Appropriate Use and Clinical Impact of Transthoracic Echocardiography

Susan A. Matulevicius, MD; Anand Rohatgi, MD, MSCS; Sandeep R. Das, MD, MPH; Angela L. Price, MD; Andres deLuna, MD; Sharon C. Reimold, MD

**IMPORTANCE** Transthoracic echocardiography (TTE) accounts for almost half of all cardiac imaging services and is a widely available and versatile tool. Appropriate use criteria (AUC) for echocardiography were developed to improve patient care and health outcomes. Prior studies have shown that most TTEs are appropriate by AUC. However, the associations among TTE, AUC, and their clinical impact have not been well explored.

**OBJECTIVES** To describe the proportion of TTEs that affect clinical care in an academic medical center overall and in subgroups defined as appropriate and inappropriate by AUC.

**DESIGN AND SETTING** Retrospective review of medical records from 535 consecutive TTEs at an academic medical center was performed. The TTEs were classified according to 2011 AUC by 2 cardiologists blinded to clinical impact and were assessed for clinical impact by 2 cardiologists blinded to AUC. Clinical impact was assigned to 1 of the following 3 categories: (1) active change in care, (2) continuation of current care, or (3) no change in care.

**PARTICIPANTS** Five hundred thirty-five patients undergoing TTE.

**EXPOSURE** Transthoracic echocardiography.

**MAIN OUTCOMES AND MEASURES** Prevalence of appropriate, inappropriate, and uncertain TTEs and prevalence of clinical impact subcategories.

**RESULTS** Overall, 31.8% of TTEs resulted in an active change in care; 46.9%, continuation of current care; and 21.3%, no change in care. By 2011 AUC, 91.8% of TTEs were appropriate; 4.3%, inappropriate; and 3.9%, uncertain. We detected no statistically significant difference between appropriate and inappropriate TTEs in the proportion of TTEs that led to active change in care (32.2% vs 21.7%; \(P = .29\)).

**CONCLUSIONS AND RELEVANCE** Although 9 in 10 TTEs were appropriate by 2011 AUC, fewer than 1 in 3 TTEs resulted in an active change in care, nearly half resulted in continuation of current care, and slightly more than 1 in 5 resulted in no change in care. The low rate of active change in care (31.8%) among TTEs mostly classified as appropriate (91.8%) highlights the need for a better method to optimize TTE utilization to use limited health care resources efficiently while providing high-quality care.
The widespread availability, negligible risk, and versatility of transthoracic echocardiography (TTE) make it a powerful and appealing diagnostic tool. However, these same characteristics may encourage overuse. In fact, the use of TTE has doubled during the past decade, constituting approximately half of all cardiac imaging services among Medicare beneficiaries. Transthoracic echocardiography accounted for 11% and more than $1.1 billion of total Medicare diagnostic imaging spending in 2010. Moreover, more than half of Medicare beneficiaries who undergo a TTE will have a second TTE within 3 years.

To respond to the dramatic increase in the use of diagnostic imaging during the past decade, the American College of Cardiology Foundation collaborated with the American Society of Echocardiography and other imaging subspecialty societies to develop appropriate use criteria (AUC) for TTE. The AUC were initially published in 2007 and later updated in 2011 to “respond to the need for the rational use of imaging services in the delivery of high-quality care” and potentially “impact physician decision making,” with an ultimate objective to “improve patient care and health outcomes.” Most TTEs performed in a wide variety of clinical settings have been judged appropriate based on AUC. Less is known, however, about the clinical impact of TTE and its relationship with appropriateness. To further elucidate this important issue, we examined the proportion of TTEs that affect clinical care in an academic medical center overall and in subgroups of TTEs defined as appropriate and inappropriate by AUC.

Methods

Study Population
All TTEs ordered from April 1 through April 30, 2011, at The University of Texas Southwestern Medical Center were retrospectively reviewed in regard to clinical impact and appropriateness. A TTE was excluded from review if it was not performed. A TTE was excluded from analysis if (1) no clinical data were available to assign an AUC or the clinical impact or (2) the TTE was performed after cardiac transplant or implantation of a left ventricular assist device (Figure 1). This study was approved by The University of Texas Southwestern Medical Center institutional review board, and a waiver of informed consent was obtained.

AUC Adjudication
Two independent general cardiologists (A.L.P. and A.dL.) reviewed all information in the electronic medical record (EMR) dated before the TTE was performed and classified each TTE according to the 2011 AUC. These reviewers were blinded to the results of the TTE and to the clinical course subsequent to the performance of TTE. The AUC were then classified as appropriate, inappropriate, or uncertain as defined by the 2011 AUC. A median score of 7 to 9 indicates that an AUC is appropriate (generally acceptable and reasonable to perform); a median score of 4 to 6, an AUC is uncertain (may be generally acceptable and reasonable to perform or significant disagreement ensued in scoring of the indication); and a median score of 1 to 3, an AUC is inappropriate (not generally acceptable or reasonable to perform) based on the available evidence. Disagreements in TTE AUC classification were resolved by consensus. Cases in which consensus could not be reached underwent definitive adjudication by a third cardiologist (S.R.D.) blinded as above.

Adjudication of Clinical Impact
Two noninvasive cardiologists (S.A.M. and A.R.) blinded to AUC classification independently assessed the clinical impact of each TTE through retrospective review of the complete EMR. Clinical impact was empirically defined a priori in one of the following 3 mutually exclusive categories: (1) active change in care, (2) continuation of current care, or (3) no change in care.
(Table 1). Reviewing cardiologists were instructed to select active change if any change occurred in response to the TTE. In exploratory analysis, TTEs that led to active change were rated on a scale of 1 to 5 by consensus as very useful (5), useful (4), neutral (3), not useful (2), or misused for affecting patient care (1) (described in the Supplement [eTable]). The EMR was reviewed through hospital discharge for all inpatients and through the next outpatient visit to the ordering provider for all outpatients. Any disagreements in clinical impact were resolved by consensus. If consensus could not be reached, a third blinded noninvasive cardiologist (S.C.R.) reviewed the cases and made a final assignment of clinical impact.

### Statistical Analysis

Agreement between reviewers in appropriateness grading and clinical impact assignment after attempted consensus was assessed with weighted κ analysis. The proportions of each of the 3 clinical impact categories were calculated for the overall cohort and by AUC grade. Differences in frequency distributions were statistically compared using the χ² test, and differences in continuous variables were compared using analysis of variance. All reported P values are 2-tailed, with P < .05 considered statistically significant. All statistical analyses were performed with commercially available software (SAS, version 9.2; SAS Institute Inc).

### Results

A total of 617 TTEs were reviewed and 535 were included in this study (Figure 1). The study population was 58.7% female, 55.3% white, 21.1% African American, and 8.2% Hispanic, with a mean (SD) age of 58 (17) years. Inpatient TTEs con-
Abbreviations: A, appropriate; AUC, appropriate use criteria; TTE, transthoracic echocardiography.

Although most TTEs (91.8%) were appropriate by AUC, less than one-third of TTEs performed at a US academic medical center constituted 57.0% of all TTEs. The specialties of general internal medicine (38.5%) and cardiology (31.2%) ordered the most TTEs. Of 617 TTEs reviewed, arbitration for AUC assignment by a third cardiologist was required for 24 (3.9% [weighted κ, 0.8]; \( P < .001 \)). Arbitration for clinical impact assignment by a third cardiologist was required for 28 TTEs (4.5% [weighted κ, 0.8]; \( P < .001 \)).

**Appropriateness**

Based on 2011 AUC, 91.8% of TTEs were classified as appropriate, 4.3% as inappropriate, and 3.9% as uncertain. Age, sex, specialty of the ordering provider, and TTE indication were similar between appropriate and inappropriate TTEs (Table 2). Slightly fewer outpatient compared with inpatient TTEs were appropriate (86.5% vs 95.7% [\( P < .001 \)]), although most outpatient and inpatient TTEs were appropriate.

The 10 most frequent AUCs, all classified as appropriate, accounted for 66.5% of TTEs resulting in an active change in care, 69.3% of TTEs resulting in no change in care, and 66.5% of TTEs resulting in continuation of care (Table 3). Clinical impact varied among these appropriate AUCs, with some AUCs being associated with active change more than 50% of the time (AUC 18 and 52), whereas others were associated with active change less than 25% of the time (AUC 1, 34, and 91). Of the 10 AUC most frequently associated with no change in care (72.8% of all TTEs resulting in no change in care), 9 were classified as appropriate and 1 was classified as uncertain (data not shown).

**Clinical Impact**

In the overall cohort, 31.8% of TTEs resulted in an active change in care, 46.9% led to continuation of current care, and 21.3% led to no change (Figure 2). Of the TTEs that led to continuation of current care, 43.0% included documented communication to the patient or physician about the results of the TTE. In an exploratory analysis of the relative impact of active change, only 18.9% of all TTEs were classified as 5 (very useful) or 4 (useful) and 6.0% were classified as 2 (not useful) or 1 (misused) in affecting patient care (Supplement eTable). Factors associated with no change compared with active change included older patients (\( P = .003 \)) and an inpatient study setting (\( P < .001 \)) (Table 2). Cardiology as the ordering specialty was associated with a lower proportion of studies resulting in no change in care (12.6%) compared with pulmonary/critical care and surgery (30.9%) and cardiology (31.5%) (\( P < .001 \)), respectively. We detected no statistically significant difference between appropriate and inappropriate TTEs that led to active change in care (32.2% vs 21.7% [\( P = .29 \)]) (Figure 2). The most common active changes were further diagnostic testing (29.4%) or subspecialty consultation (25.9%) (Table 4). In patients with a prior TTE (n = 226), 37.2% resulted in an active change in care, 40.7% resulted in continuation of care, and 22.1% resulted in no change. Similarly, in first-time TTEs (n = 309), 27.8% resulted in an active change in care, 51.5% resulted in continuation of care, and 20.7% resulted in no change. Among appropriate TTEs, we found no difference in active change resulting among TTEs rated 7, 8, or 9 by AUC (active change in 40.0%, 32.0%, and 32.0% [\( P = .79 \) for trend]).
led to an active change in clinical care, nearly half resulted in continuation of current care, and approximately one-fifth resulted in no change in care. The proportion of TTEs resulting in active change (31.8%) was markedly lower than the proportion of appropriate TTEs based on AUC (91.8%) and did not correlate with AUC classification. Similarly, the proportion of TTEs resulting in no change in care (21.3%) was markedly higher than the proportion of TTEs classified as inappropriate by AUC (4.3%), suggesting important limitations for AUC in optimizing the use of TTEs.

The AUC were developed in part to create more uniform TTE practice patterns. Our AUC grades of 91.8% appropriate, 4.3% inappropriate, and 3.9% uncertain are similar to those reported from other institutions and in a variety of clinical settings. In a US study of 384 inpatient and outpatient TTEs at a tertiary care academic medical center in New York, 92% were appropriate, 2% were inappropriate, 0.5% were uncertain, and 5.5% could not be classified; in an Italian study of 931 inpatient TTE referrals, 80% were appropriate, 15% were inappropriate, and 5% were uncertain. In a US study of 1080 inpatient TTEs in a regional community hospital, 97% were appropriate, 2% were inappropriate, and less than 1% were uncertain. In the largest study to date, performed in 1820 inpatients and outpatients from a single US Midwest academic center, 82% of TTEs were appropriate, 12% were inappropriate, 5% were uncertain, and 4% were unclassifiable.

If AUC had an effect on physician decision making, the rates of appropriate TTEs should increase and those of inappropriate TTEs should decrease after the publication of AUC. To the contrary, data from a major US academic medical center demonstrated no change in the proportion of TTEs classified as appropriate before AUC publication in 2000 and after AUC publication in 2007 (87% vs 85% [P = .58]). During this same time period, however, TTE volume increased by 85%. Similarly, inappropriate AUCs should be associated with a lower prevalence of an active change in care and a higher prevalence of no change than appropriate TTEs; however in our study, we found no difference in the prevalence of an active change or no change in care for appropriate vs inappropriate TTEs (Figure 2). These data suggest that AUC for TTE have not fulfilled some of its anticipated results, to have “a significant impact on physician decision-making,” and they did not curb the growth of TTE use since publication.

The intrinsic physical risks of TTE are negligible, and any incremental information can be seen as a benefit; therefore, the bar for accepting a TTE as appropriate would be very low unless clinical impact is also considered. Acknowledging this, the AUC working group stated that the definition of an imaging test’s appropriateness must include an explicit understanding of how the test results might lead to care that could improve the patient’s chances for survival or improved health status. However, no large studies of the association between AUC and clinical impact of TTEs in the United States have been performed, and no standard definition within the imaging community exists for assessing clinical impact. The single published report, based on a small cohort (n = 170) of only outpatient TTEs, showed an overall rate of clinical impact qualitatively similar to our study (39%). Also, when appropriate vs
Inappropriate TTEs were compared, no statistically significant difference in active change in management was found, similar to our observations. The sole large study\(^1\) (n = 917) to address this question was performed in Italy and had quite different results. Clinical impact was similarly defined as a change in diagnostic management, therapeutic decision, or follow-up planning owing to the TTE result. In that study,\(^2\) overall clinical impact (76%) was significantly higher than in the previously reported American study or our study, suggesting that potential differences between US and Italian health care practice patterns, rates of referral to TTE, and reimbursement may lead to differences in TTE use and impact of that use. In the Italian study, appropriate or uncertain TTEs were significantly more likely to result in changes in clinical care compared with inappropriate TTEs (87% vs 14% \(P < .001\)).

The potential for differences in accessibility and reimbursement affecting TTE use can also be demonstrated within the US system by comparing the use of TTEs between Medicare and the US Veterans Affairs health care system. From 1999 through 2008, use of TTE increased by 90% among Medicare beneficiaries compared with a 4% increase within the Veterans Affairs health care system during the same period (2000-2007) (Figure 3).\(^3\) The increase in use of TTE in the Medicare population may result from significant variations in physician testing thresholds, ease of access to TTEs, patient demographics, and potential differences in financial incentives associated with increased diagnostic testing.\(^4,5\) Repeated testing or lack of access to prior TTEs performed at another site may contribute to the differences seen as well because TTE reports in the Veterans Affairs population are accessible through the EMR. However, in our cohort, we found no significant difference in clinical impact between those patients with prior TTEs and those with first-time TTEs.

Several reasons explain why the AUC guidelines may not improve rational use of imaging services. The method developed by the RAND Corporation and the University of California, Los Angeles (RAND/UCLA method)\(^1\) requires that a working group review the literature and develop a list of clinical indications to be rated. A standardized literature review is then performed for each indication, and evidence tables are formed when significant evidence is available for a specific indication, recognizing that many imaging studies are observational cohort studies and may have inherent bias and that many indications may have no available evidence. These indications are then presented to an expert panel, who rate the appropriateness of each indication individually and then again after discussion with the entire panel on a scale of 1 to 9 based on the availability of published evidence, American College of Cardiology Foundation/American Heart Association clinical practice guideline recommendations, and, when published evidence or guideline recommendations are lacking, clinical experience. Ratings are tabulated and a final appropriateness score is assigned. Each indication is classified into one of the following 3 AUC categories: appropriate, inappropriate, or uncertain.\(^6\) Because the writing group and technical panel mainly consisted of physicians who specialize in imaging, including many experts in echocardiography, the consensus ratings will likely represent current clinical thinking of imagers and less likely challenge current TTE ordering practices and therefore allow for liberal use of echocardiography. Another difficulty is applying the RAND/UCLA method to a diagnostic testing modality, such as TTE instead of a therapeutic modality for which the RAND process was developed. A diagnostic test, unlike a therapeutic intervention, may be appropriate in that it can detect a certain disease process but not necessary because the clinical data without the test results may provide enough of the needed information to deliver high-quality care. Other processes of consensus decision making have tried to incorporate a consensus opinion of “necessary” into the categorization of AUC and may need to be considered to improve AUC for TTE.\(^7\)

Patient selection is a central aspect of quality in cardiac imaging.\(^7\) For cardiac imaging to be used most effectively, the test must be applied to the proper patient subset and at the optimal time, and the results of the test must be actionable. Recent data suggest that many Medicare beneficiaries undergo routine annual TTEs without proven evidence of benefit.\(^8\) In our study, 114 TTEs in 1 month led to no change in care, which equates to more than 1300 TTEs on an annual basis. If our findings are corroborated in other settings and centers, 21% (or $230 million) of the $1.1 billion of Medicare expenditure on echocardiography could have been saved if these TTEs had not been performed. Better metrics for identifying patients or scenarios when TTE is likely to result in no change in care must be developed. One striking finding in our study is the high proportion of TTEs that resulted in continuation of current care. In our study, 108 TTEs documented direct patient or physician communication about the TTE result, an important patient-centric use of TTE. Studies evaluating the role of patient communication and reassurance in patient satisfaction and medical resource utilization have been conflicting\(^19-23\); however, future research on understanding the role of diagnostic testing in patient-centered care may be a source for improved clinical impact and utility of echocardiography. Similarly, little research has been performed on the role of guideline-based TTE recommendations in patients already receiving evidence-based medicine for heart failure after myocardial infarction, another potential area for assessing necessity of testing in these subgroups rather than focusing on appropriateness.

We observed significant heterogeneity in clinical impact among common appropriate TTE indications. If validated in other studies, these findings suggest that policies targeting common indications that are currently judged appropriate but have minimal impact may help to curb TTE overuse. For example, initial evaluation of reasonably suspected valvular or structural heart disease (AUC 34) and serial reevaluations in a patient undergoing therapy with cardioactive agents (AUC 91) were among the most common appropriate TTE indications but resulted in an active change in care in fewer than 15% of studies. Alternative strategies, including performance of limited echocardiography or screening with plasma biomarkers, such as sensitive troponin and natriuretic peptide levels, may help to improve efficiency of TTE screening for these indications.

Our study has several limitations. This study is retrospective and relies on EMR review, which may have led to misclassification of impact owing to incomplete documentation. Ac-
Conclusions

Although almost all TTEs were appropriate by the 2011 AUC, only 1 in 3 resulted in an active change in care and approximately 1 in 5 resulted in no change in care. The discrepancy between appropriateness and clinical impact is striking and suggests that the AUC as currently implemented are unlikely to facilitate optimal use of TTE. Given the importance of responsible use of limited medical resources and the need to control increasing health care costs, additional research into the necessity of TTE in the process of medical care is needed and will require collaboration among hospitals, administrators, politicians, economists, the government, and patients. Prior attempts to decrease reimbursement for TTE to contain its use have not been successful because, despite cuts in reimbursements, TTE volume has continued to increase. Methods to increase incentives for definitive and patient-centered evaluation and management service in which physician time for communication of the treatment plan and expectations of care is valued may decrease the overuse of diagnostic testing. Educational programs within medical school and post-medical school training about the cost and utility of testing in the daily care of patients and professional society endorsement and commitment to programs such as “Choose Wisely” may increase patient and physician awareness and increase our commitment to be stewards of our health care resources. Promising areas of inquiry include the potential impact of incorporating necessity into the appropriateness framework, the potential effect of differences in reimbursement or accessibility on the use of TTE overall, and on the relationship between TTE appropriateness and clinical impact. In addition, quantifying the effect of TTE-based reassurance on patient-centered outcomes and specifically examining whether alternate strategies that do not involve TTE may provide similar benefit to the patient at reduced resource cost may offer the greatest potential to decrease overuse while maintaining high-quality care.
Appropriate vs Clinically Useful Diagnostic Tests

John P. A. Ioannidis, MD, DSc

Deluged by a plethora of costly diagnostic tests, many stakeholders struggle to define criteria on which tests are appropriate to perform and in what settings. Appropriateness is often decided on the basis of expert opinion and some circumstantial evidence about diagnostic performance characteristics (eg, sensitivity and specificity). However, what really matters eventually concerns what happens to the patients undergoing the tests. What other tests, invasive procedures, or treatments were ordered or aborted based on the results? How were major clinical events, quality of life, or even survival (for serious conditions) affected? Did patients fare better because of the diagnostic drilling? Often, data to answer these questions are slim or nonexistent.

This evidence gap exists even for the most popular tests in clinical medicine. Matulevicius et al report the results of a very interesting study on appropriateness vs clinical usefulness for transthoracic echocardiography (TTE). Transthoracic echocardiography is the most popular cardiac imaging study; approximately 700 TTE studies are performed annually per 1000 Medicare beneficiaries. Given its popularity, costs can easily escalate unless some restriction is set on which TTEs are deemed appropriate to perform. Indeed, 10 highly respectable professional organizations and their experts have come up with a meticulously comprehensive list of appropriate uses of echocardiography, last updated in 2011. Overall, 202 indications of echocardiography were reviewed. Of those, 97 were scored as appropriate, 34 as uncertain, and 71 as inappropriate. A perusal of the 41-page document and its 145-page appendices reveals an outstanding effort, complete and transparent in methods, processes, and conclusions. One would hope that this highly visible and widely endorsed effort would ensure that the rate of inappropriate TTE is minimized. Moreover, one would hope that appropriate TTE would optimize patient management and outcomes.

Matulevicius et al found that, among 535 consecutive TTEs performed at an academic medical center, 91.8% were indeed...