The use of noninvasive cardiac testing has risen substantially during the past 2 decades, resulting in increased costs to healthcare systems and providers.\(^1\)\(^-\)\(^3\) It is estimated that between 5% and 25% of cardiac imaging tests are performed for inappropriate indications.\(^4\)\(^-\)\(^7\) Implementation of The American College of Cardiology Foundation appropriate use criteria (AUC) across several noninvasive cardiac imaging modalities has several potential benefits.\(^8\)\(^-\)\(^1\)\(^1\) At the provider level, adherence to AUC can potentially decrease health costs by reducing the number of inappropriate tests performed. In addition, some studies have demonstrated that the prognostic value of a test differs by AUC, and that the level of a test’s appropriateness may provide additional prognostic information above diagnostic characteristics alone, suggesting that these criteria have an equally important role to improve the clinical value of cardiac diagnostic testing.\(^1\)\(^2\)

**Methods**

**Data Sources and Search Strategy**

Using a systematic search strategy, we identified studies of quality improvement initiatives aimed at reducing inappropriate cardiac imaging. The primary outcome was the proportion of inappropriate tests based on appropriate use criteria. Studies were analyzed using a random effects meta-analysis model, and heterogeneity was examined using subgroup analyses. We identified 6 observational studies and 1 randomized control trial. Most interventions (n=6) had a formal education component, and 5 included a mechanism for physician audit and feedback. Although these interventions were associated with lower odds of inappropriate testing (odds ratio, 0.44 [95% confidence interval, 0.32–0.61]; \(P<0.001\)), significant heterogeneity was observed (\(I^2=70\%\)), which was best explained by the utilization of physician audit and feedback. Interventions that employed physician audit and feedback were associated with significantly lower odds of inappropriate testing (odds ratio, 0.36 [95% confidence interval, 0.31–0.41]; \(P<0.001\); \(I^2=0\%\)), whereas those that did not had no effect (odds ratio, 0.89 [95% confidence interval, 0.61–1.29]; \(P=0.51\); \(I^2=0\%\); \(P\) value for difference <0.001). All studies had potential sources of bias that could have affected the observed estimates.

**Conclusions**

Interventions using physician audit and feedback are associated with lower odds of inappropriate cardiac testing. Further research is needed to evaluate a greater diversity of intervention types, with improved study designs.


Key Words: cardiac imaging ■ meta-analysis ■ quality improvement ■ research

The impact that AUC will have on improving cardiac test use greatly depends on how this knowledge is translated into clinical practice. To facilitate its implementation, several novel quality improvement interventions have been developed to implement AUC with the ultimate goal of reducing unnecessary cardiac testing. These initiatives vary in type and scope, and there has been no comprehensive review that examines the quality and effectiveness of these interventions. This systematic review examines the current evidence supporting quality improvement initiatives that aim to improve noninvasive cardiac test use, and quantifies the impact of these interventions on reducing the rate of inappropriate tests based on AUC. In addition, we examined the impact of specific components within these initiatives to determine, which provided the most consistent benefits.

**Methods**

**Data Sources and Search Strategy**

Using a systematic search strategy, we identified studies of quality improvement initiatives involving at least 1 of 4 common
WHAT IS KNOWN

- Previous studies have found that between 5% and 25% of noninvasive cardiac imaging tests may be performed for inappropriate indications.
- Quality improvement interventions have been developed based on American College of Cardiology Foundation appropriate use criteria with the objective of reducing inappropriate cardiac testing.

WHAT THE STUDY ADDS

- In this meta-analysis of published studies, quality improvement interventions are associated with fewer inappropriate cardiac tests.
- Physician audit and feedback appears to be an important process impacting their overall effectiveness.
- Further experimental studies are needed to determine the optimal methods to reduce inappropriate cardiac imaging.

Cardiac imaging modalities: single photon emission computed tomographic myocardial perfusion imaging, echocardiography, cardiac magnetic resonance imaging, and cardiac computed tomographic angiography. We searched MEDLINE and EMBASE for relevant studies using the following combination of key words: (myocardial perfusion imaging or single photon emission computer tomography or echocardiography or nuclear magnetic resonance imaging or computer-assisted tomography or imaging or cardiac imaging or scintiscanning) and (guidelines adherence or appropriateness criteria or appropriate use or total quality management). The search was limited to studies in adults (>18 years of age), in the English language, and published after the year 2000. The search was initially conducted on February 15, 2014 and updated on January 15, 2015. We also reviewed bibliographies of the identified articles, used the PubMed-related articles search feature, and contacted experts in the field to identify additional studies that were not captured in our initial search.

Selection Criteria

Inclusion criteria for our systematic review were the following: (1) randomized control trials or observational studies, (2) the population underwent a cardiac test using either single photon emission computed tomographic myocardial perfusion imaging, echocardiography, cardiac magnetic resonance, or cardiac computed tomographic angiography, (3) the use of a quality improvement intervention aimed at reducing inappropriate tests or increasing appropriate tests, and (4) outcomes measured using AUC.

Data Extraction and Collection

Two independent reviewers (D.C and P.J.) evaluated each study independently and discrepancies were discussed among the reviewers and settled through common consensus.

Statistical Analysis

The primary analysis was a comparison of the proportion of inappropriate cardiac tests performed in groups with quality improvement interventions compared with those without interventions. The proportion of inappropriate tests performed in the intervention and control groups were calculated for each study. The DerSimonian–Laird random-effects model was used to calculate the inverse-variance–based weighted, pooled treatment effect across studies. By the methods described by Deeks and Higgins,13 we measured heterogeneity between studies using the χ² test (with a P<0.10 denoting significant heterogeneity), and the I² statistic (with a value >50% denoting significant heterogeneity).14 Heterogeneity was further examined using subgroup analysis, with studies separated based on the presence or the absence of distinct components of the interventions, specifically: the use of formal educational interventions (eg, didactic lectures and small group sessions), the presence of a physician audit and feedback mechanism, and the presence of a point of care decision support tool. Heterogeneity between subgroups was measured using the χ² test by the methods described by Deeks and Higgins.15 A sensitivity analysis was also performed comparing the results of the observational studies to the single randomized control trial. As part of the secondary analysis examining the impact of these interventions on appropriate testing, the proportions of appropriate tests in the intervention and control groups were also calculated for each study, and the pooled treatment effect was calculated using the DerSimonian–Laird random-effects model. All analyses were performed using the Cochrane Collaboration Review Manager (version 5.3, Cochrane Collaboration, Copenhagen, Denmark), and P<0.05 was considered the threshold for statistical significance.

Results

Study Selection

Our literature search resulted in 2,254 articles (1,484 from EMBASE and 770 from MEDLINE), out of which 2,022 non-duplicate abstracts were retrieved and reviewed for inclusion (a summary of study selection process is outlined in Figure 1). Title and abstract review by P.J. and D.C. demonstrated satisfactory agreement (κ score=0.66) and yielded 12 studies for full text review. Of these 12 studies, 5 were excluded from the final analysis: 3 did not have appropriate/inappropriate...
tests as an outcome measure, 1 was an extension of another study included in the analysis, and 1 did not have adequate patient level information for meta-analysis. After full text review, 7 studies were included in the systematic review and meta-analysis.

**Study Characteristics**

Six studies were observational and 1 study was a randomized control trial. A total of 13012 participants were included in the analysis. All studies were conducted in the United States. The design and methods for each included study are summarized in the Table. Inclusion criteria for all the studies were similar. In 5 studies, interventions were focused on a specific cardiac imaging modality, whereas 2 studies examined multiple cardiac imaging modalities. Each study used the AUC criteria of the relevant cardiac imaging modality to determine appropriate and inappropriate tests.

Details of each study intervention are summarized in the Table. Most initiatives were multifaceted and included several types of interventions. Six studies included a formal educational component, 2 studies used point of care decision support tools (1 using a computer-based process as its primary intervention), and 5 studies included a process for physician audit and feedback. Lin et al also stipulated to practitioners that poor adherence to AUC could result in a loss of financial reimbursement. None of the studies described any other external limitations to ordering tests (eg, by insurance payers or radiology benefits managers) imposed on physicians.

**Study Quality and Risk of Bias**

Table I in the Data Supplement provides a summary of study quality and risk of bias using either (1) the Ottawa Newcastle tool for observational studies or (2) the Cochrane risk of bias tool for randomized studies. Among the observational studies, all rated well in terms of the risk of selection bias and the risk of outcome related bias. All studies rated poorly in terms of comparability between the cohort and the control groups, primarily because none accounted for potential biases related to temporal changes in test use in either the study design or analysis. In all the observational studies, imaging tests conducted before the quality improvement initiatives were included in the control groups, whereas those conducted after the initiatives were included in the intervention groups. Furthermore, no study adjusted for other potential confounders in the statistical analysis. All observational studies were judged to be moderate in quality, primarily because of the risk of bias associated with the comparability of the intervention and control groups. The study by Bhatia et al was the only randomized study included, and it was also judged to be moderate in quality, primarily because of the potential for increased risks of biases related to the control group being exposed to components of the intervention or members of the intervention group, the use of a nonprespecified outcome for the primary analysis, and the small number of participants within each study group.

**Results of Interventions**

Overall, quality improvement interventions were associated with significantly lower odds of inappropriate testing (odds ratio [OR], 0.44 [95% confidence interval [95% CI], 0.32–0.61]; P<0.001). However, there was significant heterogeneity observed between studies (I²=70%). Heterogeneity was best explained after studies were stratified by the presence or the absence of a physician audit and feedback mechanism (I²=0 for both groups; Figure 2). The presence of a physician audit and feedback mechanism was associated with lower odds of inappropriate testing (OR, 0.36 [95% CI, 0.31–0.41]; P<0.001), whereas studies that lacked this process had no significant effect on inappropriate testing (OR, 0.89 [95% CI, 0.61–1.29]; P=0.51; P value for difference <0.001). When studies were stratified based on the presence or the absence of a formal educational component, significant heterogeneity remained within the educational intervention group (I²=72%; Figure 1 in the Data Supplement). Stratifying by the use decision support tools demonstrated that initiatives using these interventions were also associated with significantly (OR, 0.35 [95% CI, 0.22–0.56]) and consistently (I² within group=22%) lower odds of inappropriate testing, although this effect did not significantly differ from groups not using this feature, and did not fully explain differences between interventions (Figure II in the Data Supplement). There was no significant difference in the odds of inappropriate testing between the observational studies (OR, 0.41 [95% CI, 0.36–0.47]) and the randomized control trial (OR, 0.30 [95% CI, 0.15–0.58]; P=0.38).

Quality improvement initiatives were also associated with higher odds of appropriate testing (OR, 1.67 [95% CI, 1.19–2.35]), with significant heterogeneity observed (I²=87%). This heterogeneity was also partially explained by the utilization of physician audit and feedback, although to a lesser...
Discussion

This study is the first to comprehensively evaluate quality improvement initiatives aimed at reducing the amount of inappropriate tests performed across common cardiac imaging modalities. First, we have identified components of these interventions that may provide the most consistent benefits in reducing inappropriate testing. Second, we have identified important limitations in the literature to inform further studies, which is directly relevant to determining the most effective methods of designing and evaluating initiatives aimed at reducing unnecessary cardiac testing.

Determining the optimal strategies to reduce inappropriate cardiac testing is important for both improving the quality of cardiac testing and reducing unwarranted health expenditures. Although quality improvement initiatives were associated with an overall reduction in inappropriate cardiac testing, the significant heterogeneity observed between studies in our meta-analysis signifies that the effects of these interventions vary considerably. Examining individual components of these interventions is directly relevant to determining the most effective methods of designing and evaluating initiatives aimed at reducing unnecessary cardiac testing.
interventions through subgroup analysis resulted in several important findings. First, it seems that including a mechanism of physician audit and feedback is a consistent feature that affects the overall effectiveness of these initiatives, with interventions utilizing physician audit and feedback strongly associated with a reduction in inappropriate testing, and those without such a mechanism having no apparent benefit. In fact, the presence or the absence of a physician audit and feedback mechanism largely explained the observed heterogeneity in our study. Second, although the use of decision support tools were also associated with large and consistent reductions in inappropriate tests, the observed benefit of these processes did not significantly differ from initiatives without decision support tools. Importantly, decision support tools provide real-time physician audit and feedback to clinicians about the appropriateness of the test being ordered, and the lack of a significant difference in effect compared with initiatives without this specific process may have been because of the alternate methods of physician audit and feedback that were used in some of the other studies. Finally, the overall impact of a formal educational component across these quality improvement initiatives was inconsistent, with the other processes having more consistent effects.

Our results are consistent with previous studies demonstrating that quality improvement initiatives using audit and feedback mechanisms have a positive impact on health-related outcomes or indicators of quality care. In a systematic review of 21 hospital-based studies, de Vos et al.27 examined the impact of quality indicators and observed that feedback reports were one of the most commonly used and effective tools at improving hospital care. Similarly, a systematic review of 140 randomized control trials examining audit and feedback mechanisms across a variety of health services has observed that these processes modestly improved practice outcomes, with the greatest benefits occurring when baseline practice performance was low, with continuous feedback, with multiple feedback mechanisms, and when clear targets or action plans were set.28 The impact of physician audit and feedback mechanisms may also be partly related to a Hawthorn effect, where the simple awareness of a performance audit may lead to changes in behavior to improve referral practices. In this regard, a chief concern is whether the impact of quality improvement initiatives will lessen over time, particularly if they are not maintained. Long-term follow up by Bhatia et al.26 after cessation of their quality improvement initiative showed that the rates of inappropriate scans increased to preintervention levels, emphasizing the need to develop sustainable interventions to ensure their long-term effectiveness. This is a goal of larger initiatives such as the Formation of Optimal Cardiovascular Utilization Strategies program through the American College of Cardiology Foundation, which has implemented a self-directed web-based tool consisting of multifaceted educational interventions (eg, webinars and online discussion groups) and physician audit and feedback mechanisms that can be self-sustained, and has been shown to dramatically reduced the rate of inappropriate testing in participating centers.19

Using the Grading of Recommendations, Assessment, Development and Evaluation method for assessing the quality of systematic reviews, we found that the quality of evidence for interventions to reduce inappropriate cardiac testing was low (+2) (Table II in the Data Supplement).29 Although large effect sizes were observed in interventions that used physician audit and feedback, all but 1 study were observational, and the lack of adequate accounting for temporal trends in cardiac test use was an important potential source of bias. Given that all of the observational studies used a before–after approach, it is possible that the reduction in inappropriate testing could have at least partly resulted from changes in physician behavior or referral patterns that occurred simultaneously but were unrelated to the interventions. Also, given the increasing awareness of AUC, it is possible that physicians may be improving their approach to cardiac testing without any additional interventions. McNulty
et al\textsuperscript{10} observed a reduction in the number of nuclear tests performed in California, United States from 2006 to 2011, which may be related to numerous causes such as the publication and dissemination of AUC criteria, a decrease in the incidence of coronary artery disease, and changes in health provider models. The single randomized control study identified in this review by Bhatia et al\textsuperscript{21} was not impacted by this potential bias, and still demonstrated a significant reduction in inappropriate testing with the use of a multifaceted educational and physician feedback intervention. However, this study was performed at a single-center, using a specific population (eg, residents), and it is unknown whether similar effects would be observed in the broader population.\textsuperscript{21} Moving forward, an effective method to account for this and other potential sources of confounding when evaluating quality improvement interventions may be the use of a step-wedge study randomized trial design. This method is useful for interventions that are likely to provide a benefit, and can account for temporal variations in their effects. Using this approach, a quality improvement intervention is sequentially provided to participants (or clusters) over a series of time periods, until all participants have received it. By monitoring outcomes at each stage of delivery, temporal effects can be modeled.\textsuperscript{31} Moreover, mere awareness of AUC as part of these initiatives could also reduce inappropriate studies (without further active intervention), and accounting for this potential source of bias in the control group is also necessary. Using these approaches, studies can better account for potential sources of inequality between study groups, with lower potential for study bias and greater confidence that effects are related to the interventions themselves.

**Study Limitations**

A main limitation of this review is the small number of studies that were included. This could lead to potential biases in heterogeneity estimates with the use of random effects models, particularly in subgroup analyses with small number of studies. Still, we observed large differences between the effect size point estimates in groups with or without physician audit and feedback procedures, which are consistent with our observed estimates of heterogeneity. Although our results should be considered as hypothesis generating, observations provide important insights to inform the study of future initiatives that aim to reduce inappropriate cardiac testing. Also, most initiatives were heavily focused on education-based interventions. Only Lin et al\textsuperscript{25} evaluated an alternative, real-time, computer-based decision support tool in a single, private payer setting. In this study, participating physicians were exempt from radiology benefit manager authorization of studies in lieu of the online support tool on which to base the appropriateness of a given test. Physicians were free to order any test, and a 78% reduction in the odds of ordering an inappropriate study was observed using the tool. Although limited to a selected population, the large reductions in inappropriate testing that were achieved using this strategy warrant further study to determine whether the use of similar decision support tools have comparable benefits in a broader practice population. Finally, some observational studies have observed a large increase in the use of highly reimbursed cardiac imaging tests in self-referral–based practices and independent diagnostic facilities, which has been out of keeping with trends in hospitals; and it would be of considerable interest to examine the impact of interventions that target payer models or reimbursement incentives on the appropriateness of cardiac testing.\textsuperscript{32} As previously mentioned, the overall quality of the evidence was low, primarily because of studies being observational in nature and not accounting for important sources of bias, such as temporal trends in referral patterns. We have proposed additional methods to improve the design of future studies that will better account for potential source of biases and improve the overall quality of the evidence. Our systematic review did not include 1 large, multicenter study (Formation of Optimal Cardiovascular Utilization Strategies) because sample sizes were reported at the physician or group level, rather than at the level of individual cardiac tests, and therefore could not be included in our meta-analysis.\textsuperscript{33} However, the study observed a 50% reduction in inappropriate tests with a multifaceted educational, audit and feedback intervention, which is consistent with the results of our meta-analysis. Finally, it is important to acknowledge that institutions, physicians, or practices that implement quality improvement initiatives are likely groups that inherently support such strategies, and may be more inclined to respond positively. As such, the observed benefits of these interventions may decline when applied to a broader population.

**Conclusions**

Quality improvement interventions are associated with a reduction in inappropriate cardiac testing, although these benefits seem to be closely tied to the use of physician audit and feedback mechanisms. Further studies that evaluate diverse interventions using improved designs are needed to determine the most effective strategies for reducing inappropriate cardiac testing.

**Sources of Funding**

Dr Joseph holds a McMaster University Department of Medicine Early Career Research Award.

**Disclosures**

None.

**References**
