Original Investigation

Durability of Class I American College of Cardiology/American Heart Association Clinical Practice Guideline Recommendations

Mark D. Neuman, MD, MSc; Jennifer N. Goldstein, MD; Michael A. Cirullo, BS; J. Sanford Schwartz, MD

IMPORTANCE  Little is known regarding the durability of clinical practice guideline recommendations over time.

OBJECTIVE To characterize variations in the durability of class I ("procedure/treatment should be performed/administered") American College of Cardiology/American Heart Association (ACC/AHA) guideline recommendations.


MAIN OUTCOMES AND MEASURES We abstracted all class I recommendations from the first of the 2 most recent versions of each guideline and identified corresponding recommendations in the subsequent version. We classified recommendations replaced by less determinate or contrary recommendations as having been downgraded or reversed; we classified recommendations for which no corresponding item could be identified as having been omitted. We tested for differences in the durability of recommendations according to guideline topic and underlying level of evidence using bivariable hypothesis tests and conditional logistic regression.

RESULTS Of 619 index recommendations, 495 (80.0%; 95% CI, 76.6%-83.1%) were retained in the subsequent guideline version, 57 (9.2%; 95% CI, 7.0%-11.8%) were downgraded or reversed, and 67 (10.8%; 95% CI, 8.4%-13.3%) were omitted. The percentage of recommendations retained varied across guidelines from 15.4% (95% CI, 1.9%-45.4%) to 94.1% (95% CI, 80.3%-99.3%; P < .001). Among recommendations with available information on level of evidence, 90.5% (95% CI, 83.2%-95.3%) of recommendations supported by multiple randomized studies were retained, vs 81.0% (95% CI, 74.8%-86.3%) of recommendations supported by 1 randomized trial or observational data and 73.7% (95% CI, 65.8%-80.5%) of recommendations supported by opinion (P = .001). After accounting for guideline-level factors, the probability of being downgraded, reversed, or omitted was greater for recommendations based on opinion (odds ratio, 3.14; 95% CI, 1.69-5.85; P < .001) or on 1 trial or observational data (odds ratio, 3.49; 95% CI, 1.45-8.41; P = .005) vs recommendations based on multiple trials.

CONCLUSIONS AND RELEVANCE The durability of class I cardiology guideline recommendations for procedures and treatments promulgated by the ACC/AHA varied across individual guidelines and levels of evidence. Downgrades, reversals, and omissions were most common among recommendations not supported by multiple randomized studies.

Clinical practice guidelines are ubiquitous in medical care. As adherence to recommended practices increasingly is used to measure performance, guidelines play a major role in policy efforts to improve the quality and cost-effectiveness of care.2,3

In this context, understanding the durability of individual guideline recommendations over time is of importance to clinical practice and health policy. Past research has established the importance of revising guidelines over time to address advances in research and population-level changes in health risks.4,5 Nonetheless, unwarranted variability across guidelines can reduce trust in guideline processes and complicate efforts to promote consistent use of evidence-based practices.7,8 Moreover, policies based on recommendations that prematurely endorse practices subsequently found to be ineffective can lead to waste and potential harm.9-11

Although the US Institute of Medicine12 and others13 have made recommendations for improving guideline development processes, little is known regarding the degree to which individual guideline recommendations endure or change over time. We studied the durability of class I (“procedure/treatment should be performed/administered”) recommendations across serial versions of selected American College of Cardiology/American Heart Association (ACC/AHA) guidelines. We measured how often class I recommendations were downgraded to a less determinate status, reversed to recommend against a previously endorsed treatment, or omitted altogether from the subsequent guideline version. Next, we assessed the degree to which a recommendation’s likelihood of being downgraded, reversed, or omitted varied across guidelines and across recommendations supported by different levels of evidence. Finally, we conducted additional analyses to explore the extent to which downgrades in recommendations may have been related to the emergence of new research findings vs other factors.

Methods

The ACC and AHA have jointly produced guidelines since 1984.14 ACC/AHA guidelines are reviewed annually and periodically revised; however, before 2014 there was no specified interval after which revision of an ACC/AHA guideline was required. Since 1996,15 all ACC/AHA recommendations have been assigned to 1 of 4 classes, which have undergone only minor changes over time: class I, “procedure/treatment should be performed/administered”; class IIa, “it is reasonable to perform procedure/administer treatment”; class IIb, “procedure/treatment may be considered”; class III, “procedure/treatment should not be performed.”16

We reviewed ACC/AHA guidelines that were current as of September 1, 2013, and for which there was at least 1 prior version. To be consistent with past research,17 we excluded “focused updates” that are occasionally released between ACC/AHA guideline revisions to highlight interval changes to a limited number of recommendations. Our sample included 11 guidelines addressing atrial fibrillation18,19; perioperative cardiovascular evaluation20,21; cardiac pacemakers and antiarrhythmia devices22,23; secondary prevention of coronary artery disease24,25; coronary artery bypass graft surgery16,26; cardiovascular disease prevention in women27,28; heart failure29,30; percutaneous coronary intervention31-33; chronic stable angina34,35; unstable angina and non–ST-segment elevation myocardial infarction35,36; and valvular heart disease.37,38 A 12th guideline, on ST-segment elevation myocardial infarction,39,40 was excluded because of differences in the topics addressed between versions.

We obtained the full text of the 2 most recent complete versions of each guideline from past issues of Circulation and the Journal of the American College of Cardiology; when neither journal published the full guideline, we obtained it from another journal or via the web.

Data Abstraction and Coding

For each guideline, we considered the version immediately preceding the current one to be the index; this approach was chosen over other designs that would have incorporated earlier guideline versions for analytic simplicity and to focus our analysis on topics of relevance to current practice.

We abstracted all class I recommendations from each index guideline. We focused on class I recommendations because they are among the most definitive statements regarding practice in ACC/AHA guidelines. While the ACC/AHA does not consider all class I recommendations to be appropriate for use as quality measures, class I recommendations are considered to be more appropriate than class IIa or IIb recommendations as a potential basis for such measures.41 Recommendations appeared in each guideline as boldface, numbered items. Statements that had distinct levels of evidence assigned to them were classified as discrete recommendations.

Next, we reviewed the subsequent (current) guideline version for recommendations whose text corresponded to that of an index recommendation or used alternate language to address the same content. While we avoided extrapolations beyond the literal meaning of the guideline text, we did not require a one-to-one relationship between items in the index guideline and the subsequent version. When the content of 2 index recommendations was subsumed by 1 recommendation in the subsequent version, we considered both index recommendations to correspond to the same revised recommendation. Conversely, when 1 index recommendation appeared to have been split into 2 recommendations in the subsequent version, we considered both items to correspond to the same index recommendation.

We next assigned each index recommendation to 1 of 4 outcome categories based on the text of the revised guideline. We categorized a recommendation as having been “retained” if the revised guideline contained 1 or more class I recommendations that addressed the full content of the index recommendation, allowing for wording changes and changes in cutoffs based on physiologic parameters or laboratory values. We categorized a recommendation as having been “downgraded” if part or all of its content was replaced by a class IIa or class IIb recommendation. We categorized a recommendation as having been “reversed” if part or all of its content was replaced by a class III recommendation. In cases where a recommen-
ation appeared to have been reassigned from class III to class I for purely stylistic reasons (ie, without any change in its implications for practice), we considered that recommendation to have been retained. We categorized a recommendation as having been “omitted” if we were not able to locate any corresponding recommendation in the revised guideline.

Two reviewers (M.D.N., M.A.C.) independently coded all study outcomes; initial agreement on the classification of items was 94% (κ = 83.6). Next, 2 additional reviewers (J.N.G., J.S.S.) independently evaluated all outcomes. Reviewers were not blinded as to which guideline version was being analyzed. We resolved disagreements by consensus; a formal consensus methodology was not used.

**Independent Variables**

Prior to 2008, ACC/AHA guidelines did not routinely map individual recommendations to references in the medical literature. As such, we relied on the ACC/AHA’s levels-of-evidence designations, which were introduced gradually into ACC/AHA guidelines beginning in 1998, to summarize the type of scientific evidence underlying individual index recommendations. Level of evidence A includes data derived from multiple randomized clinical trials or meta-analyses; level of evidence B includes data derived from a single randomized trial or nonrandomized studies; level of evidence C includes consensus opinion, case studies, or “standard of care” as defined by the guideline committee.30

**Statistical Analyses**

We calculated the proportions of recommendations that were retained, downgraded, reversed, or omitted out of all index recommendations in our sample and within individual guidelines. We calculated exact confidence intervals for all proportions and used the χ² test to compare the proportions of recommendations that were retained, downgraded or reversed, or omitted across individual guidelines.

For those recommendations that had available data for level of evidence, we used the Fisher exact test to compare the proportion of index recommendations within each level of evidence that were retained, downgraded or reversed, or omitted; recommendations that had missing level-of-evidence data were excluded from these analyses. Among recommendations that had available (ie, nonmissing) level-of-evidence data, we also conducted a stratified, “within-guideline” analysis to test the association of a recommendation’s level of evidence with the probability of a downgrade, reversal, or omission while holding constant all guideline-level factors. This analysis used Stata’s clogit command to fit a conditional logistic regression model, grouped by the individual guideline, to predict a binary outcome that equaled 0 for all retained recommendations and 1 for all recommendations that were downgraded, reversed, or omitted, based on each index recommendation’s own listed level of evidence. This model used robust standard errors that adjusted for clustering at the guideline level.

Lastly, we explored potential reasons related to recommendation downgrades and reversals by using descriptive statistics to characterize changes in the reported level of evidence over time for all downgraded or reversed recommendations whose initial level of evidence was B or C. A downgrade or reversal that was accompanied by a transition to a higher level of evidence (ie, a transition from C to B or A or from B to A) could potentially have been prompted by the emergence of more definitive evidence. In contrast, a downgrade or reversal not accompanied by such a transition could potentially have been prompted by factors other than the emergence of more definitive evidence. We used a value of P < .05 to indicate statistical significance. All hypothesis tests were 2-sided. Analyses used Stata version 10.0 (StataCorp).

**Results**

We identified 619 class I recommendations in 11 index guidelines published between 1998 and 2007 (Table 1). The median number of years between the index guideline and the next full revision was 6 (range, 4-10). The number of listed writing committee members for index guidelines ranged from 11 to 33 (median, 14), and the percentage of members retained between versions ranged from 0% to 75.0% (median, 30.8%). The median number of class I recommendations per guideline was 41 (range, 13-136). Out of 619 index recommendations, 495 (80.0%; 95% CI, 76.6%-83.1%) were retained in the subsequent version; 57 (9.2%; 95% CI, 7.0%-11.8%) were downgraded (55 recommendations, 8.9%; 95% CI, 6.8%-11.4%) or reversed (2 recommendations, 0.3%; 95% CI, 0.04%-1.2%). Table 2 includes selected examples; all downgraded or reversed recommendations appear in eTable 1 in the Supplement. Sixty-seven recommendations (10.8%; 95% CI, 8.4%-13.3%) were omitted across guideline versions; Table 3 includes selected examples; all omitted recommendations appear in eTable 2 in the Supplement. Within individual guidelines, the median percentage retained was 82.6% (range, 15.4%-94.1%); the median percentage downgraded or reversed was 9.8% (range, 2.9%-15.4%) and the median percentage omitted was 4.2% (range, 0%-69.2%; P < .001).

Level-of-evidence data were available for 448 of 619 index recommendations (72.4%). These data were not provided for recommendations in the 1998 valvular heart disease guideline37 or the 2002 perioperative evaluation guideline.30 These guidelines accounted for 169 of the 171 recommendations for which level-of-evidence data were unavailable; the remaining 2 came from the 2005 heart failure guideline.29

The durability of individual recommendations varied according to their underlying level of evidence (Table 4). Among the 448 index class I recommendations for which level-of-evidence data were available, 90.5% (95% CI, 83.2%-95.3%; 95/105) of recommendations that were supported by multiple trials (ie, level of evidence A) were retained in the subsequent version vs 81.0% (95% CI, 74.8%-86.3%; 158/195) of those supported by a single trial or observational data (level of evidence B) and 73.7% (95% CI, 65.8%-80.5%; 109/148) of those supported by expert opinion (level of evidence C; P = .001). Downgrades or reversals were most common among level B recommendations, occurring in 12.8% (95% CI, 8.5%-18.3%; 25/195); omissions were most common among level C recommend-
dations, occurring in 16.9% (95% CI, 11.2%-23.9%; 25/148). After accounting for guideline-level factors, the combined odds of being downgraded, reversed, or omitted was greater among level B recommendations (odds ratio, 3.14; 95% CI, 1.69-5.85; P < .001) and level C recommendations (odds ratio, 3.49; 95% CI, 1.45-5.81; P = .005) compared with level A recommendations (Table 5).

Finally, we assessed changes over time in the level of evidence of each downgraded or reversed recommendation whose initial level of evidence was B or C. Among the 39 recommendations that met these criteria, the level of evidence increased across versions for 8 (20.5%) and decreased or stayed the same for 31 (79.5%).

Discussion

The durability of class I ACC/AHA guideline recommendations for procedures and treatments varied significantly across individual guidelines and levels of evidence, with recommendations that were based on multiple clinical trials being the most likely to endure over time. Of 619 recommendations drawn from 11 ACC/AHA guidelines published between 1998 and 2007, 80% of recommendations were retained at the time of the next guideline revision; while less than 1% were reversed, 9% were downgraded to a less determinate status, and 11% were omitted. After accounting for guideline-level factors, the odds of a downgrade, reversal, or omission were more than 3 times greater for recommendations based on a single trial, observational data, consensus opinion, or standard of care than for recommendations based on multiple randomized trials.

Increases in the level of evidence were uncommon among recommendations that were downgraded or reversed. We hypothesize that the classification of many of these recommendations may have changed in response to the reevaluation of available research; alternately, these recommendations may have been downgraded in response to new research that was insufficient by itself to alter the reported level of evidence. As the membership of many guideline committees changed substantially over time, our results also may reflect variability in the grading of evidence between different groups of experts.

To our knowledge, this is the first study to empirically evaluate the durability over time of individual guideline recommendations. Past work has suggested that guidelines should be reassessed for validity every three to five years to incorporate new clinical evidence; our study extends this prior work by systematically quantifying how individual recommendations drawn from a sample of prominent cardiovascular disease guidelines actually changed over time. As such, our findings offer a novel quantitative validation of contemporary emphases on making randomized trial evidence the primary basis for guideline recommendations. Further, they provide a basis for future efforts to understand the relative effect of emergent clinical evidence vs social and organizational factors on changes in guidelines over time.

Our findings also offer practical insights related to the application of guideline recommendations to clinical care and health policy. While our results highlight the overall durability of cardiovascular disease guideline recommendations, they also emphasize that particular subsets of recommendations may be more fragile than others as a basis for changes in practice and policy. For example, 1 of 8 recommendations that was based on a single trial or observational data was either downgraded or reversed in the subsequent guideline version, vs 1 of 26 recommendations based on 2 or more randomized trials. Such variations in durability might relate to differences in the actual validity of recommendations across levels of evidence, differences in the availability of new research over time, or both. Nonetheless, such information may aid clinicians and policy makers in quantifying the potential risks of measuring...

Table 1. Durability of 619 Class I Recommendations Abstracted From 11 ACC/AHA Guidelines Across 2 Guideline Versions

<table>
<thead>
<tr>
<th>Guideline Topic</th>
<th>Year</th>
<th>Class I Recommendations, No.</th>
<th>Index Guideline</th>
<th>Status of Recommendation in Revised (Current) Guideline*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>2001</td>
<td>46</td>
<td>2005</td>
<td>Retained, No. (%), Downgraded or Reversed, No. (%), Omitted, No. (%)</td>
</tr>
<tr>
<td>Preoperative cardiovascular evaluation</td>
<td>2002</td>
<td>13</td>
<td>2007</td>
<td>38 (82.6), 7 (15.2), 1 (2.2)</td>
</tr>
<tr>
<td>Pacemakers and implantable rhythm-management devices</td>
<td>2002</td>
<td>34</td>
<td>2008</td>
<td>2 (15.4), 2 (15.4), 9 (69.2)</td>
</tr>
<tr>
<td>Chronic stable angina</td>
<td>2002</td>
<td>79</td>
<td>2012</td>
<td>59 (74.7), 10 (12.7), 10 (12.7)</td>
</tr>
<tr>
<td>Secondary prevention of ischemic heart disease</td>
<td>2006</td>
<td>38</td>
<td>2011</td>
<td>35 (92.1), 3 (7.9), 0</td>
</tr>
<tr>
<td>Coronary artery bypass graft procedures</td>
<td>2004</td>
<td>39</td>
<td>2011</td>
<td>32 (82.1), 4 (10.3), 3 (7.7)</td>
</tr>
<tr>
<td>Percutaneous coronary intervention</td>
<td>2005</td>
<td>41</td>
<td>2011</td>
<td>36 (87.8), 4 (9.8), 1 (2.4)</td>
</tr>
<tr>
<td>Management of coronary artery disease in women</td>
<td>2007</td>
<td>24</td>
<td>2011</td>
<td>20 (79.2), 3 (12.5), 1 (4.2)</td>
</tr>
<tr>
<td>Unstable angina/non-ST-segment elevation myocardial infarction</td>
<td>2002</td>
<td>83</td>
<td>2007</td>
<td>75 (90.4), 6 (7.2), 2 (2.4)</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>2005</td>
<td>66</td>
<td>2013</td>
<td>36 (54.6), 5 (7.6), 25 (37.9)</td>
</tr>
<tr>
<td>Valvular heart disease</td>
<td>1998</td>
<td>156</td>
<td>2006</td>
<td>130 (86.3), 12 (7.7), 14 (9.0)</td>
</tr>
<tr>
<td>All topics</td>
<td></td>
<td>619</td>
<td></td>
<td>495 (80.0), 57 (9.2), 67 (10.8)</td>
</tr>
</tbody>
</table>

Abbreviations: ACC/AHA, American College of Cardiology/American Heart Association.

* P < .001 for difference in percentages retained, downgraded/reversed, or omitted across guidelines.
<table>
<thead>
<tr>
<th>Index Guideline</th>
<th>Index Recommendation</th>
<th>Class and LOE</th>
<th>Revised (Current) Guideline</th>
<th>Corresponding Revised (Current) Recommendations</th>
<th>Class and LOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001: Guidelines for the management of patients with atrial fibrillation</td>
<td>Screening for the presence of thrombus in the left atrium or left atrial appendage by transesophageal echocardiography is an alternative to routine postanesthesia... for cardioversion of AF.</td>
<td>Class I, LOE B</td>
<td>2006: Guidelines for the management of patients with atrial fibrillation</td>
<td>As an alternative to cardioversion prior to cardioversion of AF, it is reasonable to perform transesophageal echocardiography in search of thrombus in the left atrium or left atrial appendage.</td>
<td>Class IIa, LOE B</td>
</tr>
<tr>
<td>2002: Guideline update on perioperative cardiovascular evaluation for noncardiac surgery</td>
<td>Preoperative noninvasive evaluation of LV function: patients with current or poorly controlled HF. (If previous evaluation has documented severe left ventricular dysfunction, repeat preoperative testing may not be necessary.)</td>
<td>Class I, LOE not provided</td>
<td>2007 Guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery</td>
<td>Preoperative noninvasive evaluation of LV function: ...is reasonable for patients with worsening dyspnea or other change in clinical status... if not performed within 12 mo.</td>
<td>Class IIa, LOE C</td>
</tr>
<tr>
<td>2002: Guideline update for implantation of cardiac pacemakers and antiarrhythmia devices</td>
<td>Implantable cardioverter-defibrillator therapy: spontaneous sustained VT in patients without structural heart disease not amenable to other treatments.</td>
<td>Class I, LOE C</td>
<td>2008: Guidelines for device-based therapy of cardiac rhythm abnormalities</td>
<td>Implantable cardioverter-defibrillators: ICD implantation is reasonable for patients with sustained VT and normal or near-normal ventricular function.</td>
<td>Class IIa, LOE C</td>
</tr>
<tr>
<td>2006: Guidelines for secondary prevention for patients with coronary and other atherosclerotic vascular disease</td>
<td>ACE inhibitors: consider for all other patients [those without left ventricular ejection fraction &lt;40% and those without hypertension, diabetes, or chronic kidney disease].</td>
<td>Class I, LOE B</td>
<td>2011: Secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease</td>
<td>ACE inhibitors: It is reasonable to use ACE inhibitors in all other patients [those without left ventricular ejection fraction &lt;40% and those without hypertension, diabetes, or chronic kidney disease].</td>
<td>Class IIa, LOE B</td>
</tr>
<tr>
<td>2004: Guideline update for coronary artery bypass graft surgery</td>
<td>CABG is recommended in patients with stable angina who have 2-vessel disease with significant proximal LAD stenosis and either EF less than 0.50 or demonstrable ischemia on noninvasive testing.</td>
<td>Class I, LOE A</td>
<td>2011: Guideline for coronary artery bypass graft surgery</td>
<td>CABG to improve survival is reasonable in patients with mild-to-moderate LV systolic dysfunction (EF 35% to 50%) and significant... multivessel CAD or proximal LAD coronary artery stenosis, when viable myocardium is present.</td>
<td>Class IIa, LOE B</td>
</tr>
<tr>
<td>2007: Evidence-based guidelines for cardiovascular disease prevention in women</td>
<td>Lifestyle and pharmacotherapy should be used as indicated in women with diabetes to achieve an HbA1c &lt;7% if this can be accomplished without significant hypoglycemia.</td>
<td>Class I, LOE C</td>
<td>2011: Evidence-based guidelines for cardiovascular disease prevention in women</td>
<td>Lifestyle and pharmacotherapy can be used in women with diabetes mellitus to achieve an HbA1c &lt;7% if this can be accomplished without significant hypoglycemia.</td>
<td>Class IIa, LOE B</td>
</tr>
<tr>
<td>2005: Guideline update for the diagnosis and management of chronic heart failure in the adult</td>
<td>Coronary arteriography should be performed in patients presenting with HF who have angina or significant ischemia unless the patient is not eligible for revascularization....</td>
<td>Class I, LOE B</td>
<td>2013: Guidelines for the management of heart failure</td>
<td>When ischemia may be contributing to HF, coronary arteriography is reasonable for patients eligible for revascularization.</td>
<td>Class IIa, LOE C</td>
</tr>
<tr>
<td>2005: Guidelines on percutaneous coronary intervention</td>
<td>PCI after successful fibrinolysis: patients whose anatomy is suitable, PCI should be performed when there is objective evidence of recurrent MI.</td>
<td>Class I, LOE C</td>
<td>2011: Guideline for percutaneous coronary intervention</td>
<td>PCI is reasonable in patients with STEMI and clinical evidence for fibrinolytic failure or infarct artery reocclusion.</td>
<td>Class IIa, LOE B</td>
</tr>
<tr>
<td>2002: Guideline update for the management of patients with unstable angina and non-ST-segment myocardial infarction</td>
<td>Morphine sulfate intravenously [is recommended] when symptoms are not immediately relieved with nitroglycerin or when acute pulmonary congestion and/or severe agitation is present.</td>
<td>Class I, LOE C</td>
<td>2007: Guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction</td>
<td>In the absence of contraindications to its use, it is reasonable to administer morphine sulfate intravenously to UA/NSTEMI patients if there is uncontrolled ischemic chest discomfort despite nitroglycerin....</td>
<td>Class IIa, LOE B</td>
</tr>
</tbody>
</table>

Abbreviations: ACC/AHA, American College of Cardiology/American Heart Association; ACE, angiotensin-converting enzyme; AF, atrial fibrillation; AR, aortic regurgitation; CABG, coronary artery bypass graft procedure; CAD, coronary artery disease; EF, ejection fraction; HbA1c, hemoglobin A1c; HF, heart failure; ICD, implantable cardioverter-defibrillator; LAD, left anterior descending artery; LOE, level of evidence; LV, left ventricle; MI, myocardial infarction; MPI, myocardial perfusion imaging; PCI, percutaneous coronary intervention; SIHD, stable ischemic heart disease; STEMI, ST-segment elevation myocardial infarction; UA/NSTEMI, unstable angina/non-ST-segment elevation myocardial infarction; VT, ventricular tachycardia.

* Recommendations have been edited for length; see eTable 1 in the Supplement for full text.
physician performance based on adherence to recommendations derived from limited clinical evidence.

One of 9 ACC/AHA class I recommendations in our sample was omitted across guideline versions. Guideline texts rarely stated the reasons for these omissions, which may have related to changes in the prevalence of specific medical conditions, changes in clinical evidence or opinion related to the risks and benefits of particular interventions, or changes in the perceived relevance of a topic to the scope of a given guideline. As each of these potential reasons for omission carries distinct implications for practice, our findings stress the importance of communication on the part of guideline-producing bodies regarding the reasons that specific recommendations are removed from guidelines, as well as any changes in practice that might be implied by their removal.

This work has limitations. The guidelines we examined came from 1 organization, focused on cardiovascular diseases, and primarily addressed procedures and treatments. It
is possible that an analysis of guidelines that were produced by other organizations, that focused on other areas of medicine, or that dealt with other aspects of care might yield different findings. Level-of-evidence data were not available for 28% of the recommendations reviewed here, potentially limiting the generalizability of our findings to the topics addressed by those recommendations. The available data do not allow us to quantify the health consequences of adherence to guideline recommendations that were reversed, or changes in practice that may have resulted from downgrades in recommendations, several of which may have related to subtle but potentially important changes in emphasis regarding treatments' benefits and harms. Further, the ACC/AHA's guideline development process has undergone changes since the release of the Institute of Medicine's 2011 report on the development of trustworthy clinical practice guidelines. The guidelines that were available to us for review did not permit us to assess the durability of recommendations whose development incorporated independent, systematic evidence reviews as recommended in this report. However, our work may nonetheless serve as a basis for future research.

In addition, nearly all of the recommendations we identified underwent some degree of change over time. While many of these changes involved minor grammar or wording revisions, others involved more substantial changes in content or scope. As a result, our efforts to match recommendations across guideline versions necessitated some degree of interpretation. Nonetheless, we aimed throughout to adhere as closely as possible to the literal meaning of the guideline text and sought to limit bias by validating our outcome coding across multiple reviewers.

Despite these limitations, our results may have important implications for health policy and medical practice. The categorization of medical evidence, through guidelines, into stronger and weaker recommendations, influences definitions of good medical practice and informs efforts to measure the quality of care on a large scale. Our findings stress the need for frequent reevaluation of practices and policies based on guideline recommendations, particularly in cases where such recommendations rely primarily on expert opinion or limited clinical evidence. Moreover, our results suggest that the effectiveness of clinical practice guidelines as a mechanism for quality improvement may be aided by systematically identifying and reducing unwarranted variability in recommendations. Finally, our work emphasizes the importance of greater efforts on the part of guideline-producing organizations to communicate the reasons that specific recommendations are downgraded, reversed, or omitted over time.

Conclusions

The durability of class I cardiology guideline recommendations for procedures and treatments promulgated by the ACC/AHA varied across individual guidelines and levels of evidence. Downgrades, reversals, and omissions were most common among recommendations not supported by multiple randomized studies.

### Table 4. Durability Class I ACC/AHA Guideline Recommendations With Differing Levels of Underlying Scientific Evidence Among 448 Index Recommendations for Which Level-of-Evidence Data Were Available

<table>
<thead>
<tr>
<th>Status of recommendation in revised (current) guideline, No. (%)</th>
<th>Level of Evidence A: Multiple Randomized Clinical Trials or Meta-analyses</th>
<th>Level of Evidence B: Single Randomized Trial or Nonrandomized Studies</th>
<th>Level of Evidence C: Consensus Opinion, Case Studies, or Standard of Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained</td>
<td>(n = 105)</td>
<td>(n = 159)</td>
<td>(n = 148)</td>
</tr>
<tr>
<td>Downgraded or reversed</td>
<td>95 (90.5)</td>
<td>158 (81.0)</td>
<td>109 (73.7)</td>
</tr>
<tr>
<td>Omitted</td>
<td>4 (3.8)</td>
<td>25 (12.8)</td>
<td>14 (9.5)</td>
</tr>
</tbody>
</table>

### Table 5. “Within-Guideline” Analysis: Relative Odds of a Downgrade, Reversal, or Omission According to Level of Evidence for 448 Index Recommendations That Had Available Level-of-Evidence Data, Holding Constant All Guideline-Level Factors

<table>
<thead>
<tr>
<th>Level of evidence A: multiple randomized clinical trials or meta-analyses</th>
<th>Odds Ratio for Downgrade, Reversal, or Omission (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>1 [Reference]</td>
<td></td>
</tr>
<tr>
<td>Level of evidence B: single randomized trial or nonrandomized studies</td>
<td>3.14 (1.69-5.85)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Level of evidence C: consensus opinion, case studies, or standard of care</td>
<td>3.49 (1.45-8.41)</td>
<td>.005</td>
</tr>
</tbody>
</table>

### Article Information

**Author Contributions:** Dr Neuman had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Neuman, Cirullo, Schwartz. Acquisition, analysis, or interpretation of data: Neuman, Goldstein, Cirullo. Drafting of the manuscript: Neuman, Schwartz. Critical revision of the manuscript for important intellectual content: Neuman, Goldstein, Cirullo, Schwartz. Statistical analysis: Neuman, Schwartz. Obtained funding: Neuman. Administrative, technical, or material support: Neuman, Cirullo, Schwartz. Study supervision: Neuman, Goldstein, Schwartz.

**Conflict of Interest Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr
Schwartz reported receiving personal fees from Allergan, Bayer, the Blue Cross and Blue Shield Associations, General Electric, and UCB and grants from Pfizer outside the submitted work during the 36 months prior to submission of the research manuscript; no funding from any of these sources was obtained for the present study. No other disclosures were reported.

**Funding/Support:** Dr Neuman received funding from the National Institute on Aging (K08AG043548-02).

**Role of the Sponsors:** The National Institute on Aging had no role in the design or conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Contributions:** Lee A. Fleisher, MD (University of Pennsylvania), Sankey Williams, MD (University of Pennsylvania), Brian Kavanagh, MB, BSc (University of Toronto), and Amir Qaseem, MD, PhD (American College of Physicians), provided input regarding study design and selection of guidelines for analysis. None of these individuals were compensated for their contributions. We also thank the editor and reviewers for their constructive comments and suggestions.

**REFERENCES**


27. Mosca L, Banica CL, Benjamin EJ, et al; Expert Panel/Writing Group; American Heart Association; American Academy of Family Physicians; American College of Obstetricians and Gynecologists;
American College of Cardiology Foundation; Society of Thoracic Surgeons; American Medical Women’s Association; Centers for Disease Control and Prevention; Office of Research on Women’s Health; Association of Black Cardiologists; American College of Physicians; World Heart Federation; National Heart, Lung, and Blood Institute; American College of Nurse Practitioners. Evidence-based guidelines for cardiovascular disease prevention in women: 2007 update. Circulation. 2007;115(11):1481-1501.


