Perspective

Designing Radiology Outcomes Studies—Essential Principles

Stella K. Kang, MD, MSc, Alvin I. Mushlin, MD, ScM

Health outcomes research is essential to align radiology with current standards of high-value patient care, through the assessment of end results of diagnostic tests, interventions, or policy on patient health. To bridge studies of diagnostic test accuracy and health outcomes research, key considerations include: (1) how to determine when a diagnostic test merits evaluation of impact on outcomes, (2) when study of intermediate/surrogate outcomes can be useful, (3) how to consider the possible harms as well as potential benefits of a test, and (4) how to integrate evidence of an imaging test's efficacy/effectiveness with clinical data to assess outcomes. Due to challenges in conducting studies of long-term outcomes consequent to imaging use, intermediate health outcomes may capture a test's impact on successful diagnosis and therapy, and can provide readily measurable, incremental insights into the role of imaging in health-care delivery and efficiency. In an era marked by recognition of quality and value of care, outcomes research will provide essential evidence to inform radiologists' guidance of imaging use toward improved patient care, creation of clinical guidelines, and policy decisions.

Key Words: outcomes research; diagnostic test performance; evidence synthesis; receiver operating characteristic curve.

INTRODUCTION

The development and dissemination of new and improved diagnostic imaging tests have thrived over the past several decades. Disease detection and characterization for medical decision-making increasingly depend upon imaging tests. As in other medical fields, the accumulation of published literature in radiology has become the basis of evidence-based practice or evidence-based radiology. The rapid evolution of imaging technology places unique demands upon this research effort, for re-evaluation of reproducibility, diagnostic test accuracy, and comparative performance against existing imaging tests—in the setting of varying patient populations and techniques for each imaging test. Additionally, the current era of health-care policy, with its emphasis on not only efficacy and effectiveness, but also value, requires evidence about how imaging technology translates into qualitative and quantitative changes in health outcomes and efficiency of clinical care.

Value can be simply defined as the health outcomes achieved per dollar spent (1). Because of the concern that indiscriminate diagnostic imaging use contributes to high health-care costs and potentially lower quality of care, policies intended to encourage evidence-based use of imaging are being enacted. For example, the Centers for Medicare and Medicaid Services will require ordering physicians to consult appropriate use criteria for imaging tests such as those endorsed by national medical specialty societies, including the American College of Radiology, to avoid reimbursement deductions, and also requires documentation of shared decision-making with patients regarding the benefits and harms of lung cancer screening (2,3). In efforts to enhance quality and value in the health-care system, decision-makers from day-to-day practitioners to expert panels and policy-making bodies are evaluating evidence on the impact of diagnostic tests on health outcomes for patients and health-care systems.

The purpose of this article is to provide an overview of some of the methodological issues and approaches that bridge diagnostic test accuracy and health outcomes research: (1) how to determine when a diagnostic test merits further evaluation, (2) when study of surrogate outcomes can be useful, (3) how to consider the possible harms as well as potential benefits of a test, and (4) how to integrate evidence of an imaging test's efficacy/effectiveness with clinical data to assess outcomes and value.

WHEN DOES A DIAGNOSTIC TEST MERIT FURTHER EVALUATION THROUGH OUTCOMES RESEARCH?

The assessment of imaging tests encompasses several stages, and may be described as a hierarchy of study types, all

Acad Radiol 2016; ■:■■–■■

Department of Radiology, NYU School of Medicine, 550 First Ave, New York, NY 10016Department of Population Health, NYU School of Medicine, New York, New YorkDepartment of Healthcare Policy and Research, Weill Cornell Medical College, New York, New York Received December 16, 2015; revised March 1, 2016; accepted March 6, 2016. Research Support: Stella Kang (NYU Medical Center) and/or this investigation was supported (in part) by an Association of University Radiologists—GE Radiology Research Academic Fellowship Award. **Address correspondence to:** S.K.K. e-mail: stella.kang@nyumc.org

http://dx.doi.org/10.1016/j.acra.2016.03.001

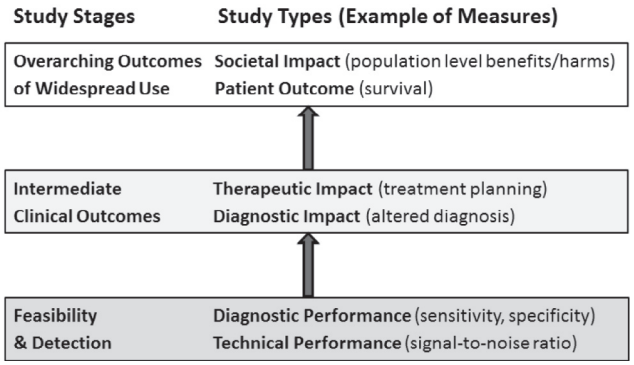

Figure 1. Progression of evidence and study types contributing to outcomes research.

providing evidence in support of, and ideally before, a test's widespread clinical adoption. At the top, patient outcomes (eg life expectancy) and societal impact (eg cost-effectiveness) capture the overarching effect of a diagnostic test on patients and align its use with the ideals of improved population health and sometimes, enhanced value over existing alternatives. In descending order, the study types in the hierarchy include therapeutic and diagnostic impact, diagnostic performance, and technical performance (Fig 1) (4–6).

Following a logical sequence of steps to evaluate a diagnostic test is critical. Initial studies should establish the technical feasibility and reliability of an imaging technique for diagnostic testing. The new diagnostic test may play a potentially valuable role when: (1) studies suggest clinically significant improvement in diagnostic test accuracy, (2) the test offers similar diagnostic performance to existing options at less cost, or (3) the test is more widely available or applicable/usable compared to the current standard for disease detection or characterization. To determine whether the test truly warrants further evaluation, decision analysis can be used to assess its potential impact on decision-making and outcomes and to help target the needed information for future research (7).

Decision-analytic modeling can provide a challenge receiver operating characteristic (ROC) curve, which depicts the performance characteristics of the new test in ROC space that would be required to contribute greater cost-effectiveness compared to the current diagnostic test. The challenge ROC curve is composed of pairs of true-positive and false-positive ratios that form a threshold for cost-effectiveness, and the space above and to the left of this curve represents the performance characteristics of the new test that would make it cost-effective compared to the standard test (7). If the purported accuracy meets the threshold demonstrating potential added value by using the new test versus the current standard, these findings support more complete assessment of its diagnostic test accuracy. The assessment would then serve to document the performance characteristics of the test as represented by its sensitivity, specificity, and ROC curve. The optimal sensitivity and specificity along the ROC curve of a test may also be probed using decision-analytic modeling. Specifically, the point on the curve that maximizes benefits (eg effectiveness or cost-effectiveness) may be determined by weighing the outcomes associated with true and false results (8).

Decision-analytic methods can also be used to quantify the potential monetary or net health benefits once test accuracy has been established. In addition, a value of information analysis can guide understanding of the overall value of research to evaluate the test, by estimating the expected benefit of obtaining this information. For example, a value of information analysis on colorectal cancer screening identified a large monetary net benefit to society in determining the optimal screening method among several options, including optical colonoscopy and computed tomography (CT) colonography (9). This analysis also demonstrated that the decision regarding the optimal test was largely impacted by adherence rates to screening recommendations, which were higher with less invasive testing, and by the rate of carcinoma development in colonic polyps, supporting further research in these areas (9).

After a test's accuracy has been estimated, its actual impact on clinical decision-making and patient care should be assessed in terms of patient clinical outcomes. This can be done via several study designs ranging from observational methods to randomized trials targeting either long-term outcomes or intermediate consequences such as diagnosis and treatment planning, or surrogate markers of disease progression (10). An understanding of the clinicians' preferred markers for disease status, and patients' preferences and quality of life, guides the selection of particularly relevant outcomes to measure in such studies.

WHEN IS STUDY OF INTERMEDIATE OR SURROGATE OUTCOMES USEFUL?

Progressing from studies of feasibility, diagnostic test accuracy, and its potential impact on decision-making to population-level health benefits and harms, and overall value to the healthcare system, is dependent upon a number of clinical factors, some of which are not always well known. Therefore, there may be sizable challenges to understanding the long-term effects of imaging. Even after the accurate detection of a particular disease, patients are subject to variable clinical courses given coexisting medical comorbidities, treatment choices or timing, or individual differences in disease genotypes and phenotypes. Surrogate markers or intermediate outcomes may provide more direct and readily measured information regarding the potential impact of a new test, of a new test compared to an existing test, or of test use in different patient populations.

The major limitation of a surrogate outcome study is that the selected marker may in fact not be directly linked to the clinical effect of interest (11). However, given the challenges of designing and implementing studies of imaging tests' impact on survival or net costs, selection of well-established