Appropriate Use Criteria For Stress Single-Photon Emission Computed Tomography Sestamibi Studies
A Quality Improvement Project

Raymond J. Gibbons, MD; J. Wells Askew, MD; David Hodge, MSc; Beth Kaping, RN; Damita J. Carryer, RN; Todd Miller, MD

Background—We previously reported the application of the 2005 American College of Cardiology Foundation appropriate use criteria for stress single-photon emission computed tomography (SPECT) imaging to patients at Mayo Clinic (Rochester, MN) in 2005 and 2006. A subsequent internal quality improvement project focused on physician education in an attempt to reduce the rate of inappropriate SPECT studies.

Methods and Results—Our 2008 physician education effort, focused on 4 specific indications that accounted for 88% of the inappropriate SPECT studies, included a presentation at medical grand rounds, a publication in the staff newsletter, meetings with physician administrators, and focused presentations to departments/divisions with many ordering physicians. We then remeasured the appropriateness of SPECT studies using previously published methods. The general categories of study indications, eg, after revascularization, were similar in 273 SPECT patients in 2008 and in our 2005 (n=284) and 2006 (n=284) cohorts. There was a trend suggesting a change in the overall classification of appropriateness over time (P=0.08) and a significant change in the rate of inappropriate studies over time (P=0.018). Inappropriate studies decreased from 14.4% in 2005 to 7.0% in 2006 before initiation of the quality improvement project. After completion of the quality improvement project, inappropriate studies increased to 11.7% (P=0.06). The 95% confidence limits for the 4.7% increase in inappropriate studies after the quality improvement project included a decrease of 0.2% and an increase of 9.6%.

Conclusions—This quality improvement project, focused on feedback, physician education, and remeasurement, did not reduce the rate of inappropriate stress SPECT studies in a single academic medical center. Similar limited interventions focused on physician education alone may have limited benefit. More extensive intervention may be necessary to improve the quality of care with appropriateness criteria. (Circulation. 2011;123:499-503.)

Key Words: coronary artery disease ▪ imaging ▪ radioisotopes

The rapid increase in imaging services, in particular cardiac imaging services, has been the source of increasing societal attention. Between 1993 and 2001, the average annual increase in cardiac imaging stress tests (stress single-photon emission computed tomography [SPECT] studies and stress echocardiograms) was 6.1%, which exceeded the rates of increase for cardiac catheterization or percutaneous coronary intervention.1 The annual rate of increase for outpatient SPECT studies accelerated to ≥15% between 1998 and 2006, but the rate of increase appears to have slowed since 2005.2,3 Imaging (primarily noncardiac) involving magnetic resonance, computed tomography, and positron emission tomography also increased at a rate of 16%/y between 1995 and 2005.4

Editorial see p 471
Clinical Perspective on p 503

In response to this concern, the American College of Cardiology Foundation (ACCF), along with various partners, has developed appropriateness criteria (AUC) for a variety of imaging modalities.5,6 In 2 previous studies,7,8 we applied the 2005 ACCF AUC for stress SPECT imaging to patients who underwent such imaging at Mayo Clinic (Rochester, MN) in May 2005 (before publication of the ACCF AUC) and in October 2006 (after publication of the ACCF criteria but before internal review or dissemination of our 2005 results). The 2005 ACCF criteria for SPECT were the only ones available at that time, although they were later updated in 2009.9 We subsequently initiated an internal quality improvement project in an attempt to educate physicians and to reduce the rate of inappropriate studies. This article reports the results of that effort.

Methods
Quality Improvement Initiative
The results of our first study applying the 2005 AUC were presented at several small internal research conferences in early 2007. The

Received July 6, 2010; accepted November 15, 2010.
From the Mayo Foundation Rochester, MN.
Correspondence to Raymond J. Gibbons, MD, Mayo Foundation, 200 First St SW, Rochester, MN 55905. E-mail gibbons.raymond@mayo.edu
© 2011 American Heart Association, Inc.
Circulation is available at http://circ.ahajournals.org
DOI: 10.1161/CIRCULATIONAHA.110.975995

499
A 1-page summary of the study, focusing on the 4 inappropriate components:

1. The principal findings of our first study were the subject of a well-advertised and well-attended (standing room only) medical grand rounds in January 2008. This presentation was available over the succeeding weeks to any staff physician who was unable to attend these grand rounds.

2. A 1-page summary of the study, focusing on the 4 inappropriate indications in Table 1, were distributed to all members of the Mayo Cardiovascular Division and published in the Mayo staff newsletter, which is distributed weekly to all Mayo staff.

3. One of the investigators (R.J.G.) reviewed the findings of our first study (including Table 1) in a personal meeting with the physician director of the cardiac education clinic, which handles most of our internal outpatient cardiology consults. The physician director subsequently met with the staff nurses, who review the records of all patients before their appointment in that clinic and preorder additional necessary studies before the patient visit.

4. Departments and divisions within the institution whose physicians were well represented in the ordering physicians for our original study were contacted. These included the divisions of general internal medicine, hospital internal medicine, and executive and preventive medicine within the Department of Internal Medicine, as well as the Department of Family Medicine. One of the investigators appeared at a subsequent staff meeting of each of these groups to review the findings of our study and the 4 most common inappropriate indications.

We announced our intention to remeasure appropriateness after these efforts were completed.

**Determination of Appropriateness**

In this study, our general methodology for the determination of appropriateness of patient studies was virtually identical to that in our 2 previous studies, except when specifically noted below, in an effort to minimize variability resulting from methodology. The study was approved by the Mayo Institutional Review Board.

**Database**

The Mayo Rochester Nuclear Cardiology Laboratory database was used for all 3 studies. This database prospectively records information on all patients undergoing stress radionuclide procedures, including symptoms of chest pain or dyspnea, and a categorization of chest pain, using the criteria of Diamond et al.,**10** as typical angina, atypical angina, or noncardiac pain. This database has been used extensively in multiple previous publications from this laboratory.**11,12**

**Assumptions**

As reported in both of our previous publications, we used a number of assumptions in an attempt to apply the AUC in a standardized fashion. These are briefly summarized here.

1. Patients whose primary symptom was dyspnea rather than chest pain were categorized as symptomatic with “atypical angina.”

2. Coronary heart disease risk was assessed with a Framingham score that determines the risk of future hard cardiac events. If patients were taking statins for hyperlipidemia, they were assigned a value of $+2$ for the low-density lipoprotein cholesterol component of this score.

3. The patient’s exercise tolerance was assessed as accurately as possible from clinical notes through the use of standardized definitions.

4. Minimally invasive surgical procedures that were not specifically identified in the relevant American College of Cardiology/American Heart Association preoperative testing guidelines were regarded as low-risk surgical procedures.

5. Patients who could not be classified using the 2005 ACCF AUC for SPECT myocardial perfusion imaging**6** or could be assigned to different indications within the same table of those criteria (which would result in different levels of appropriateness) were considered “unclassified.”

6. When prior percutaneous coronary intervention and coronary artery bypass grafting had both occurred in an individual patient, only the most recent revascularization procedure was used to classify the patient.

7. The multiple tables of the AUC were applied in a predetermined fashion.

8. Because of the difficulty in applying the AUC for follow-up testing in our first study, we developed a table to guide our cardiovascular study nurses.**8**

**Study Group**

The study group consisted of all patients who underwent stress SPECT rest/stress studies at Mayo Clinic in Rochester from March 16 to April 4, 2008. This 3-week period was required to have similar numbers of eligible patients because of the impact of the American College of Cardiology meeting and the absence of Mayo physicians during the Rochester school system spring break on laboratory volumes. We used the same exclusions used in our previous 2 studies: patients who did not grant research authorization under Minnesota State law and patients who underwent stress SPECT studies outside Rochester but within the Mayo Regional Health System as part of our laboratory outreach program.

**Patient Classification**

Each patient was classified using the same methodology that we described previously. In brief, the same 2 experienced cardiovascular nurse-abstractors (B.K. and D.M.J.), who are not affiliated with the Nuclear Cardiology Laboratory, reviewed each patient. Patients were designated as unclassified (as defined above) or were classified as appropriate, uncertain, or inappropriate. When the 2 nurse assessments agreed, their classification was final. When they did not, a single staff nuclear cardiologist (R.J.G.) resolved the discrepancy. In our original study,**7** a consensus of 2 staff physicians was used. On the basis of our subsequent experience,**9** we thought that this procedure could be simplified without compromising the quality of the analysis. The results that are presented subsequently represent the final classification of each patient.

**Statistical Analysis**

The results of the present study and the 2 previous studies were compared by use of a $\chi^2$ test for the overall comparison of appropriateness. When this test showed a nonsignificant trend ($P=0.08$) suggesting an overall difference, an exploratory analysis of
the results of each of the individual categories was performed. Continuous variables were compared among the 3 groups through the use of ANOVA. \( \kappa \) Statistics were used to investigate the agreement between the nurses’ evaluations.

### Results

#### General

There were 273 eligible stress SPECT myocardial perfusion image patients between March 16 and April 4, 2008. A 3-week period was required to have a number of patients comparable to the 284 studies that were eligible in 2 weeks in 2005 and in 2 weeks in 2006. The demographics for the 2008 patients are shown in Table 2 and compared with the demographics in each of the 2 previous studies. The 2008 cohort was significantly younger. There were no other significant differences between the 3 cohorts in gender, risk factors, or cardiac history.

#### Indication for Testing

The most common general indication for testing of the patients in the present study was referral for testing after revascularization \( (n=88, \ 32\%) \), followed by referral to evaluate symptoms \( (n=71, \ 26\% \) ), follow-up testing after prior SPECT myocardial perfusion imaging \( (n=61, \ 22\%) \), and screening for suspected coronary artery disease in asymptomatic patients without other indications \( (n=38, \ 14\%) \). These percentages were all similar to those seen in our 2 previous studies. In our institution, preoperative assessment before noncardiac surgery is performed more commonly with stress echocardiography.

#### Observer Agreement

The overall level of agreement between the 2 cardiovascular nurse-abstractors for the 2008 data was very good with a \( \kappa \) of 0.738 (95% confidence interval, 0.669 to 0.808).

### Table 2. Patient Characteristics

<table>
<thead>
<tr>
<th>Year</th>
<th>Present Study, March–April 2008</th>
<th>Previous Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age, y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Women, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diabetes mellitus, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hypertension, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hyperlipidemia, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smoking history, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prior MI, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prior PCI or CABG, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dyspnea, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rest ECG normal, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI ≥30 kg/m², %</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Present Study, March–April 2008</th>
<th>Previous Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age, y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Women, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diabetes mellitus, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hypertension, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hyperlipidemia, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smoking history, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prior MI, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prior PCI or CABG, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dyspnea, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rest ECG normal, %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI ≥30 kg/m², %</td>
<td></td>
</tr>
</tbody>
</table>

Age is shown as mean ± standard deviation.

MI indicates myocardial infarction; PCI, percutaneous coronary intervention; CABG, coronary artery bypass graft surgery; and BMI, body mass index.

#### Table 3. Classification of Appropriateness by Study Cohort

<table>
<thead>
<tr>
<th>Year</th>
<th>Total, n</th>
<th>Appropriate, n (%)</th>
<th>Uncertain, n (%)</th>
<th>Inappropriate, n (%)</th>
<th>Unclassified, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>284</td>
<td>182 (64.1)</td>
<td>30 (10.6)</td>
<td>41 (14.4)</td>
<td>31 (10.9)</td>
</tr>
<tr>
<td>2006</td>
<td>284</td>
<td>188 (66.2)</td>
<td>33 (11.6)</td>
<td>20 (7.0)</td>
<td>43 (15.1)</td>
</tr>
<tr>
<td>2008</td>
<td>273</td>
<td>164 (60.1)</td>
<td>36 (13.2)</td>
<td>32 (11.7)</td>
<td>41 (15.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Total, n</th>
<th>Appropriate, n (%)</th>
<th>Uncertain, n (%)</th>
<th>Inappropriate, n (%)</th>
<th>Unclassified, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>284</td>
<td>182 (64.1)</td>
<td>30 (10.6)</td>
<td>41 (14.4)</td>
<td>31 (10.9)</td>
</tr>
<tr>
<td>2006</td>
<td>284</td>
<td>188 (66.2)</td>
<td>33 (11.6)</td>
<td>20 (7.0)</td>
<td>43 (15.1)</td>
</tr>
<tr>
<td>2008</td>
<td>273</td>
<td>164 (60.1)</td>
<td>36 (13.2)</td>
<td>32 (11.7)</td>
<td>41 (15.0)</td>
</tr>
</tbody>
</table>

Overall Appropriateness

The overall classification of the stress SPECT studies in the present study compared with our previous studies is shown in Table 3. There was a nonsignificant trend toward a change in the overall classification of appropriateness over time \( (P=0.081 \) by \( \chi^2 \) statistic). There was no significant difference in the unclassified patients \( (P=0.25) \), the appropriate patients \( (P=0.31) \), or the studies of uncertain appropriateness \( (P=0.63) \). There was a significant change in the rate of inappropriate studies over time \( (P=0.018; \) the Figure). However, although (as previously reported) there was a decrease in inappropriate studies between 2005 and 2006 \( (P=0.004) \), there was an increase between 2006 and 2008 that was of borderline statistical significance \( (P=0.06) \). The change in inappropriate studies between 2006 and 2008 was 4.7% (95% confidence limits, −0.2 to 9.6). Overall, between 2005 and 2008, there was no significant difference in inappropriate studies \( (P=0.34) \), indicating that the significant change that we previously reported in 2006 was not maintained in 2008.

### Discussion

We sampled 3 weeks in 2008 to have a number of patients comparable to our earlier 2-week samples from 2005 and 2006. As noted earlier, this primarily reflected the effects of external events on laboratory volumes in March 2008. The number of eligible stress SPECT studies per month between March and October was 512 in 2005, 534 in 2006, and 458 in 2008. The overall broad indications for stress SPECT imaging did not change over time. There was only a trend toward
a change in the overall classification of appropriateness over time, but our exploratory analysis suggested that there was a significant difference in inappropriate studies across the 3 samples, although we did not adjust for multiple comparisons. A decrease in inappropriate studies occurred between 2005 and 2006, before our quality improvement project, but this was not sustained in 2008. The rate of inappropriate studies actually increased after our quality improvement effort, although this increase was only borderline significant. Thus, the overall finding was that our quality improvement effort, consisting primarily of audit and feedback at a group level and physician education, was not successful in reducing the rate of inappropriate studies.

We sought to apply the same methodology throughout all 3 time periods. The agreement between the same 2 cardiovascular study nurses in categorizing the 2008 patients was very good. However, the application of rigorous methodology to our retrospective review of the patients was resource intensive and, along with our limited funding, limited the sample size studied. This limited our power to detect modest differences, particularly because the rate of inappropriate studies had already declined from 14% to 7% before initiation of our project. (We were not yet aware of this decline when this project was designed.) However, the 95% confidence limits on the observed 4.7% increase in inappropriate studies that we observed between 2006 and 2008 ranged from a decrease of 0.2% to an increase of 9.6%. Thus, even with our limited sample size, a substantial decrease in inappropriate studies as a result of our quality improvement effort is unlikely.

Our negative results may reflect the limited intervention (physician education and group feedback) that we performed. We did not perform prepresentation and postpresentation testing to assess learning. Bradley et al identified 8 features of hospitals that successfully reduced door-to-balloon time in treating ST-elevation myocardial infarction. Our effort included only 3 of these 8 features. We had an explicit goal (reducing the rate of inappropriate studies), the visible support of the leadership of both the departments of medicine and radiology (the departments jointly funded our initial study), and clearly identifiable individual thought leaders who made multiple presentations. We did not use innovative standardized protocols or collaborative interdisciplinary teams. The limited scope of this project did not test the organizational culture. Noticeably absent from our effort was rapid, patient-specific data feedback at either a point-of-order or a point-of-service location. Our data feedback was provided only on a broad laboratory level by retrospective review rather than on a real-time patient-specific and ordering physician— or nurse-specific basis. Although Mayo Clinic in Rochester has a mature electronic medical record, it was not configured to provide clinical decision support at the time of our study. Previous Cochrane reviews of both continuing medical education and audit and feedback had found generally modest effects for both of these interventions. Solberg et al reported that experienced guideline implementers favor as many as 25 strategies that include the organizational capabilities for change rather than more limited efforts focused on the individual physician.

Our negative results may also reflect “competing” external information, including physician approval or resistance to ongoing national discussions of imaging appropriateness and healthcare reform. Our study design did not permit a contemporary control group to test for this effect.

To the best of our knowledge, only 1 previous study has examined trends in the rates of inappropriate SPECT studies over time in clinical practice. Using a point-of-service Internet-based tool, Hendel et al reported on 6351 patients enrolled in 6 centers in a study supported by the ACCF and UnitedHealthcare. Reports of appropriate, uncertain, and inappropriate tests were made available online in real time at each of the sites that participated. A summary report was also provided to each site. Only 4 of the 6 sites had sufficient data to examine inappropriate tests at 3 different time periods. Only 1 of these 4 sites had a significant decrease in inappropriate studies over time, from 22% (40 of 182) at baseline to 13.3% (34 of 256) at the end, reflecting the efforts of a highly motivated management team that held group meetings and discussions to educate physicians. The rate of inappropriate testing at baseline in this center was the highest in the study, appreciably higher than ours. However, the total number of patients studied was actually smaller than in our study.

Previous studies of interventions to improve the appropriate use of imaging in noncardiac situations have had mixed results. A recent Cochrane review identified 28 intervention studies to improve the appropriate use of imaging in patients with musculoskeletal conditions. In the management of osteoporosis, there was a modest benefit of intervention (10% improvement). For studies of low back pain, the effects were variable. For other musculoskeletal conditions, the effects were negative. Solberg et al reported the use of an electronic medical record decision support tool on the ordering of computed tomography and magnetic resonance imaging of the head and magnetic resonance imaging of the lumbar spine. They demonstrated an increase in appropriate studies (from 79% to 89%) after implementing decision support, without any change in the proportion of tests with positive findings or likely impact on patients. Radiology benefit managers may slow the rate of growth of imaging, but their impact on appropriateness is less well established.

In addition to the limitations of sample size, the absence of real-time data feedback using decision support tools, and the absence of a contemporary control group, our study was limited to a single academic medical center in the northern Midwest where the physicians are salaried and have no financial incentives to order tests. The rate of SPECT imaging reported in the Dartmouth Healthcare Atlas in this region is well below the national average. Our findings therefore may not apply to other practices or other regions. We used the 2005 AUC for SPECT perfusion imaging because they were the only AUC available at the time of this study. We have already applied the 2009 criteria to the 2005 patient cohort and reported that these updated criteria increased the rate of studies that were classified as inappropriate. The rate of inappropriate studies reported here for 2006 and 2008 is therefore likely an underestimate of the true rate based on the 2009 AUC.
Conclusions
Our quality improvement project devoted to physician education and group feedback did not reduce the rate of inappropriate stress SPECT studies in a single academic medical center. Our results suggest that interventions focused on physician education and measurement alone may have limited benefit. Further studies are needed to explore additional measures, including decision support with an electronic medical record, to achieve the full potential of AUC to improve the quality and efficiency of cardiovascular care.

Disclosures
Dr Askew has a research grant from General Electric Medical Imaging, but it is not for support of this study. Dr Miller has a research grant from Bristol-Myers Squibb, but it is not for support of this study. Dr Gibbons serves as a consultant to Molecular Insight Pharmaceuticals, Cardiovascular Clinical Studies (WOMEN Study), and Lantheus Medical Imaging. The other authors report no conflicts.

References

CLINICAL PERSPECTIVE
In a previous study, we reported the application of the published American College of Cardiology Foundation appropriate use criteria for stress single-photon emission computed tomography (SPECT) imaging to patients at Mayo Clinic (Rochester, MN) in 2005 and 2006. This study reports the results of a subsequent internal quality improvement focused on physician education, including a presentation at Medical Grand Rounds, a publication in the staff newsletter, meetings with physician administrators, and focused presentations to pertinent department/divisions of ordering physicians. Remeasurement of the appropriateness of SPECT studies was then performed using previously published methods. Unfortunately, the rate of inappropriate SPECT studies did not decrease as hoped but in fact showed a borderline increase (P=0.06). Thus, this quality improvement project did not achieve the desired reduction in the rate of inappropriate stress SPECT studies in a single academic medical center. More extensive intervention may be necessary to achieve quality improvement with appropriateness criteria.