testes diagnósticos. Há a necessidade de se ponderarem as consequências desejáveis e indesejáveis dos resultados dos testes para os pacientes, de modo a favorecer a tomada de decisão racional acerca de qual teste recomendar na prática clínica.

PALAVRAS-CHAVE: Tomada de decisão clínica. Testes diagnósticos de rotina. Prática clínica baseada em evidências. Sensibilidade e especificidade. Valor preditivo dos testes. Equipamentos para diagnóstico.

REFERENCES

- White S, Schultz T, Enameh YAK. Synthesizing evidence of diagnostic accuracy. Philadelphia: Lippincott Williams and Williams; 2011.
- Bouza E, Torres MV, Radice C, Cercenado E, Diego R, Carrillo-Sánchez C, et al. Direct E-test (AB Biodisk) of respiratory samples improves antimicrobial use in ventilator-associated pneumonia. *Clin Infect Dis*. 2007;44(3):382-7.
- Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*. 2006;332(7549):1089-92.
- Hsu J, Brozek JL, Terracciano L, Kreis J, Compalati E, Stein AT, et al. Application of GRADE: making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implement Sci*. 2011;6:62.
- Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*. 2012;344:e686.
- Hofmann B, Welch HG. New diagnostic tests: more harm than good. *BMJ*. 2017;358:j3314.
- Bossuyt PM, Lijmer JG, Mol BW. Randomized comparisons of medical tests: sometimes invalid, not always efficient. *Lancet*. 2000;356(9244):1844-7.
- Theron G, Zijenah L, Chanda D, Clowes P, Rachow A, Lesosky M, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet*. 2014;383(9915):424-35.
- Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174(4):469-76.
- Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem*. 2005;51(8):1335-41.
- Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36.
- Altman DG. *Practical statistics for medical research*. London: Chapman & Hall; 1991.
- Davidson M. The interpretation of diagnostic test: a primer for physiotherapists. *Aust J Physiother*. 2002;48(3):227-32.
- Bobo JK, Lee NC, Thames SF. Findings from 752,081 clinical breast examinations reported to a national screening program from 1995 through 1998. *J Natl Cancer Inst*. 2000;92(12):971-6.
- Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008;149(12):889-97.
- Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med*. 2006;144(11):850-5.
- Roberts E, Ludman AJ, Dworzynski K, Al-Mohammad A, Cowie MR, McMurray JJ, et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. *BMJ*. 2015;350:h910.
- Sabarudin A, Subramaniam C, Sun Z. Cerebral CT angiography and CT perfusion in acute stroke detection: a systematic review of diagnostic value. *Quant Imaging Med Surg*. 2014;4(4):282-90.
- Fletcher RH, Fletcher SW, Fletcher GS. *Diagnostics*. In: *Clinical epidemiology: the essentials*. 5th ed. Baltimore: Wolters Kluwer; 2012.
- Hamon M, Biondi-Zoccai GG, Malagutti P, Agostoni P, Morello R, Valgimigli M, et al. Diagnostic performance of multislice spiral computed tomography of coronary arteries as compared with conventional invasive coronary angiography: a meta-analysis. *J Am Coll Cardiol*. 2006;48(9):1896-910.
- Segal JB. Choosing the important outcomes for a systematic review of a medical test. *J Gen Intern Med*. 2012;27(Suppl 1):S20-7.

