Comparison of Clinical and Administrative Data Sources for Hospital Coronary Artery Bypass Graft Surgery Report Cards

David M. Shahian, MD; Treacy Silverstein, BS; Ann F. Lovett, RN, MA; Robert E. Wolf, MS; Sharon-Lise T. Normand, PhD

Background—Regardless of statistical methodology, public performance report cards must use the highest-quality validated data, preferably from a prospectively maintained clinical database. Using logistic regression and hierarchical models, we compared hospital cardiac surgery profiling results based on clinical data with those derived from contemporaneous administrative data.

Methods and Results—Fiscal year 2003 isolated coronary artery bypass grafting surgery results based on an audited and validated Massachusetts clinical registry were compared with those derived from a contemporaneous state administrative database, the latter using the inclusion/exclusion criteria and risk model of the Agency for Healthcare Research and Quality. There was a 27.4% disparity in isolated coronary artery bypass grafting surgery volume (4440 clinical, 5657 administrative), a 0.83% difference in observed in-hospital mortality (2.05% versus 2.88%), corresponding differences in risk-adjusted mortality calculated by various statistical methodologies, and 1 hospital classified as an outlier only with the administrative data–based approach. The discrepancies in volumes and risk-adjusted mortality were most notable for higher-volume programs that presumably perform a higher proportion of combined procedures that were misclassified as isolated coronary artery bypass grafting surgery in the administrative cohort. Subsequent analyses of a patient cohort common to both databases revealed the smoothing effect of hierarchical models, a 9% relative difference in mortality (2.21% versus 2.03%) resulting from nonstandardized mortality end points, and 1 hospital classified as an outlier using logistic regression but not using hierarchical regression.

Conclusions—Cardiac surgery report cards using administrative data are problematic compared with those derived from audited and validated clinical data, primarily because of case misclassification and nonstandardized end points. (Circulation. 2007;115:1518-1527.)

Key Words: coronary artery bypass surgery • databases • health policy • outcome assessment (health care) • statistics

Cardiac surgeons have consistently led the effort to collect, analyze, and apply outcomes data for the benefit of patients. Cardiac surgery also is the paradigm for public performance reporting.1–3 Its predominant procedure, coronary artery bypass grafting (CABG), is the most frequently performed of all complex surgical operations, has well-defined and measurable end points, and has substantial public health and cost implications.

Determining the optimal statistical methodology for risk adjustment and provider profiling has been controversial.1,3 No statistical technique, however, regardless of its sophistication, can compensate for flawed data, and it is this concern that motivates the present study. Healthcare data may be derived from administrative or clinical sources,4,5 the latter including both retrospective chart abstraction and prospectively maintained databases such as the Society of Thoracic Surgeons National Adult Cardiac Database. Administrative data, typically derived from discharge billing forms, are the most inexpensive and readily available source of information regarding acute care hospitalizations. Although not originally intended for this purpose, such data also have been used to assess healthcare provider performance and are currently the basis for some public “report cards.”

Using administrative data to evaluate the quality of provider performance is plausible when they are the sole available source of information and when it has been demonstrated that their use results in conclusions similar to those derived from clinical data.6 However, are public...
reports based on hospital discharge data an acceptable alternative when more accurate but also more costly and resource-intensive options are available, as in the case of CABG surgery.

The present study, prompted by recent experience in Massachusetts, investigates the usefulness of a national algorithm based on administrative data, compared with results derived from prospectively maintained clinical data, in assessing hospital CABG outcomes. Beginning in 2002, Massachusetts law mandated that data on all cardiac surgical procedures be collected using the Society of Thoracic Surgeons National Adult Cardiac Database. These data are submitted to the Massachusetts Data Analysis Center (Mass-DAC) based at the Department of Health Care Policy at Harvard Medical School, where they are subjected to internal and external audit and validation. Analyses are then performed using hierarchical logistic regression in which random intercepts are included. Analyses are then performed using hierarchical logistic regression in which random intercepts are included for each hospital. The first CABG report card was published in October 2004 and was based on calendar year 2002 data.

Because of its desire to bring information to the public more quickly than was possible with audited clinical data, the Massachusetts Executive Office of Health and Human Services published its own fiscal year 2004 CABG results on their Internet site in 2006. They were based on hospital discharge billing data collected by the Massachusetts Division of Health Care Finance and Policy using a national model proposed by the Agency for Healthcare Research and Quality (AHRQ). The Massachusetts Executive Office of Health and Human Services also published the calendar year 2003 Mass-DAC isolated CABG results. This afforded a side-by-side, although not strictly contemporaneous, comparison.

Notably, the CABG cohort based on AHRQ CABG criteria as applied to Massachusetts administrative data consisted of 5315 procedures for fiscal year 2004, whereas Mass-DAC recorded 4393 confirmed isolated CABG cases for calendar year 2003. This 21% discrepancy in volume (as high as 40% at 1 hospital) cast doubt on the accuracy of the AHRQ algorithm, particularly because the volume of isolated CABG surgery declined rather than increased in Massachusetts between 2003 and 2004.

Because we had access to both audited clinical data and administrative data for fiscal year 2003, we conducted a comprehensive, contemporaneous comparison of these 2 algorithms and the hospital quality profiles derived from them.

Methods

Creation of Cohorts
Two cohorts were created: 1 based on clinical data (Mass-DAC) and 1 that used hospital discharge billing data (AHRQ). The AHRQ CABG model includes patients on the basis of discharge dates; the Mass-DAC model includes patients on the basis of surgery dates. To obtain strictly contemporaneous groups, the Mass-DAC cohort for the present study was selected on the basis of discharge date, including patients who were discharged from a nongovernmental Massachusetts acute care hospital with a diagnosis of CABG surgery between October 1, 2002, and September 30, 2003.

Clinical Cohort (Mass-DAC)
Hospitals submit clinical information to Mass-DAC for all adult patients (age ≥18 years) having cardiac surgery in a Massachusetts non-US governmental hospital. Isolated CABG cases are identified by excluding cases that combine CABG with other significant procedures such as valve replacement or carotid endarterectomy. For certain procedures with problematic classification, additional exclusion criteria were used on the basis of clinical data submitted by the hospital. Cases coded by hospitals as “CABG+other” are reviewed by an adjudication committee of senior cardiac surgeons, and a determination is made as to whether the case is legitimately in the “other” category. This review seeks to mitigate gaming of the system in which cases with unfavorable outcomes are shifted into an unreported category.

Administrative Cohort (AHRQ)
Patients ≥40 years of age discharged with International Classification of Diseases, 9th revision, clinical modification (ICD-9-CM) codes 36.10 to 36.19 are selected. Thus, the total number of discharges invariably includes cases that are not isolated CABG such as CABG plus valve replacement. Discharges with missing discharge disposition or those involving transfer to a short-term general hospital are excluded.

The gold standard for this comparison is the audited database and the analytical methodology used by Mass-DAC.

Definition of Outcomes
Mass-DAC uses 30-day all-cause mortality as the primary patient end point. Hospital-reported 30-day status is verified by linking all cardiac surgery cases reported to the Massachusetts Registry of Vital Records and Statistics. The AHRQ model uses all-cause in-hospital mortality reported in hospital billing data.

The primary hospital end point is standardized mortality incidence rate (SMIR), interpreted as the mortality at each hospital standardized to the overall mortality experience in the population of hospitals under study. Confidence intervals are constructed when estimating logistic regression models, and probability intervals are constructed when estimating hierarchical logistic regression models. For ease of exposition, we use the terms interval estimates or confidence intervals interchangeably.

Risk Adjustment
Mass-DAC uses the Society of Thoracic Surgeons National Adult Cardiac Database instrument and definitions, as well as a hierarchical logistic regression model, to determine standardized mortality rates. Risk factors for the model were selected by expert consensus from the best existing clinical models. A hospital-specific random intercept was included for each hospital in the sample.

The AHRQ model uses standard logistic regression, and its risk factors include demographic information and factors based on All Patient Refined Diagnosis-Related Group codes.

Statistical Analyses
Comparison of Cardiac Surgery Volumes With CABG
Volumes and With Observed CABG Mortality Rates
CABG volumes and observed mortality rates were examined overall and by hospital using graphical methods. We plotted ratios of CABG volumes in the original AHRQ and Mass-DAC cohorts as a function of total cardiac surgery procedures to determine whether the disparities between cohorts varied by hospital program size. Total cardiac surgery volume was obtained from the Mass-DAC database. A similar plot was constructed to examine the relationship of the difference between observed in-hospital mortality in the original AHRQ and Mass-DAC cohorts as a function of total cardiac program volume. Scatterplot smoothers were constructed to illustrate these relationships.
Comparison of Discharges Included in the AHRQ and Mass-DAC Models

We identified the number of discharges meeting inclusion criteria of both approaches, the number included in the AHRQ but not the Mass-DAC model, and the number included in the Mass-DAC but not in the AHRQ model. The administrative data and Mass-DAC data were merged using a combination of medical record number, discharge date, admission data, and date of birth. When possible, we determined the specific reason that discharges were included in 1 model but not the other. Because we had in-hospital mortality for all patients, we also calculated the observed inpatient mortality rate by exclusion category.

Comparison of Hospital Standardized Mortality and Outlier Determination

In addition to misclassification errors, we hypothesized other potential reasons for differences in the AHRQ and Mass-DAC approaches, including risk models, mortality end points, and statistical methodology. To isolate differences not attributable to classification errors, additional comparative analyses were performed using a common cohort of patients who met both AHRQ and Mass-DAC inclusion criteria.

For this common cohort, we estimated the predicted risk of in-hospital mortality for each patient using simple logistic regression models and the model-specific risk predictors and determined the areas under the receiver-operating characteristics curves. Second, we computed point and 95% CI estimates of in-hospital SMIRs using AHRQ and Mass-DAC predictors, estimated by both standard and hierarchical logistic regression. The standardized rate is obtained by risk adjusting for the case mix of the hospital, indirectly standardizing its risk to an expected risk, and then multiplying by the state average. If the interval estimate for the hospital was below the state average, we concluded that mortality was lower than expected. If the interval estimate included the state average, then the mortality was as expected. If the interval estimate was above the state average, then mortality was higher than expected. Finally, we determined 30-day SMIRs using Mass-DAC predictors and both standard and hierarchical logistic regression. Models were fitted using the WinBUGS (Medical Research Council Biostatistics Unit, Cambridge, United Kingdom), SAS (SAS Institute, Cary, NC), and S-Plus (Insightful Corp, Seattle, Wash) software systems.

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

Results

CABG Volumes (Original Cohorts)

The fiscal year 2003 Mass-DAC cohort consisted of 4440 isolated CABG patients who were discharged between October 1, 2002, and September 30, 2003 (inclusive). In contrast, the contemporaneous AHRQ cohort consisted of 5657 patients identified as CABG patients using the criteria noted above, which represents a 27.4% disparity. Figure 1 (left) depicts the ratio of AHRQ CABG volume to Mass-DAC CABG volume on the y axis and hospital total cardiac surgery volume (Mass-DAC) on the x axis. Superimposed on both plots are scatterplot smoothers.
program). Higher-volume hospitals presumably perform a higher percentage of complex CABG cases (eg, CABG and valve, CABG and aneurysm repair) that are excluded from the Mass-DAC cohort but inappropriately included in the AHRQ cohort.

**Observed CABG Mortality (Original Cohorts)**
Overall observed in-hospital and 30-day mortality rates in the Mass-DAC cohort were 2.05% (91 of 4440) and 2.27% (101 of 4440). The in-hospital mortality observed in the AHRQ cohort was 2.88%, exceeding the corresponding Mass-DAC value by 0.83% (range, 0% to 1.63%). In-hospital mortality from the AHRQ-defined cohort is always higher than that from the Mass-DAC cohort, and the absolute difference increases directly with hospital total cardiac surgery volumes (Figure 1, right).

**Risk-Adjusted and Standardized Mortality (Original Cohorts and Methodologies)**
Figure 2 depicts hospital-specific 30-day SMIRs (95% probability intervals) using hierarchical logistic regression for the original Mass-DAC cohort (n=4440) and hospital-specific SMIR point estimates (95% CIs) using logistic regression for the original administrative cohort (n=5657). The AHRQ cohort has generally higher mortality rates (overall, 2.88% versus 2.27%). The smoothing effect of the Mass-DAC hierarchical model reduces interhospital variability, as demonstrated by the narrower 95% CIs compared with the AHRQ model. Hospital 14 had no in-hospital mortalities, and the AHRQ model–estimated SMIR is 0% (95% CI undefined), whereas the hierarchical model estimate is 2.30% (95% CI, 1.24 to 3.71). No hospital is identified as a statistical outlier based on the Mass-DAC algorithm, whereas hospital 4 is identified as having statistically higher-than-expected 30-day mortality using the AHRQ algorithm.

**Comparison of AHRQ and Mass-DAC Cohorts**
Figure 3 depicts 3 groups of patients: (1) 1264 patients found only in the AHRQ cohort having an in-hospital observed mortality of 5.93%, (2) 47 patients found only in the Mass-DAC cohort having an observed mortality of 4.25%; and (3) 4393 patients found in both cohorts, with 2.03% mortality in the Mass-DAC cohort (confirmed as accurate) and 2.00% in the AHRQ cohort. Most of the patients (42 of 47) classified in group 2 met the age criteria for Mass-DAC but not for AHRQ, had missing data, or had been transferred. Among the 1264 patients in group 1, about half (n=596) had CABG combined with valve procedures. This misclassification could be corrected with a more refined administrative algorithm. However, an additional 663 cases, ≈10% of the original AHRQ cohort, were coded as CABG plus other. Even with a clinical database and well-defined exclusion criteria, adjudication of such cases by an expert panel typically is required. It would be difficult, if not impossible, to correctly categorize such procedures (isolated versus nonisolated CABG) from administrative data using only discharge codes. This subset of CABG plus other cases was associated with high mortality (7.39%) and clearly had substantial impact on the overall AHRQ mortality estimate.

**Risk Model Differences (Common Cohort, N=4393)**
Tables 1 and 2 list risk factors, frequencies, and associations with in-hospital or 30-day mortality for 4393 patients meeting criteria for both the AHRQ and Mass-DAC models. For example, 2.0% of patients in the Mass-DAC cohort presented with cardiogenic shock, of which 23% died in-hospital and
19.5% died within 30 days. In-hospital mortality exceeds 30-day mortality in this and several other instances in Tables 1, 2 and 3 because the Society of Thoracic Surgeons definition of in-hospital mortality includes all in-hospital deaths, regardless of timing.

**TABLE 1. Distribution of Risk Factors and Relationship With Mortality for AHRQ Model Using 4393 Discharges in Common Cohort**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency, %</th>
<th>In-Hospital Mortality (2.00), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female age group, y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26–60</td>
<td>26.6</td>
<td>2.91</td>
</tr>
<tr>
<td>40–64</td>
<td>8.4</td>
<td>0.82</td>
</tr>
<tr>
<td>65–69</td>
<td>3.6</td>
<td>3.14</td>
</tr>
<tr>
<td>70–74</td>
<td>5.3</td>
<td>1.28</td>
</tr>
<tr>
<td>75–79</td>
<td>5.7</td>
<td>4.44</td>
</tr>
<tr>
<td>80–84</td>
<td>2.6</td>
<td>7.89</td>
</tr>
<tr>
<td>≥85</td>
<td>1.1</td>
<td>6.38</td>
</tr>
<tr>
<td>Age group (males and females), y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–64</td>
<td>41.2</td>
<td>0.77</td>
</tr>
<tr>
<td>65–69</td>
<td>13.8</td>
<td>1.64</td>
</tr>
<tr>
<td>70–74</td>
<td>18.0</td>
<td>1.77</td>
</tr>
<tr>
<td>75–79</td>
<td>16.5</td>
<td>3.86</td>
</tr>
<tr>
<td>80–84</td>
<td>7.9</td>
<td>5.19</td>
</tr>
<tr>
<td>≥85</td>
<td>2.6</td>
<td>3.57</td>
</tr>
<tr>
<td>Cardiac valves, extreme risk</td>
<td>0.02</td>
<td>100</td>
</tr>
<tr>
<td>CABG, catheterization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major risk</td>
<td>21.8</td>
<td>1.15</td>
</tr>
<tr>
<td>Extreme risk</td>
<td>6.3</td>
<td>16.6</td>
</tr>
<tr>
<td>CABG, no catheterization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major risk</td>
<td>11.4</td>
<td>0.80</td>
</tr>
<tr>
<td>Extreme risk</td>
<td>2.2</td>
<td>13.3</td>
</tr>
</tbody>
</table>

*Mean mortality.

Comparisons of Hospital Risk-Standardized Mortalities (Common Cohort, N=4393)

Table 3 lists the individual hospital volumes and observed mortality rates for the 4393 patients common to both the AHRQ and Mass-DAC cohorts.

Figure 4 depicts hospital-specific SMIRs (with 95% posterior intervals) for the common cohort estimated by hierarchical logistic regression using Mass-DAC and AHRQ predictors. The 95% Mass-DAC intervals are generally larger compared with those based on AHRQ predictors. Figure 5 depicts a similar comparison using logistic regression. There were 3 hospitals with no in-hospital mortalities for which the logistic model produces SMIRs of 0%.

Figure 6 depicts the results for 30-day mortality using Mass-DAC predictors estimated using both hierarchical and standard logistic regression models. Overall, there was an absolute 0.18% (2.21% versus 2.03%) difference in average mortality, depending on whether in-hospital (Figure 5) or 30-day (Figure 6) mortality was used as the end point. This 9% relative difference in mortality suggests that choice of an appropriate, standardized end point is not a trivial concern. The smoothing effect of the Mass-DAC hierarchical model shrinks the estimates toward the state mean (especially for low-volume providers), reduces inter-hospital variability, and narrows the 95% intervals.

Overall, these results demonstrate substantially both higher average mortality at 30 days compared with in-hospital mortality and the shrinkage effect of the hierarchical logistic regression model on point and 95% interval estimates. Notably, using in-hospital mortality, hospital 5 was clearly an outlier using the AHRQ or Mass-DAC predictors, estimated via logistic regression. It was not an outlier using hierarchical regression and either set of predictors.

Comparative Risk Model Discrimination (Common Cohort, N=4393)

The Mass-DAC model using in-hospital mortality as an outcome had an area under the receiver-operating character-
AHRQ model had an area under the curve of 0.89 (not significantly different from the Mass-DAC model; P = 0.05) with observed mortalities in the highest and lowest risk deciles of 15.6% and 0.4%. Finally, the Mass-DAC model assigned 0.2% of the sample to a risk of dying; the AHRQ assigned no one to this high-risk group.

**Discussion**

These analyses demonstrate the substantial variability in CABG profiling results that occurs as a result of differences in data source, algorithms to identify isolated CABG, non-standardized end points, and different risk models and statistical methodologies. In the present study, it appears that the predominant factor is misclassification of isolated CABG cases using an algorithm based on administrative data. Many of these classification errors could not have been corrected even with more sophisticated software algorithms and would require case-by-case review. Because of these classification errors, many of these programs have their true isolated CABG volumes and mortality rates inflated by the inclusion of combined procedures that are inherently higher risk, and this may disproportionately affect higher-volume tertiary centers. Our analyses also demonstrate the smoothing effect of the hierarchical model, which compensates for small sample sizes and clustering. This is manifested graphically as a shift in adjusted mortality rates toward the state average for low-volume programs for which there is relatively little information.

When used as the basis for public report cards, outcomes data must be of the highest possible reliability and validity, a requirement emphasized by the Institute of Medicine and the American Heart Association. Criteria have been proposed for profiling initiatives that evaluate risk-adjusted outcomes in cardiac surgery. Clinical data registries generally incorporate these desirable features, whereas administrative databases do not.

Because clinical data are unavailable for many specialties and because of the emphasis currently being placed on outcomes transparency and public accountability, there has been renewed interest in using administrative data for these purposes. Administrative data originally were intended primarily for reimbursement, although there has been a resurgence of interest in their use for healthcare research (despite having been abandoned by the federal government in 1993 as the basis for publicly reporting Medicare hospital mortality and the American Heart Association). Recent reports suggest that experienced researchers can, with careful design and sophisticated statistical methodology, develop models based on administrative data for certain conditions that may be appropriate for profiling.

---

**TABLE 2. Distribution of Risk Factors and Relationship With Mortality for Mass-DAC Model Using 4393 Discharges in Common Cohort**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency, %</th>
<th>In-hospital (2.03)*</th>
<th>30-Day (2.21)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>26.6</td>
<td>3.01</td>
<td>3.52</td>
</tr>
<tr>
<td>Years over 65 y*</td>
<td>1.85*</td>
<td>6.96</td>
<td>6.67</td>
</tr>
<tr>
<td>Renal failure</td>
<td>7.01</td>
<td>6.49</td>
<td>6.49</td>
</tr>
<tr>
<td>Hypertension</td>
<td>78.3</td>
<td>2.15</td>
<td>2.38</td>
</tr>
<tr>
<td>PVD</td>
<td>17.5</td>
<td>3.51</td>
<td>3.38</td>
</tr>
<tr>
<td>Diabetes</td>
<td>37.7</td>
<td>1.81</td>
<td>2.47</td>
</tr>
<tr>
<td>History of CABG</td>
<td>3.12</td>
<td>3.65</td>
<td>4.38</td>
</tr>
<tr>
<td>History of PCI</td>
<td>18.3</td>
<td>2.24</td>
<td>2.86</td>
</tr>
<tr>
<td>EF, %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF &lt;30 or missing</td>
<td>13.2</td>
<td>3.45</td>
<td>3.11</td>
</tr>
<tr>
<td>EF 30–39</td>
<td>12.5</td>
<td>4.74</td>
<td>4.93</td>
</tr>
<tr>
<td>EF ≥40</td>
<td>74.4</td>
<td>1.32</td>
<td>1.59</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>49.7</td>
<td>1.19</td>
<td>1.56</td>
</tr>
<tr>
<td>≤6 h</td>
<td>8.22</td>
<td>8.33</td>
<td>8.33</td>
</tr>
<tr>
<td>7–24 h</td>
<td>2.03</td>
<td>7.87</td>
<td>6.74</td>
</tr>
<tr>
<td>1–7 d</td>
<td>23.9</td>
<td>3.57</td>
<td>3.17</td>
</tr>
<tr>
<td>8–21 d</td>
<td>5.35</td>
<td>3.83</td>
<td>4.26</td>
</tr>
<tr>
<td>&gt;21 d</td>
<td>19.2</td>
<td>0.95</td>
<td>1.43</td>
</tr>
<tr>
<td>Cardiogenic shock</td>
<td>2.0</td>
<td>23.0</td>
<td>19.5</td>
</tr>
<tr>
<td>Status of surgery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective</td>
<td>32.5</td>
<td>0.98</td>
<td>1.33</td>
</tr>
<tr>
<td>Urgent</td>
<td>64.3</td>
<td>2.09</td>
<td>2.27</td>
</tr>
<tr>
<td>Emergent</td>
<td>3.2</td>
<td>10.1</td>
<td>8.63</td>
</tr>
<tr>
<td>Salvage</td>
<td>0.10</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>Pre-op IABP</td>
<td>10.9</td>
<td>6.49</td>
<td>5.65</td>
</tr>
</tbody>
</table>

*PVD indicates peripheral vascular disease; PCI, percutaneous coronary intervention; EF, ejection fraction; and Pre-op IABP, preoperative intra-aortic balloon pump.

*Mean.

---

**TABLE 3. Hospital CABG Volumes and Observed Mortality Rates for 4393 Discharges in Common Cohort**

<table>
<thead>
<tr>
<th>Hospital, Volume, n</th>
<th>Mass-DAC, %</th>
<th>AHRQ In-Hospital OM</th>
<th>30-d OM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>524</td>
<td>15 (2.86)</td>
<td>16 (3.05)</td>
</tr>
<tr>
<td>2</td>
<td>346</td>
<td>4 (1.16)</td>
<td>5 (1.45)</td>
</tr>
<tr>
<td>3</td>
<td>385</td>
<td>8 (2.08)</td>
<td>9 (2.34)</td>
</tr>
<tr>
<td>4</td>
<td>365</td>
<td>10 (2.74)</td>
<td>13 (3.56)</td>
</tr>
<tr>
<td>5</td>
<td>345</td>
<td>12 (3.48)</td>
<td>10 (2.90)</td>
</tr>
<tr>
<td>6</td>
<td>95</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>7</td>
<td>548</td>
<td>12 (2.19)</td>
<td>10 (1.82)</td>
</tr>
<tr>
<td>8</td>
<td>353</td>
<td>5 (1.42)</td>
<td>6 (1.70)</td>
</tr>
<tr>
<td>9</td>
<td>181</td>
<td>3 (1.66)</td>
<td>4 (2.21)</td>
</tr>
<tr>
<td>10</td>
<td>664</td>
<td>13 (1.96)</td>
<td>13 (1.96)</td>
</tr>
<tr>
<td>11</td>
<td>281</td>
<td>5 (1.78)</td>
<td>8 (2.85)</td>
</tr>
<tr>
<td>12</td>
<td>129</td>
<td>0 (0)</td>
<td>1 (0.78)</td>
</tr>
<tr>
<td>13</td>
<td>150</td>
<td>2 (1.33)</td>
<td>2 (1.33)</td>
</tr>
<tr>
<td>14</td>
<td>27</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>ALL</td>
<td>4393</td>
<td>89 (2.03)</td>
<td>97 (2.21)</td>
</tr>
</tbody>
</table>

Values are expressed as n (%) unless otherwise indicated. OM indicates observed mortality.
but experts have generally cautioned against this practice.\textsuperscript{4,12,18–29} Administrative data should be used to profile hospitals only after rigorous verification of their validity for such purposes, as was recently demonstrated for mortality after AMI\textsuperscript{16} and heart failure.\textsuperscript{17}

Concerns regarding administrative data sources for profiling include the following:

- Incorrect number of cases for analysis. This results from errors in diagnosis or procedure coding and from the use of software algorithms that cannot reliably identify isolated CABG cases.\textsuperscript{30}
- Restricted study populations. Performance results based on only 1 group (eg, Medicare patients), more common in studies based on administrative databases, are not necessarily representative of overall program quality and may bias the results of profiling.\textsuperscript{30,31}
- Nonstandardized mortality end points. Most administrative data registries record only in-hospital mortality, whereas clinical databases often record 30-day mortality. This may disadvantage providers having fewer options for early transfer to extended care facilities.\textsuperscript{32} The additional statistical and clinical implications of nonstandardized time intervals and the increasing ability to prolong time to death of critically ill postoperative CABG patients have recently been reviewed.\textsuperscript{33}
- Misalignment of data sources with intended use. Administrative databases were originally designed for claims sub-

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{SMIRs for common cohort (n=4393) estimated by hierarchical logistic regression using Mass-DAC predictors (solid lines) and AHRQ predictors (dashed lines).}
\end{figure}
mission and benefits coordination. Coding practices and software algorithms were developed for such reimbursement issues, not clinical outcomes profiling.23,34

- Coding accuracy and completeness. Administrative data coding accuracy may be problematic because of imprecise or ambiguous definitions4,12,23,25,31,35 or because the coding is not supported by the clinical record.36 In complex and lengthy hospitalizations, particularly if the patient ultimately dies, there is also a tendency to undercode chronic or asymptomatic comorbidities when using administrative data,28,34,35,37,38 with the result that these may paradoxically appear to be protective.26,35

- Absence of critical clinical variables. Much of the predictive power in surgical risk models is derived from a limited number of critical clinical variables that typically are not available in administrative databases.39,40 Addition of a few such clinical variables to administrative models may substantially improve their performance.25,35,41

- Complications and comorbidities. With administrative data, it is difficult to accurately map ICD-9-CM codes to clinically relevant comorbidities or complications, to consistently capture all complications and comorbidities, and to distinguish comorbidities from complications. In general, undercoding of comorbidities in administrative databases seems to be a much more common problem than overcoding, although either type of misclassification may occur despite the use of sophisticated coding algorithms.4,42–51

As a consequence of failing to differentiate complications from comorbidities, the performance of risk models based on administrative data may be exaggerated. This results from including predictors in the risk model that are actually late-hospitalization, preterminal events and thus highly predictive of subsequent mortality.4,14,19,29

Using an indicator for condition present at admission or date stamping of secondary diagnoses49 may improve differentiation of complications from comorbidities in administrative databases. This is currently the practice in at least 2 states, New York and California.4,52

- Outlier determination. Several studies have shown differences in the outliers determined using administrative versus clinical data, even when the administrative databases were relatively sophisticated.35,41,52–54

**Study Limitations**

In the present study, we used the actual AHRQ algorithm used by the Massachusetts Executive Office of Health and Human Services to generate CABG profiles for their Web site in 2006. Better administrative algorithms could certainly be developed, especially with regard to the correct identification of isolated CABG cases. The results from administrative-based profiling also could be enhanced by adding a few critical clinical variables and using more appropriate statistical techniques such as hierarchical models.

**Conclusions**

Government agencies and payers will be increasingly tempted to use administrative data for provider profiling because it is inexpensive and available within a short time frame. Given the inaccuracies demonstrated in our Massachusetts CABG analysis, this practice should be discouraged. Efforts should be made, as they are currently in Massachusetts, to reduce the lag time for producing more accurate reports based on clinical data.
Source of Funding
This work was supported by a contract from the Massachusetts Department of Public Health (No. 620022A4PRE).

Disclosures
Dr Shahian receives an honorarium for providing clinical advice to Mass-DAC. T. Silverstein, A. Lovett, R. Wolf, and Dr Normand receive salary support from a Massachusetts Department of Public Health contract.

References


---

**CLINICAL PERSPECTIVE**

Demand is increasing for public accountability in health care through the use of “report cards.” Numerous states have implemented coronary artery bypass grafting surgery (CABG) report cards. Hospitals are concerned that only the most accurate data and statistical methodologies are used for such purposes because their reported outcomes may substantially affect referral practices. Ideally, public report cards should be based on prospectively collected, audited, and validated clinical data, using the most appropriate risk-adjustment models and statistical approaches. The present study was motivated by the observation that results from a Massachusetts public CABG report card based on these principles differed substantially from those derived from a Massachusetts administrative database. Administrative data are widely accessible at low cost, and analyses based on them are generally available more quickly than from audited clinical registries. We determined that much of the difference between the 2 Massachusetts CABG report cards was attributable to case misclassification, typically with some combined procedures categorized as isolated CABG. Because combined operations have higher risk, their inclusion inflates the apparent mortality rate of a hospital. High-volume tertiary centers may be particularly disadvantaged by procedure classification errors because of their larger proportion of combined procedures. About half of the misclassified cases could not have been correctly classified through the use of more sophisticated coding algorithms and would require case-by-case adjudication. Administrative data are valuable for population-based studies, for quality improvement activities, and for profiling some diseases. When possible, however, audited data registries are preferable for public profiling of CABG surgery outcomes.
Comparison of Clinical and Administrative Data Sources for Hospital Coronary Artery Bypass Graft Surgery Report Cards
David M. Shahian, Treacy Silverstein, Ann F. Lovett, Robert E. Wolf and Sharon-Lise T. Normand

_Circulation_. 2007;115:1518-1527; originally published online March 12, 2007;
doi: 10.1161/CIRCULATIONAHA.106.633008
_Circulation_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2007 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/115/12/1518

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/