Methods and Limitations of Assessing New Noninvasive Tests
Part I: Anatomy-Based Validation of Noninvasive Testing

Rory Hachamovitch, MD, MSc; Marcelo F. Di Carli, MD

The introduction and dissemination of new technology provide the potential for enhancing and expanding our understanding of disease processes (eg, atherosclerosis, myocardial dysfunction) and extending our treatment options while providing a tool for monitoring therapeutic responses. However, the growth of cardiac imaging has profound cost implications that will be exacerbated if newer technology is widely disseminated and used freely without appropriate validation. Hence, technology validation has become an important consideration in today’s healthcare reality. Our goal is to provide a critical review of the methods and challenges inherent to the validation of existing or emerging noninvasive imaging technologies.

How Should Noninvasive Testing Be Viewed?
Historically, imaging has been considered in the context of anatomic end points. A shift from anatomy-based to outcomes-based assessments of testing has been accepted. More recently, a further shift occurred from imaging for risk identification to imaging for identification of patients’ optimal therapeutic management, ie, identifying a therapeutic approach associated with optimal survival or improved well being after a given test result for a patient. This review focuses on diagnostic approaches; the second part focuses on test validation using risk and benefit end points.

What Is Technology Assessment for Imaging Modalities?
In the context of assessing cardiac imaging, several factors must be considered. First, the assessment of a new modality is not a simple determination of sensitivity and specificity. Rather, it is a stepwise, multifactorial process incorporating diagnostic, prognostic, therapeutic, resource use, cost-effectiveness, and other end points that considers the perspectives of patients, payers, ordering physicians, and the healthcare system. A series of questions drive this process: Does the modality work? For which end point? In which patient? At what cost? How does it compare with other modalities? Can it be used to improve clinical outcomes?

Importantly, the end product of this assessment process must remain practical. Although “scientific” differences may be found between tests, the clinical implications and relevance of these differences must be considered; do “prettier images,” those with superior resolution and image quality, necessarily translate to clinically relevant information, improved patient care, or better outcomes?

Study Design and Sources of Error
The literature validating noninvasive testing has 2 serious limitations: the paucity of well-conducted and well-designed prospective randomized trials guiding the performance and optimal use of testing and its domination by observational studies, usually small, single-center, retrospective studies. The latter studies are analytically challenging because of the various biases introduced when patient cohorts are drawn from “routine testing.”

Bias
Bias is generally defined as a systematic error in the design or execution of a study that results in an inaccurate estimate of test accuracy. Importantly, bias and/or confounding are potential alternative explanations of any study result; thus, seeking out and explaining these factors is crucial for study validity. To temper these threats to study validity, various statistical techniques—matching, restriction, stratification, and multivariable modeling—can be used to control or limit these potential sources of error (for review, see elsewhere).

Numerous biases relevant to imaging may be introduced by both pretest (eg, patient selection, data collection, pattern of test ordering) and posttest (eg, image interpretation, referral to gold standard, posttest therapeutics) factors (Table 1). Because referral to noninvasive testing occurs through many pathways (Figure 1), studies using patients referred for “routine testing” are intrinsically biased because they do not consider the “denominator” of patients from which their cohort is drawn or the specific reason why their patients were referred to that specific test rather than alternative tests. Intersite variability in referral patterns further compromises the generalizability of these results.
Image interpretation may introduce additional biases, most notably the use of clinical data by the reader (eg, likelihood of coronary artery disease [CAD] and/or symptoms). For example, mild anterior wall ischemia on stress single-photon emission computed tomography (SPECT) in a woman may be interpreted as an abnormality rather than attenuation if the reader is informed of a recent left anterior descending coronary artery territory percutaneous coronary intervention. This introduces a bias and results in compromised generalizability and likely overestimation of the test value. Furthermore, studies using blinded visual or quantitative software core laboratory readings will likely have dissimilar results compared with studies using routine readings because the accuracy of the “data-enhanced” visual readings will probably be more accurate.

The most pervasive and important bias is introduced by selective posttest use of catheterization, the gold standard of diagnostic testing. Because limited posttest catheterization is performed, only a nonrandom subset of all patients referred to testing will have anatomic data available. This bias (partial verification bias) results in reduced numbers of subjects who are false or true negatives with relative increased numbers of true and false positives, resulting in increased sensitivity and markedly reduced specificity (Figure 2).

A prognostic counterpart to the diagnostic partial verification bias has been reported recently. Because imaging results dictate the intensity of posttest patient management and intervention, patients with abnormal (particularly ischemic) tests will be preferentially referred to revascularization procedures that will, in turn, alter the natural history of their CAD so that their risk is reduced. Hence, even if these revascularized patients are removed (censored) from analyses, survival rates in nominally higher-risk subsets will be attenuated, and the observed prognostic value of the test will be reduced. The implications of this bias are discussed in the context of prognostic validation of testing.

**Confounding**

Although bias creates an incorrect association, confounding generates an association that is correct but misleading (and possibly unique to the study population). For example, a study assessing sex-related post-SPECT resource use differ-

### TABLE 1. Partial Listing of Bias Often Encountered in Studies of Noninvasive Testing

<table>
<thead>
<tr>
<th>Type of Bias</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectrum bias</td>
<td>Types of cases and controls are not representative of the population</td>
</tr>
<tr>
<td>Limited challenge</td>
<td>The exclusion of more challenging diagnostic patients, resulting in increased observed test accuracy</td>
</tr>
<tr>
<td>Partial verification bias (work-up, posttest referral, validation, sampling, or selection bias)</td>
<td>Study inclusion only of patients who have undergone a gold-standard test</td>
</tr>
<tr>
<td>Test interpretation bias</td>
<td>Available information distorts the reported test results</td>
</tr>
<tr>
<td>Recall bias</td>
<td>Individuals with disease recall and report their condition or events differently than do controls</td>
</tr>
<tr>
<td>Interviewer bias</td>
<td>Systematic error in soliciting, recoding, or interpreting subject data</td>
</tr>
<tr>
<td>Misclassification bias</td>
<td>Subjects erroneously categorized with respect to either exposure or disease status</td>
</tr>
</tbody>
</table>

![Figure 1. Patient flow through testing strategies. After clinical evaluation, patients can be identified as needing further testing or being low risk (medical management only; A). The former may be evaluated in several ways, including referral to lower-cost/complexity testing (B), higher-cost/complexity testing (C), or a gold standard (D). At each stage, further testing, referral to the gold standard, or no testing is possible. Dashed lines show an optimal sequential testing strategy. The variety of potential evaluation pathways highlights the potential biases and limited generalizability of studies drawn from routine testing. For example, a study examining the diagnostic accuracy of stress imaging would enroll only patients in pathway E, not considering those in F (for the denominator E + F) or even the larger denominator B + C + D. GP indicates general practitioner; FM, family medicine physician; and IM, internist.](image1)

![Figure 2. Impact of posttest referral bias. Because of the disproportionate referral to catheterization (gold standard) of patients with abnormal vs normal tests, studies of test accuracy usually find relatively lower quantities of true negatives (TN) and false-negative (FN) studies and more true positive (TP) and false-positive (FP) studies. This results in the pattern of findings associated with partial verification bias, a slightly elevated sensitivity with a marked reduction in specificity.](image2)
In the context of validation, several classes of tests can be defined. A new test similar to, but an enhancement of, a previously validated test (eg, adding a contrast agent to echocardiography) or use of a common approach but in a different modality (eg, stress perfusion with echocardiography) or use of a common approach but in a different domain (eg, coronary stenoses or to invasive intravascular ultrasound for assessing atherosclerosis).

Types of Testing

In the context of validation, several classes of tests can be defined. A new test similar to, but an enhancement of, a previously validated test (eg, adding a contrast agent to echocardiography) or use of a common approach but in a different modality (eg, stress perfusion with echocardiography or cardiac magnetic resonance) is one class. These can be assessed with respect to their accuracy compared with a gold standard (eg, a “new” versus an “old” test for detection of catheterization-identified CAD) or their agreement with a validated test (stress perfusion with echocardiography versus stress perfusion with SPECT as a gold standard “equivalent”). The use of 2 end points concurrently, one to ascertain its equivalence clinically (eg, diagnostic or prognostic accuracy) and a second to assert superiority in a different domain (eg, faster performance time, reduced cost or radiation exposure, enhanced reproducibility), may be useful. A key factor to the validation of a test is revealing what is new, important, or advantageous and identifying an end point to capture this information.

Alternatively, a new-modality imaging structure or process previously not feasible defines a new class of test. For example, magnetic resonance spectroscopy has no previously validated test with which it can be compared; hence, its validation in patients could be problematic. Computed tomography angiography (CTA), although a “first in its class” with respect to noninvasive atherosclerosis imaging, can be compared with invasive angiography for detecting epicardial coronary stenoses or to invasive intravascular ultrasound for assessing atherosclerosis.

Selecting the Correct End Point

End points or outcomes used in studies vary considerably, but diagnostic and/or prognostic end points are most commonly used. A modality may perform dissimilarly with respect to these 2 end points. Similarly, different metrics from a test may have different associations with different outcomes. For example, ischemia is more closely associated with softer end points (myocardial infarction), whereas left ventricular function is more predictive of cardiac death. For the purposes of this review, we address several categories of clinical test validation: using a diagnostic end point, using an outcomes end point, and studies of agreement and reproducibility (the last 2 are in part 2 of this review). Preclinical studies of imaging modalities generally use a range of study designs and end points that are beyond the scope of this review.

In specific situations, selecting the optimal end point for validation may be challenging. For example, when assessing the use of imaging in patients with chest pain presenting to an emergency room, what is the optimal end point? Identification of acute myocardial infarction? Recurring admissions? Short- or intermediate-term death? Because of power issues, some studies have used posttest resource use as the end point.

For tests of subclinical disease, optimal end points would assess CAD development and progression, whereas for a new stress perfusion test (or stress imaging agent or stressor), the demonstration of comparable posttest resource use may be equivalent to demonstration of similar prognostication (eg, a similar pattern of posttest referral to catheterization or revascularization as a function of the test results for both the “new” and “old” tests). It is important to consider each study individually and to focus on the specific questions being addressed, particularly in the context of how the investigators believe the test will fit into a clinical strategy. Furthermore, most tests are validated in, and recommended for, specific patient populations. In the case of stress testing, the population is those patients at intermediate to high likelihood of CAD or risk of adverse events.

Beyond diagnostic and anatomic end points, various end points can be used to assess cardiac structure, function, or perfusion (or their changes) or quality-of-life domains and are valid but surrogate end points. Surrogate end points and their limitations are discussed in the second part of this review. An imaging test may serve as a potential screening test (tests performed in asymptomatic patients without clinical indication of disease but at risk for developing the disorder).
Hennekens and Buring\textsuperscript{12} provide a discussion of screening test validation and assessment.

**Anatomy-Based Validation of Diagnostic Testing**

Historically, CAD has been defined by anatomic criteria (eg, absent, present, single vessel, multivessel). Consequently, initial reports of new imaging modalities assess diagnostic accuracy, the agreement between the results of a test and those of a reference standard.\textsuperscript{13} Alternatively, detection of sufficient disease to justify revascularization (eg, left main CAD \([\geq 50\%\) stenosis], 3-vessel CAD) also has been used as an end point. Thus, diagnostic-based validations can use a number of end points or combinations thereof. The greatest challenge for these studies, as discussed above, is identifying a group of subjects who have been selected with minimal pretest and posttest referral biases.

**Methods for Reporting of Diagnostic Accuracy**

The basic measures of diagnostic accuracy are sensitivity and specificity. For clinicians, positive and negative predictive values carry considerably more relevance, expressing the expected likelihood that the results of a test represent the patient’s disease status. Hence, with a negative predictive value of 95\%, a negative test result suggests a 5\% likelihood that disease is present. It must be noted that predictive values are determined both from sensitivity and specificity and from prevalence. Thus, a test with sensitivity and specificity of 90\% will have positive and negative predictive values of 95\% and 79\%, respectively, when the prevalence is 70\% but 83\% and 94\% when the prevalence is 35\%.

**Aggregate Measures of Diagnostic Accuracy**

It is convenient to express the performance characteristics of a test or to compare the performance of 2 tests with a single metric. Several measures incorporate sensitivity and specificity into a single metric.\textsuperscript{13} Test accuracy, defined as the proportion of all tests that are correct (true positives plus true negatives divided by all patients) is commonly used to express the likelihood that the test result is correct. Its limitations include the fact that it is a prevalence-weighted average of sensitivity and specificity; thus, patient mix will influence its value. For example, when a very low-prevalence population is tested, merely assuming all tests to be negative will yield a very high accuracy. Furthermore, 2 tests with the same accuracy, despite very different sensitivity and specificity (2 tests with sensitivities of 100\% and 0\% and specificities of 0\% and 100\%), will yield identical accuracies in the setting of a prevalence of 50\%.

Receiver-operating characteristics (ROC) curves define the ability of a test to discriminate between disease presence and absence or to compare the discriminative properties of \(\geq 2\) tests (for details on this method, see the discussion by Zou et al\textsuperscript{14}). ROC curves represent the tradeoff between sensitivity and the false-positive rate (1–specificity) across decision thresholds, thereby defining test performance across these thresholds and identifying the optimal decision threshold for test abnormality (generally, the point on the curve closest to the upper left corner of the plot).\textsuperscript{14} ROC analysis is particularly meaningful when the value of an imaging test is considered in the context of clinical data. An example is the use of ROC curves to compare the predictive value of coronary artery calcium plus Framingham Risk Score (area under the ROC curve, 0.68) with Framingham Risk Score (area under the ROC curve, 0.63; \(P<0.001\)) alone for the identification of risk for myocardial infarction or cardiac death.\textsuperscript{15}

ROC curves have several limitations. They assume clinical equivalence of false-negative and false-positive results. For example, given a new test to diagnose acute myocardial infarction, a false positive may result in an unnecessary catherization, whereas a false negative may result in an untreated myocardial infarction, a missed diagnosis, and its sequela. Clinically, the latter may be of greater significance and hence should be weighted more than the former. ROC curve application also must be tempered by clinical reality; although it is advantageous to assess test discrimination across all diagnostic thresholds, all thresholds may not have clinical relevance. For example, a clinician may be disinterested in test sensitivity when specificity falls below a specific threshold. To counter this limitation, 2 approaches exist: the sensitivity at a fixed false-positive rate\textsuperscript{13} and, of greater value, the determination of the partial area underneath the ROC curve. The latter defines a clinically relevant range of values between 2 false-positive rates (hence, specificities) and limits the ROC area to that range.\textsuperscript{13}

The likelihood ratio is another single index of diagnostic accuracy that is calculated as a ratio of the probability of a specific test result occurring in patients with the known condition to the probability that the same result would occur in patients without the condition. Although likelihood ratio values >1 indicate a test result associated with the presence of disease and values <1 are associated with the absence of disease, only at values >10 and <0.1, respectively, is there strong evidence to “rule in” or “rule out” the presence of disease.\textsuperscript{16} These thresholds notwithstanding, representative positive and negative likelihood ratios are 1.3 and 0.5 (bias corrected, 2.3 to 2.5 and 0.7 to 0.8, respectively)\textsuperscript{17} for stress echocardiography and 1.1 and 0.15 (bias corrected, 2.0 and 0.44, respectively)\textsuperscript{18} for SPECT. These values are not dissimilar to those based on data from a recent meta-analysis (positive likelihood ratio, 2.4 to 3.7; negative likelihood ratio, 0.19 to 0.20).\textsuperscript{19} Hence, diagnostically, these commonly used tests fall below accepted thresholds for testing in general.

**Can Noninvasive Testing Be Validated With an Anatomy-Based Approach?**

Although diagnostic-based assessment of testing is generally accepted, that the limitations of this approach may compromise its validity is increasingly appreciated. Two important limitations to this approach must be noted. Invasive anatomic measures are a “tarnished” gold standard in that they are not closely associated with either fractional flow reserve or intravascular ultrasound results.\textsuperscript{2} In addition, as discussed above, numerous biases threaten the validity of anatomy-based studies. In practice, posttesting patient management is driven by testing results. After an abnormal result, catheterization is very likely; after a normal test, additional testing is unlikely. As mentioned, this bias lowers specificity and
increases sensitivity compared with a sample without this bias (Figure 2).8

The magnitude of the impact of this bias is generally underappreciated, and several approaches to alleviate this problem have been proposed. First is the normalcy rate (the frequency of a normal study among low [<5%] -CAD-likelihood patients), a surrogate for specificity. Although used in imaging, normalcy has not been formally validated, nor does it appear in the epidemiology literature. Understanding this metric is problematic. Why were these low-likelihood patients referred to the test? To whom can their results be generalized? Did unmeasured covariates drive referral to testing? Finally, whether normalcy and specificity rates are associated and whether the association persists across likelihood of disease are as yet undefined. Consequently, the value and validity of this metric are unclear.

A second approach to avoiding bias is to refer patients for catheterization after testing regardless of test results. Although not without issues, this approach is limited to rigorously defined and executed investigations.14,20 Furthermore, recruitment must be limited to candidates for testing rather than stable patients referred for catheterization and recruited for post hoc testing. An alternative design for comparing 2 techniques with a gold standard is for patients to undergo both tests, and if either test is abnormal, the patient can justifiably undergo catheterization. If both tests are negative, it is unlikely that the patient has significant disease. Providing that neither test has unacceptable false-positive rates, this approach, although not validated, may prove helpful.

Finally, formula-based methods to correct for referral bias have been proposed. Generally, these methods are used in studies of postimaging patients, a subset of whom underwent catheterization. The correction is based on the results of this subset, which are then extrapolated to “correct” observed accuracies in the overall cohort,20 although more elaborate modifications exist7,8,18,21 (for details, elsewhere”). The impact of the corrections is dramatic17,18 (Table 2). Because design-related bias in diagnostic studies is ubiquitous,8 reports of test accuracies without elimination of (by study design) or correction for (by formulas) biases are suspect. However, certain assumptions underlying these corrections are questionable.20 Although a reasonable first step, the validity and accuracy of these methods are undefined. Newer algorithms of increasing sophistication and accuracy are available.7,21 Ideally, future verification bias corrections will incorporate more robust multivariable modeling.

### Understanding Early Reports of Diagnostic Accuracy

Reports of the diagnostic accuracy of CTA are an excellent example for reviewing the challenges in understanding the accuracy of a new modality. Although early reports comparing CTA with catheterization reported very high sensitivities and specificities (≥ 90% for both), their analyses were limited to larger-caliber vessels (≥1.5 to 2.0 mm) and excluded nonvisualized vessels.1 Inclusion of smaller vessels, necessary to define the accuracy of the test, reduces sensitivity,1 whereas reclassifying nonvisualized vessels as abnormal (because disease cannot be excluded and further evaluation is necessary) reduces observed specificity.1

Typically, these studies consist of stable patients referred for elective catheterization and recruited to undergo CTA. These are generally higher-risk patients with greater CAD prevalences. Generalizing accuracy estimates from these patients to lower-risk cohorts (more likely to undergo CTA) is problematic. Based on pooled data (sensitivity, 93%; specificity, 81%; prevalence, 63%), the calculated positive and negative predictive values (86% and 88%, respectively) change substantially in lower-prevalence cohorts: positive and negative predictive values for a prevalence of 30% are 68% and 96%, respectively, and for a prevalence of 15% are 46% and 98%, respectively.1 Finally, reported sensitivities and specificities also must be considered in the context of known referral biases associated with particular study designs. For example, when studies report high sensitivity and lower specificity with CTA,1,7,8 the presence of partial verification bias must immediately be suspected. The sensitivity must be considered an overestimate, and the specificity must be thought of as an underestimate.

### Understanding the Role of Meta-Analysis

Meta-analytic techniques are frequently applied to combine data from multiple studies to yield pooled estimates of test accuracy with greater power and possibly generalizability compared with single-site studies. Although meta-analyses often are cited to support the value of a modality or to demonstrate the superiority of one modality over another, it must be emphasized that the results of a meta-analysis are inherently limited by the limitations of the original data, and as evidenced by a recent meta-analysis,19 errors can be introduced. If 2 technologies are compared, equivalence in the “age” of the methods used must be ensured (eg, recent studies of technology A versus older studies of technology B). Furthermore, if comparisons of technologies for detecting CAD are made, it is critical to adjust for differences in the characteristics of patients between studies. In addition, differences in interstudy resource use must be accounted for because they determine the intensity of the referral biases present that likely vary between sites and that may corrupt the results. Finally, careful attention must be paid to the statistical methods used to avoid methodological error.22
Table 3. Results of Sample Size Calculations for Studies Comparing 2 Isotopes With Respect to Identification of Anatomic CAD

<table>
<thead>
<tr>
<th>Isotope A</th>
<th>Isotope B</th>
<th>Delta Sensitivity, %</th>
<th>Needed per Group, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>88</td>
<td>2</td>
<td>5142</td>
</tr>
<tr>
<td>90</td>
<td>85</td>
<td>5</td>
<td>918</td>
</tr>
<tr>
<td>90</td>
<td>80</td>
<td>10</td>
<td>266</td>
</tr>
<tr>
<td>85</td>
<td>80</td>
<td>5</td>
<td>1212</td>
</tr>
<tr>
<td>80</td>
<td>75</td>
<td>5</td>
<td>1464</td>
</tr>
</tbody>
</table>

Calculations assume power of 90%, α of 0.05, and equal numbers in the 2 groups.

Validation of New Agents, Tracers, and Methods: Issues of Statistical Power

Numerous diagnostic validation studies comparing newer and older methodologies—imaging methods (eg, SPECT versus positron emission tomography), stress agents (eg, adenosine versus dipyridamole), isotopes, or use of contrast (stress echocardiography)—have been reported. Given the relatively small size of these studies, whether adequate power is present is a concern, especially as many reports do not include a power analysis.

As an example, numerous studies have compared the predictive value of various radioisotopes for the identification of anatomic CAD in catheterized cohorts. These studies compare accuracies with either 2 cohorts, each of whom underwent testing with 1 isotope, or 1 cohort that underwent 2 tests (1 with each isotope). With the first approach (assuming patient randomization to minimize biases), assessing the superiority of a new tracer when both new and old traces work reasonably well and the difference between them is small is problematic, as evidenced by the sample sizes needed (Table 3). Even with the second approach, a paired study would require >400 patients to detect a 5% difference in sensitivity. Furthermore, examining sensitivity differences assumes that changes in tracer-associated defect size result in observable differences in accuracies. This is more likely the case in patients with milder CAD (in whom defect size reduction may translate to defect elimination) but not in all CAD patients.

More subtle intertracer differences could be detected using fewer patients by comparing intertracer defect size for the same amount of anatomic CAD. Even with a small defect size difference (eg, 15% versus 12%; power, 90%; effect size, 0.25; α = 0.05) anticipated, only 171 patients would be required. A limitation of this approach is the need to recruit patients for a second SPECT. Alternatively, another approach would be to study consecutive patients undergoing stress SPECT with either agent (preferably randomized) and to use multivariable modeling to assess the association between the agent used and defect size after adjustment for confounders.

Conclusions

Despite the pitfalls discussed here, the diagnostic assessment of testing will continue to be reported as an early indicator of test efficacy. Because ideal diagnostic validation studies are unlikely to be performed, clinicians and investigators must be aware of the limitations inherent in this approach. We must remember that test validation consists of a series of studies constituting a body of evidence. These studies necessarily require patients from a spectrum of risk categories and likelihoods of CAD. This will probably be associated with a variety of study designs and sampling schemes. As discussed in part 2 of this review, risk- and benefit-based approaches are superior, especially with respect to identifying cost-effective patient management strategies.

Disclosures

Drs Hachamovitch and Di Carli have received grant support from Bracco Diagnostics, Astellas Pharma US, GE Healthcare, and Siemens Medical Solutions and material support from Vital Images. They are on the speakers’ bureau for Astellas Pharma and GE Healthcare. Dr Hachamovitch is on the speakers’ bureau for Lantheus Medical Imaging and consults for King Pharmaceuticals and Lantheus Medical Imaging. Dr Di Carli is on the speakers’ bureau for Bracco Diagnostics and is on the advisory board for Bracco Diagnostics, GE Healthcare, and Lantheus Medical Imaging.

References


**KEY WORDS:** epidemiology ■ heart diseases ■ imaging ■ statistics ■ tests
Methods and Limitations of Assessing New Noninvasive Tests: Part I: Anatomy-Based Validation of Noninvasive Testing

Rory Hachamovitch and Marcelo F. Di Carli

_Circulation._ 2008;117:2684-2690
doi: 10.1161/CIRCULATIONAHA.107.708586
_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2008 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/117/20/2684

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/