Review

Beyond Diagnostic Accuracy:
The Clinical Utility of Diagnostic Tests

Patrick M.M. Bossuyt,[1*] Johannes B. Reitsma,[1,2] Kristian Linnet,[3] and Karel G.M. Moons[2]

Like any other medical technology or intervention, diagnostic tests should be thoroughly evaluated before their introduction into daily practice. Increasingly, decision makers, physicians, and other users of diagnostic tests request more than simple measures of a test's analytical or technical performance and diagnostic accuracy; they would also like to see testing lead to health benefits. In this last article of our series, we introduce the notion of clinical utility, which expresses—preferably in a quantitative form—to what extent diagnostic testing improves health outcomes relative to the current best alternative, which could be some other form of testing or no testing at all. In most cases, diagnostic tests improve patient outcomes by providing information that can be used to identify patients who will benefit from helpful downstream management actions, such as effective treatment in individuals with positive test results and no treatment for those with negative results. We describe how comparative randomized clinical trials can be used to estimate clinical utility. We contrast the definition of clinical utility with that of the personal utility of tests and markers. We show how diagnostic accuracy can be linked to clinical utility through an appropriate definition of the target condition in diagnostic-accuracy studies.

© 2012 American Association for Clinical Chemistry

In the clinical evaluation of diagnostic tests and markers, diagnostic accuracy plays a pivotal role. Previous articles in this series (1–3) have explained how accuracy is defined by comparing the results of the test or marker under evaluation with those of the clinical reference standard in the same patients. The reference standard is the best method for establishing the pres-

ence or absence of disease or, more generally, the target disease or condition.

Increasingly, decision makers, physicians, and other users of diagnostic tests request more information than simple measures of the technical and analytical performance of the tests and their diagnostic accuracy. Healthcare policymakers have long called on manufacturers to shift from a narrow technical or biomedical perspective to a wider one, one that considers whether the diagnostic technology improves final outcomes in typical patient populations (4). Before recommending the use of diagnostic tests and markers and before deciding on their reimbursement, decision makers and users now want to see evidence that testing actually improves outcomes in relevant patient populations or that it enhances healthcare quality, efficiency, and cost-effectiveness.

We introduce the notion of "clinical utility," and we discuss how randomized trials and other study designs can be used to evaluate the clinical utility of diagnostic tests.

Clinical Utility: Effect on Health Outcomes

Numerous authors have proposed a hierarchical system for looking at the evidence supporting the use of diagnostic tests and phased approaches to their evaluation (5). Diagnostic accuracy always figures prominently in these systems, but other desiderata occur at higher levels. The utility of diagnostic testing is usually introduced at these higher levels, or it is positioned at a later stage in the evaluation of tests or markers. Awareness that testing should not only be informative but also be *usefully* informative has been increasing, especially in this age of genomic and metabolic testing.

The concept of utility has been defined somewhat ambiguously in evaluations of healthcare. At one extreme, we find measures of how well clinicians appreciate the information of diagnostic testing or perceive it as useful. In the oft-quoted Fryback–Thornbury hierarchy of diagnostic testing, which had initially been introduced for imaging, this measure is called "diagnostic thinking efficacy" (level 3) and "therapeutic efficacy" (level 4) (6). These perceptions have been criticized because of their subjective nature and their limited validity. Intended behavior, which is expressed when clinicians announce that the information from

[1] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; [2] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands; [3] Section of Forensic Chemistry, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.

* Address correspondence to this author at: Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Rm. J1b-217, P.O. Box 22700, 1100 DE Amsterdam, the Netherlands. Fax +31-20-691-2683; e-mail p.m.bossuyt@amc.uva.nl.

diagnostic testing would change their practice and decision-making, is not always seen to reflect actual behavior (5). Furthermore, a change in thinking or practice may be a prerequisite for a change in patient outcome, but it is not necessarily sufficient.

Essentially, the utility of diagnostic testing is defined as the degree to which actual use of the corresponding test in healthcare is associated with changing health outcomes, such as preventing death and restoring or maintaining health. This utility is called "patient outcome efficacy" (level 5) in the Fryback–Thornbury hierarchy. Tests then have to be evaluated by looking at the degree to which their use leads to actual changes in outcomes that matter to patients. This means that tests are used to identify patients for whom randomized trials or other types of evidence have shown that a specific form of clinical management is more effective than the alternative. An example is the diagnosis of pulmonary embolism, for which anticoagulation therapy has been shown to be effective in reducing mortality and morbidity.

Defining changes in health outcome as the criterion for decisions about diagnostic tests puts them at the same level as any other intervention in healthcare. In the spirit of evidence-based medicine, we find it self-evident these days that drugs should be safe and effective before they are introduced into the market; we take it for granted that guideline developers look at the effects of drugs on patient outcomes before recommending their use; we endorse that decisions about reimbursement consider value for the money and cost-effectiveness. Similarly emphasizing the utility of diagnostic testing in decision-making puts diagnostic tests on the same table: Decisions about diagnostics do not differ from those about other interventions in healthcare.

As the Fryback–Thornbury hierarchy shows, different terms, not just "clinical utility," are used to refer to this concept. Other authors have referred to the value of testing when describing the effects on outcomes, or they have referred to its usefulness or to the benefits and gain of testing. In decision analysis, the utility of an option is a quantitative expression of the chance-adjusted value of its possible consequences. In economic evaluations in healthcare, the expected utility of an intervention is often expressed in terms of the expected number of life years gained and often adjusted for the quality of those years. Because of this association with decisions and an explicit evaluation of their consequences—preferably expressed explicitly and in a quantitative sense—we prefer the term "utility" in this article. The addition of the qualification "clinical" is explained below.

Table 1. Key features of clinical utility.	
Elements of clinical utility	Explanation
Health outcomes	Health outcomes are outcomes that matter to patients and society: to prevent premature death, to restore or maintain functional health.
Strategy	Outcomes are generated not only by testing only but also by a management strategy that starts with testing but includes all downstream consequences of subsequent clinical management.
Probabilistic	Not all outcomes will be observed in everyone tested; evaluations will be made at the group level and expressed in terms of a distribution of outcomes.
Comparative	Utility is defined relative to a comparator strategy: current best standard practice.

Clinical Utility: Defining Features

From these elements, we can define several key features for evaluations of the utility of diagnostic testing. These elements are summarized in Table 1. First, the definition of utility should be in line with the general purpose of healthcare. Without entering into a philosophical or political debate, one could argue that the central purpose of healthcare in our world is to allow people to maintain or regain functional health. The primary purposes for using diagnostic tests—and all other forms of testing, including screening and monitoring tests—should therefore be to prevent premature death and suffering and to restore functional health.

The majority of tests simply generate information, and information itself does not generate a health benefit. In most cases, improvements in health outcomes from diagnostic testing will be generated by the way test results are used to guide downstream management, such as decisions to initiate, modify, stop, or withhold treatment. Clinical care is effective if it uses the right form of treatment in the right group of patients. Testing can help to select the most effective form of treatment and to identify the appropriate group of patients. The mechanisms for generating improved outcomes after testing are manifold, however, and include several other pathways (7).

Clinical utility is not defined only in terms of the potential benefits of subsequent or directed treatments. It also requires an evaluation of the full range of the effects that tests or markers may have on patients. Elsewhere we have described how these effects cover emotional, social, cognitive, and behavioral effects (7). Emotional effects from diagnostic testing are probably best studied: They cover the anxiety and stress associ-

ated with testing and waiting for results. Social effects relate to social relationships: Testing could lead to stigmatization and social isolation. Cognitive effects include patients' beliefs, perceptions, and understanding about their diagnostic test result and condition. Behavioral effects cover people's engagement in other health behaviors, such as following a healthy diet, participating in exercise, and smoking.

Thus, a test may have clinical utility in the absence of effective clinical treatment. Testing can lead to an early resolution of uncertainty, for example, which can itself be positive (or negative). A test that leads to a diagnosis—and the associated prognosis—may help patients cope with their illness, thereby improving daily functioning and quality of life. These benefits may extend to the patient's partner, friends, and relatives, as in a diagnosis of Alzheimer disease.

These effects on health outcome can be positive or negative, and they can be intended or unintended. They not only may have relevance to patients themselves but also may influence the cost-effectiveness of clinical management. DNA testing, for example, may increase the effectiveness of a program targeted at healthy behavior, but it can also reduce patients' expectations that behavioral changes would be effective (8).

It is possible that a diagnostic test has clinical utility without actually improving the primary health outcomes targeted by subsequent treatment. If the introduction of a test or marker leads to health outcomes comparable to those obtained with current standard diagnostic–treatment strategies but these outcomes are achieved in a simpler way, there may be clinical utility. Most likely, patients will be prevented from undergoing unnecessary additional diagnostic testing or ineffective treatment, as well as the associated burden.

A further defining feature of clinical utility is that it will necessarily be probabilistic. No matter how reliable the diagnostic test itself is, there is no certainty that testing will lead to the desired outcomes. That may be due to limitations in the testing procedure itself, but it can also be caused by variability in the test results (not everyone gets the same result). Differences in management (not everyone with the same result is managed in the same way) and variability in response (not everyone with the same management gets the same outcome) further convolute this issue. At best, we can describe the utility in aggregate terms as the expected outcome in a group of people for whom we consider using the test. Again, this feature is not unlike other forms of evidence-based medicine.

A final but essential element is that the definition of clinical utility has to be a relative one. The utility of a diagnostic testing strategy cannot be defined in an absolute sense; it has to be evaluated relative to that a comparator strategy. In most cases, this comparator strategy will be the current best practice for dealing with presenting patients. This strategy could be the current best testing strategy for the same purpose or not using any form of diagnostic testing all. In thinking about the utility of a test or marker, we do not just evaluate whether using the test leads to the outcomes desired. We also compare the outcomes of using the test with the outcomes of using the current best alternative, in similar patients. The utility of fecal occult blood–based population screening for colon cancer is defined by comparing it to, say, no screening (9). The utility of high-sensitivity C-reactive protein testing to aid in the identification of an acute coronary syndrome in patients admitted to a chest pain unit can be defined only relative to the standard best practice, according to the clinical assessment and measurements of conventional cardiac markers.

The clinical utility of a diagnostic test is both consequential and contextual. It refers to the consequences of using tests or markers. It also refers to and depends on the clinical context: With changes in the available management options and with new and alternative forms of testing, the utility of a diagnostic test may change (10–13). The same consideration applies to advances in knowledge, which could lead to changes in thinking about effective treatment and could affect the clinical utility of testing, although the testing technology itself may be unaffected. New interventions could lead to the disappearance of a previously useful diagnostic test, but they could also make a new test interesting, owing to the potential for the test and the management strategy concerned with improving health outcomes.

Clinical utility is often discussed in terms of the validation or qualification of medical tests, but these terms are slightly more general, in that they refer to a related but different evidentiary basis. The term "validation" encompasses many different aspects of diagnostic test and marker development. Thus, we can distinguish between analytical and clinical validation. Clinical validation is the process through which one shows that test results are clinically meaningful, i.e., finding whether the test is able to detect or predict the disorder or condition of interest in targeted patient groups (14).

The "qualification" of a test refers to the process through which one concludes that its results can be relied on to have a specific interpretation within a stated context of use (15). This definition was proposed by the US Food and Drug Administration Critical Path Initiative in 2004 for the use of biomarkers in drug development and regulatory review. "Qualification" then means evaluating whether the medical test or biomarker is "fit for purpose," i.e., whether it can be applied for a specific proposed use.

No Need for Personal Utility

As healthcare has expanded in the Western world, the aims of healthcare have also diversified, and authors have pleaded that the definition of the utility of diagnostic tests be widened. Pleas for more novel genetic and molecular markers are especially heard. These discussions have led to the introduction of a new term, "personal utility," which is sometimes also referred to, somewhat confusingly, as "social utility" *(16, 17)*. We believe that this new terminology is unnecessary and potentially misleading.

This definition seems to have been introduced to explore different aspects of novel markers and tests: first, to help justify tests that do not seem to have clinical utility as we have defined it above and, second, for personal decisions, rather than clinical ones, about ordering or undergoing these tests.

Some tests at first do not seem to have clinical utility, but their use seems justified because of other benefits. Testing for Huntington disease, for example, offers no options for effective clinical management but can be of help in life planning and reproductive decisions. Other tests may help to clarify the origins or risk factors for a specific disease or condition. Test results can then help patients to better understand their disease or risk of disease *(18)*.

In our view, both the purposes for testing and the corresponding effects of that testing can be subsumed under the definition of the clinical utility of medical testing proposed earlier. Intended or unintended cognitive effects of using diagnostic and other medical tests should be included in evaluations of clinical utility. In this sense, there is no need for a different definition. That does not mean that all tests have clinical utility. If there are no effects on patient outcome, i.e., effects that could be noted by patients in their personal diaries, then there cannot be any form of utility.

The second aspect of personal utility refers to the right of individuals to decide what is useful for them and what is not *(16, 17)*. Some have argued that it would be rather paternalistic for clinicians to presume to know what is in a person's best interest *(19)*. If someone is willing to pay for the marker test, then it seems the patient has a right to have it done. Nevertheless, testing may have downstream healthcare consequences, with societal costs and the potential for incorrect or misguided understanding, misdiagnosis, and mistreatment for the individual.

Consideration of what to include in an evaluation of testing clearly touches on a very basic discussion about the purposes of healthcare and the boundaries between the individual and the societal perspective. Most likely this debate will continue in the coming years, with those emphasizing that there is only a single definition of utility—as defined above—holding the strongest cards.

## Study Designs for Evaluating Utility: Randomized Trials

The defining elements of utility in Table 1 can be used to consider and select study designs for evaluating the utility of diagnostic and other types of medical tests. In developing a management strategy, we need to look at health outcomes in the broader sense across groups of patients and compare these outcomes with those in similar patients for which the best available alternative strategies have been used. In this approach, randomized trials would be the eminent type of study for evaluating the utility of tests.

Randomized clinical trials always compare 2 or more management strategies in groups of people. The strategies concentrate on the same outcome while keeping as many other things as similar as possible via the use of placebos and/or sham procedures and with blinded, objective assessments of outcomes. Randomized trials suffer less from bias than most other types of study design and offer the evidence needed for decision-making: the effects on patient outcomes from diagnostic (or other types of) testing. Fig. 1 offers an example of a randomized clinical trial of testing.

Mueller and colleagues, for example, compared 2 diagnostic strategies for patients who present to the emergency department with acute dyspnea *(20)*. One involved measuring B-type natriuretic peptide concentrations with the use of a rapid bedside assay, and the other diagnostic strategy used the conventional workup. As the report in this journal showed, there was no difference in cumulative all-cause mortality at 720 days (37% vs 36%) *(21)*.

To be informative, randomized trials of testing should have a well-defined protocol that links test results to specific clinical responses, such as "treat if test is positive; discharge if negative," or "additional testing if test is positive; discharge if negative." The protocol should be based on the current best evidence—from randomized trials or other strong study designs—about the effectiveness of these downstream forms of clinical management. Without such a protocol, a randomized trial of testing may not be informative and generalizable, because it depends on local opinion, thereby reflecting variability in practice *(11)*.

There are a number of additional concerns, disadvantages, and inefficiencies when designing trials for laboratory tests or other forms of testing *(10–13)*. Trials of tests evaluate not only the consequences of the diagnostic test but also the full testing strategy by which the test is integrated into an approach for a clinical-management protocol that specifies how the tests are

Fig. 1. A schematic representation of a randomized trial comparing 2 test strategies.
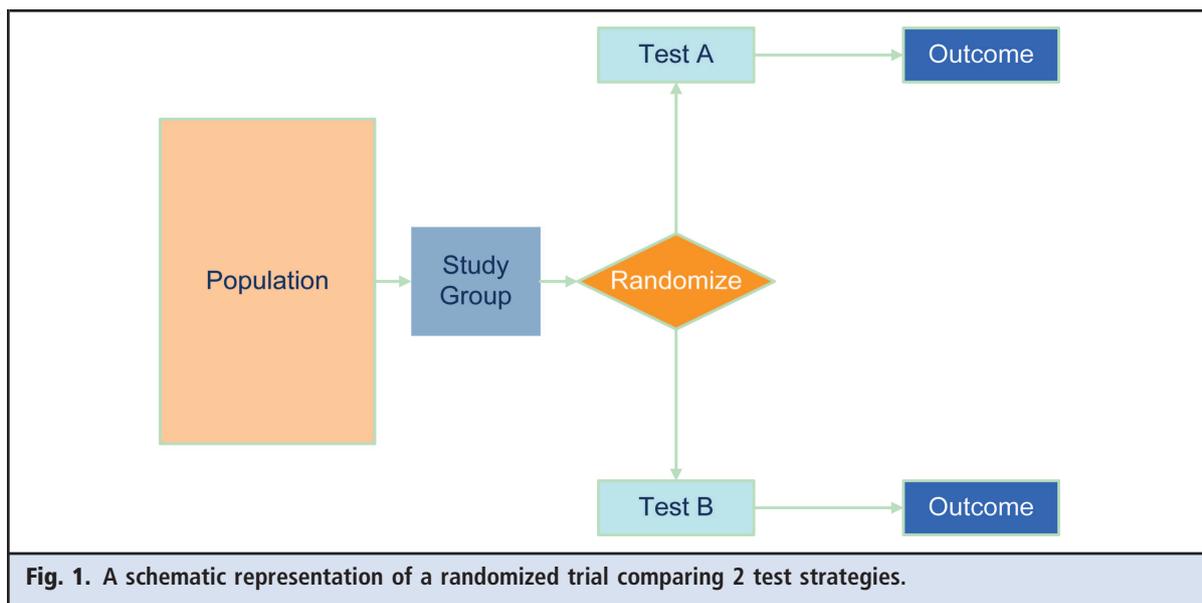
used to guide downstream management decisions. Therefore, the effects of testing will depend not just on the test but also on the effectiveness of downstream management. That is, in itself, a good thing, but a possible result could be that a mediocre test improves outcomes or cost-effectiveness when it is coupled with effective management. Conversely, a quality test could fail to improve outcomes in the absence of effective management. Fecal occult blood screening, mentioned earlier, is an example of the former, and tests for Alzheimer disease are an example of the latter.

Trials of medical tests or markers can also be rather inefficient *(11)*. In a randomized trial of a new drug, everyone in the control group receives, say, a placebo, and every participant in the experimental group receives the active drug. The contrast between the 2 groups is maximal, but that does not mean that everyone in the experimental group will benefit from the drug. In a trial of preventive treatment, only those at risk for the event of interest will have the potential to benefit from the intervention.

Trials of tests and markers, on the other hand, usually do not offer such a maximal contrast. If patient outcome depends on clinical management and test results are used to guide decision-making, then only patients with discordant test results will be managed differently, depending on the outcome of the randomization procedure. In a trial of a screening test, for example, the initial test—such as fecal immunochemical testing in screening for colorectal cancer—is used to decide on further management. Only individuals positive for the test, not all screening-trial participants, will be referred for colonoscopy. Usually, that implies that screening trials and other trials of medical tests, includ-

ing diagnostic tests, need large numbers of participants. Achieving these numbers can be difficult. The MARVEL (Marker Validation for Erlotinib in Lung Cancer) trial sponsored by the National Cancer Institute had intended to recruit 1200 patients in 4 years but had to permanently close the trial because of poor accrual, even though 300 sites had institutional review board approval to enroll patients.

Alternative and potentially more efficient randomized designs for testing trials are possible. Such designs include randomizing only individuals with discordant test results in a comparison of 2 forms of testing, or randomizing only individuals testing positive to different treatment strategies in evaluations of a single test *(11)*.

Study Designs for Evaluating Utility: Models

Given the complexities in designing randomized trials of diagnostic tests, some researchers have turned to other study designs for evaluating the utility of tests and markers. One way to obtain estimates of the expected outcomes in a group of patients after testing relative to another strategy is to use modeling *(13, 22, 23)*. Decision analysis models rely on parameter estimates obtained from other study designs. Starting from the available management options, model builders define links between testing and other actions, health states, future events, and outcomes. They then specify probabilities for corresponding transitions across these links. Using the resulting model, researchers can estimate the probabilities of the final health outcomes, as well as the average or expected outcome, for specific management strategies. Fig. 2 presents a
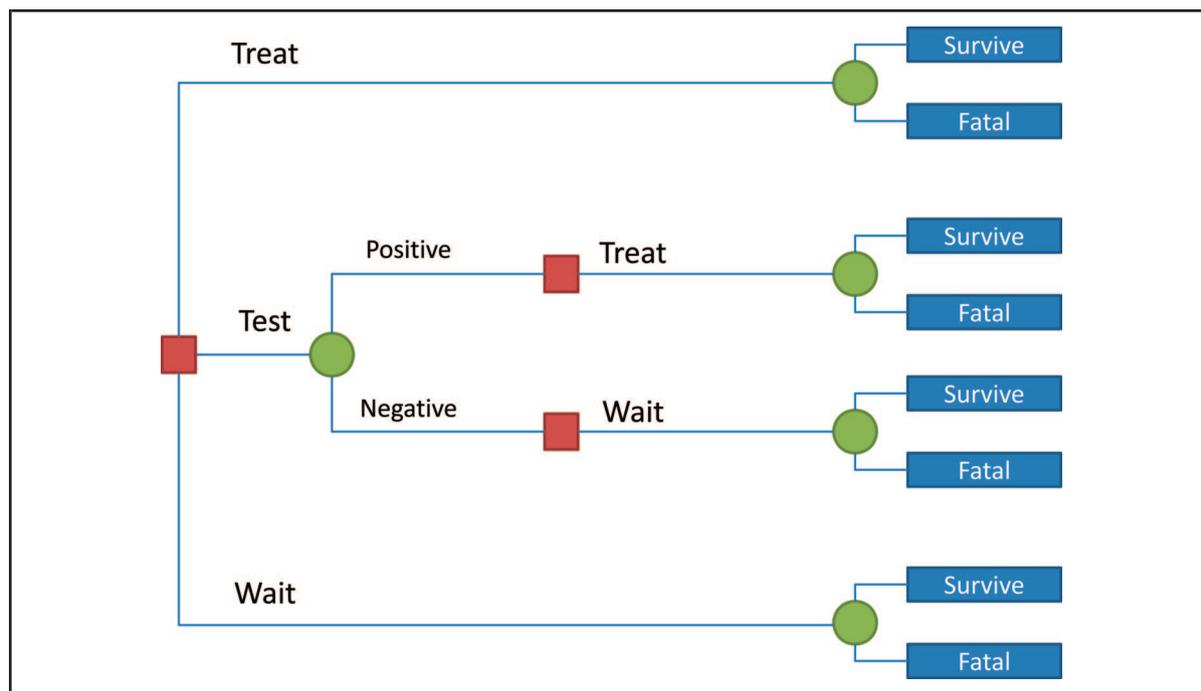
Fig. 2. A simple generic decision model to compare 3 strategies.

very simple, iconic decision model for comparing 3 strategies: a test-and-treat strategy, a treat-all strategy, and a wait-in-all strategy.

Such decision analysis models have been used extensively in evaluating and comparing cancer-screening programs (24). Models can be complex and time-consuming to build but may offer efficient alternatives to costly trials. Downsides are the critical nature of assumptions and model builders' limited ability to include the unexpected. Table 2 highlights key differences between randomized trials and models.

Table 2. Key differences between trials and models for evaluating the utility of medical tests.

Randomized trials	Models
Can compare only a few strategies	Can compare many strategies
Evaluates intended and unintended effects	Evaluates modeled effects only
No assumptions	Assumptions necessary
Expensive	Less costly
Time for follow-up needed	Model-building time only

Target Condition: Linking Accuracy to Utility

Amidst discussions about how to manage rising health-care costs and the role of new technologies in driving this trend is an ongoing debate over the evidence standards that should be applied for introducing new diagnostics into routine clinical practice. The differences in opinion concern the extent to which information other than findings from randomized controlled trials can be used to support the use of tests and markers. Do we need randomized trials, do we require mathematical models, or do mechanistic reasoning and findings from pharmacologic and observational studies suffice? As elsewhere in medicine, randomized trials are mandatory in principle; however, randomized trials can be justified only by the principle of equipoise: genuine uncertainty in the expert medical community over whether testing is beneficial.

Given that randomized controlled trials can be costly and often provide answers only after an intolerable delay, the challenge is to find shortcuts. If one test produces results that are identical to those of an existing but more laborious test, do we need evidence from randomized trials that introducing the test does not compromise patient outcome? Probably not. A full discussion of the need for randomized trials to document the clinical utility of testing is beyond the scope of this article (11, 12, 23).

How does diagnostic accuracy, as defined in the previous articles in this series, relate to clinical utility? For diagnostic tests, accurately identifying the patients with the suspected disease is usually a necessary condition for clinical utility *(10)*. Identification of disease, however, does not equate with health benefit. Severity and clinical consequences vary across cases and management strategies for many diseases and are rarely uniform. Not all forms of disease are consequential, and not every disease case may benefit from treatment. Furthermore, the accuracy of a diagnostic test varies according to the spectrum of disease being examined and on the medical context in which the test is placed, as we have discussed in a previous article in this series *(1)*.

We recommend that evaluations of diagnostic accuracy include a definition of the target condition (i.e., a clinically relevant disease), rather than just expressing the accuracy of diagnostic tests *(25)*. We recommend that this definition always be based on the best available evidence about the threshold or criteria for the intervention that the test will be used to guide. For a single test, test sensitivity thus represents the proportion of patients with the condition who will receive appropriate management, and test specificity represents the proportion of patients without the condition who will not undergo further unnecessary tests or treatment.

Not all test evaluations will be helped by considering diagnostic accuracy. With an aging population, increased longevity, and improved detection, chronicity is on the rise, and management of chronic disease has overtaken the care of acute disease. Tests and markers are often proposed for purposes other than diagnosis, such as the selection of treatment, monitoring treatment, dose adjustment, detecting side effects, prognosis, surveillance, screening, and others. It is not always clear how to define diagnostic accuracy for these other, nondiagnostic purposes of testing, and unconditional fixation on definitions of clinical sensitivity and specificity may prevent more informative evaluations of the clinical utility of novel markers and other medical tests.

Concluding Remarks

In this article we have discussed clinical utility without explicitly addressing costs. When making decisions and considering recommendations about diagnostic tests, clinicians and other decision makers will also have to consider what resources are needed to produce the desired improvements in patient outcomes. Studies about cost-effectiveness will then necessarily have to follow or accompany studies of clinical utility *(22, 23)*.

We are aware that in daily clinical practice tests are ordered in the absence of any documented net clinical utility. A test may be ordered to please the patient, for purely financial reasons, or because of liability concerns (an instance of defensive medicine). Tests may be performed and reacted to even if the profession is aware that they cannot be interpreted (for lack of proper research data).

These days, laboratory medicine cannot escape the ongoing paradigm shift, in which evidence that diagnostic testing improves patient outcomes is becoming a requirement before the cost of a test can be reimbursed and the test can be used in practice. In the future, such evidence may become necessary for new tests to receive approval for marketing. Innovators, developers, and manufacturers of new technology will have to identify the intended use of new markers and produce the evidence of their impact on healthcare outcomes *(26)*. These are interesting times.

References

1. Linnet K, Bossuyt PM, Moons KG, Reitsma JB. Quantifying the accuracy of a diagnostic test or marker. Clin Chem 2012;58:1292–301.
2. Reitsma JB, Moons KG, Bossuyt PM, Linnet K. Systematic reviews of studies quantifying the accuracy of diagnostic tests and markers. Clin Chem 2012;58:1534–45.
3. 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.
4. Neumann PJ, Tunis SR. Medicare and medical technology—the growing demand for relevant outcomes. N Engl J Med 2010;362:377–9.
5. Lijmer JG, Leeflang M, Bossuyt PM. Proposals for a phased evaluation of medical tests. Med Decis Making 2009;29:E13–21.
6. Gazelle GS, Kessler L, Lee DW, McGinn T, Menzin J, Neumann PJ, et al. A framework for assessing the value of diagnostic imaging in the era of comparative effectiveness research. Radiology 2011;261:692–8.
7. Bossuyt PM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making 2009;29:E30–8.
8. Marteau TM, Weinman J. Self-regulation and the behavioural response to DNA risk information: a

theoretical analysis and framework for future research. Soc Sci Med 2006;62:1360–8.

9. Heitman SJ, Hilsden RJ, Au F, Dowden S, Manns BJ. Colorectal cancer screening for average-risk North Americans: an economic evaluation. PLoS Med 2010;7:e1000370.

10. 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;144:850–5.

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

12. Biesheuvel CJ, Grobbee DE, Moons KG. Distraction from randomization in diagnostic research. Ann Epidemiol 2006;16:540–4.

13. Hunink MG, Krestin GP. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. Radiology 2002;222:604–14.

14. Kelloff GJ, Sigman CC. Cancer biomarkers: selecting the right drug for the right patient. Nat Rev Drug Discov 2012;11:201–14.

15. Wagner JA, Williams SA, Webster CJ. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. Clin Pharmacol Ther 2007;81:104–7.

16. Foster MW, Mulvihill JJ, Sharp RR. Evaluating the utility of personal genomic information. Genet Med 2009;11:570–4.

17. Grosse SD, Khoury MJ. What is the clinical utility of genetic testing? Genet Med 2006;8:448–50.

18. Pletcher MJ, Pignone M. Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. Circulation 2011;123:1116–24.

19. Wizemann T, Berger AC, eds., Institute of Medicine (US) Roundtable on Translating Genomic-Based Research for Health. The value of genetic and genomic technologies: workshop summary. Washington, DC: National Academies of Press/Institute of Medicine; 2010.

20. Mueller C, Scholer A, Laule-Kilian K, Martina B, Schindler C, Buser P, et al. Use of B-type natriuretic peptide in the evaluation and management of acute dyspnea. N Engl J Med 2004;350:647–54.

21. Breidthardt T, Laule K, Strohmeyer AH, Schindler C, Meier S, Fischer M, et al. Medical and economic long-term effects of B-type natriuretic peptide testing in patients with acute dyspnea. Clin Chem 2007;53:1415–22.

22. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. Circulation 2009;119:2408–16.

23. Moons KG. Criteria for scientific evaluation of novel markers: a perspective. Clin Chem 2010;56:537–41.

24. Clarke LD, Plevritis SK, Boer R, Cronin KA, Feuer EJ. A comparative review of CISNET breast models used to analyze U.S. breast cancer incidence and mortality trends. J Natl Cancer Inst Monogr 2006;:96–105.

25. Lord SJ, Staub LP, Bossuyt PM, Irwig LM. Target practice: choosing target conditions for test accuracy studies that are relevant to clinical practice. BMJ 2011;343:d4684.

26. Price CP, Christenson RH. Evaluating new diagnostic technologies: perspectives in the UK and US. Clin Chem 2008;54:1421–3.