Original Investigation

Using the Analytic Hierarchy Process for Prioritizing Imaging Tests in Diagnosis of Suspected Appendicitis

Maria Agapova, PhD, Brian W. Bresnahan, PhD, Ken F. Linnau, MD, Louis P. Garrison Jr., PhD, Mitchell Higashi, PhD, Larry Kessler, ScD, Beth Devine, PhD, PharmD, MBA

Rationale and Objectives: In clinical guideline or criteria development processes, such as those used in developing American College of Radiology Appropriateness Criteria (ACR AC), experts subjectively evaluate benefits and risks associated with imaging tests and make complex decisions about imaging recommendations. The analytic hierarchy process (AHP) decomposes complex decisions into structured smaller decisions, incorporates quantitative evidence and qualitative expert opinion, and promotes structured consensus building. AHP may supplement and/or improve the transparency of expert opinion contributions to developing guidelines or criteria.

Materials and Methods: To conduct an empirical test using health services research tools, we convened a mock ACR AC panel of emergency department radiology and nonradiology physicians to evaluate by multicriteria decision analysis, the relative appropriateness of imaging tests for diagnosing suspected appendicitis. Panel members selected benefit-risk criteria via an online survey and assessed contrast-enhanced computed tomography, magnetic resonance imaging, and ultrasound using an AHP-based software. Participants were asked whether the process was manageable, transparent, and improved shared understanding. Priority scores were converted to rankings and compared to the rank order of ACR AC ratings.

Results: When compared to magnetic resonance and ultrasound imaging, participants agreed with the ACR AC that contrast-enhanced computed tomography is the most appropriate test. Contrary to the ACR AC ratings, study results suggest that magnetic resonance is preferable to ultrasound. When compared to nonradiologists, radiologists' priority scores reflect a stronger preference for computed tomography.

Conclusions: Study participants addressed decision-making challenges using a relatively efficient data collection mechanism, suggesting that AHP may benefit the ACR AC guideline development process in identifying the relative appropriateness of imaging tests. With additional development, AHP may improve transparency when expert opinion is used in clinical guideline or appropriateness criteria development.

Key Words: Diagnostic imaging; appropriateness criteria; analytic hierarchy process; appendicitis

INTRODUCTION

The American College of Radiology (ACR) publishes evidence-based and opinion-based criteria outlining appropriate uses of imaging tests (1). ACR Appropriateness Criteria (ACR AC) are developed and revised every 3 years by panels composed of 10–16 volunteer ACR members (2,3). Panel members rate imaging tests on a scale that ranges from 1 to 9 (1–3, inappropriate; 4–6, equivocal; 7–9, appropriate) using the RAND/UCLA Appropriateness Method (4). The ACR AC quantitative ratings represent qualitative reconciliation of benefits and risks into one measure that supports the ultimate decision (or recommendation) of the radiologist expert panel. For a given clinical scenario, referred to as a topic variant, panel members assess the risks of each test against the benefits of performing the procedure. Expert opinion is used to address evidence gaps and supplements existing evidence (3).

Decision scientists have shown that individuals struggle with complex decisions involving multiple objectives with uncertain trade-offs (5). As the number of alternatives and criteria judgments increases, individuals' decision-making capabilities degrade (6). In this context, ACR AC expert panel members may face several challenges: (1) rating multiple imaging alternatives for any given clinical indication; (2) assignment of importance to multiple potential benefits and risks; (3) as volunteers, working with limited financial resources and time; (4) decision making in an environment of high uncertainty with regard to benefits and risks across alternatives; and (5) given divergent views, arriving at a single metric representing the benefit-risk balance, or appropriateness. While the modified Delphi consensus approach of the RAND/UCLA Appropriateness Method assists panel members with reaching consensus, support for decomposing complex decisions, individually or in groups, is lacking.

Acad Radiol 2017; ■:■■–■■

From the Pharmaceutical Outcomes Research and Policy Program, Department of Pharmacy, School of Pharmacy, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195 (M.A., B.W.B., L.P.G., B.D.); Department of Radiology, School of Medicine, University of Washington, 325 Ninth Ave, Seattle, WA 98104 (B.W.B., K.F.L.); Department of Health Services, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195 (B.W.B., L.K.); General Electric Healthcare, Waukesha, Wisconsin (M.H.). Received March 24, 2016; accepted January 5, 2017. **Address correspondence to:** B.W.B. e-mail: bres@u.washington.edu

http://dx.doi.org/10.1016/j.acra.2017.01.001

Multicriteria decision analysis (MCDA) methods are one of many benefit-risk assessment approaches (7,8). MCDA methods are particularly useful in organizing and weighting multiple, often conflicting, criteria (e.g., intervention effects) (9). MCDA is applied broadly in healthcare research, with underpinnings in operations research and decision theory (10). There are several reasons for choosing MCDA to facilitate ACR AC development. First, MCDA is well suited to meet the need in diagnostic imaging to merge quantitative and qualitative evidence and expert opinion. Second, MCDA assists participants in articulating values, applying the values rationally, and documenting the results across alternative strategies. Third, the steps in an MCDA are similar to steps taken by ACR AC panels (Fig 1). The added structure of MCDA methods promises to improve the efficiency of meetings to the extent that the marginal time the MCDA process consumes becomes negligible. Fourth, if decision makers are unsure about their judgments, varying decision makers' inputs to determine whether results are robust across a plausible range of values for one or more inputs can be assessed in sensitivity analyses. Fifth, when a lack of consensus requires additional Delphi consensus rounds, MCDA may offer an alternative approach to decomposing the decision problem and building consensus piecemeal.

While, in theory, MCDA methods appear well suited for the problems faced by ACR AC, the feasibility of using MCDA in ACR AC guideline development has not been previously explored or tested. Using the information gathered for ACR AC deliberations, we sought to explore whether MCDA analyses would yield, without a cost to transparency or efficiency, similar conclusions using an existing ACR AC use case: the diagnosis of lower quadrant pain, suspected appendicitis. Among computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound, it is unclear which modality has the most favorable benefit-risk profile. In the United States, CT is considered the gold standard for diagnosis of the classical presentation of suspected appendicitis in adults (11). In Europe, standard practice is to use ultrasound first (12). Further, some propose MRI as an alternative to CT when ultrasound findings are equivocal (13). The objective of this study was to use the analytic hierarchy process (AHP) approach to assess the relative appropriateness of these diagnostic imaging tests in diagnosing suspected appendicitis.

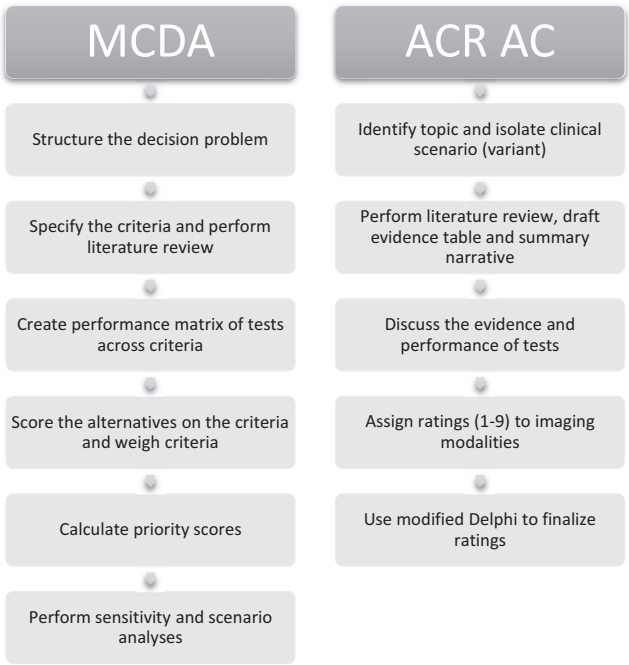

Figure 1. Steps of the MCDA and the ACR AC processes. ACR AC, American College of Radiology Appropriateness Criteria; MCDA, multicriteria decision analysis.

METHODS

Selection and Definition of Clinical Use Case

We used the following ACR AC case definition for appendicitis, last updated in 2013: Lower quadrant pain-suspected appendicitis (NGC-10146) variant 1: A patient arrives complaining of lower quadrant pain. Fever, leukocytosis, and other signs point to a classic case of clinical appendicitis. The 10 diagnostic modalities assessed by the ACR AC panel are listed on the ACR AC website (14). We abstracted information from the following ACR AC supporting documentation: (1) ACR AC narrative written by a panel member serving as the topic author describing the evidence base used in the decision-making process (Appendix File S1); and (2) the evidence table of the studies cited in the narrative portion of the ACR AC including details of study design, summary of study results, and an evaluation of study quality (15).

Selection of MCDA Method

Two methods within MCDA—the AHP and Multiple Attribute Utility (Value) Theory Analysis—are designed to facilitate group decision making (16). We chose AHP because pairwise comparisons are intuitive and cognitively less burdensome than using direct elicitation. Scoring imaging tests in relative terms, a feature of AHP, also obviates the need to assign each imaging test a measure of performance: this is a strong advantage when limited evidence is supplemented with expert opinion.

Mock ACR AC Panel Participant Recruitment

The recruitment population pool comprised radiologists and nonradiologists with a clinical specialty in emergency medicine who had participated in a previously conducted, related study, and who had expressed interest in participating in the mock ACR AC panel activity. Participants received an invitation to participate via email. This study received approval from The University of Washington's Institutional Review Board.