Effect of Provider Volume on the Accuracy of Hospital Report Cards
A Monte Carlo Study

Peter C. Austin, PhD; Mathew J. Reeves, PhD

Background—Hospital report cards, in which outcomes after the provision of medical or surgical care are compared across healthcare providers, are being published with increasing frequency. However, the accuracy of such comparisons is controversial, especially when case volumes are small. The objective was to determine the relationship between hospital case volume and the accuracy of hospital report cards.

Methods and Results—Monte Carlo simulations were used to examine the influence of hospital case volume on the accuracy of hospital report cards in a setting in which true hospital performance was known with certainty, and perfect risk-adjustment was feasible. The parameters used to generate the simulated data sets were obtained from empirical analyses of data on patients hospitalized with acute myocardial infarction in Ontario, Canada, in which the overall 30-day mortality rate was 11.1%. We found that provider volume had a strong effect on the accuracy of hospital report cards. However, provider volume had to be >300 before ≥70% of hospitals were correctly classified. Furthermore, hospital volume had to be >1000 before ≥80% of hospitals were correctly classified.

Conclusions—Producers and users of hospital report cards need to be aware that, even when perfect risk adjustment is possible, the accuracy of hospital report cards is, at best, modest for small to medium-sized case loads (ie, 100–300). Hospital report cards displayed high degrees of accuracy only when provider volumes exceeded the typical annual hospital case load for many cardiovascular conditions and procedures. (Circ Cardiovasc Qual Outcomes. 2014;7:299-305.)

Key Words: health services research | risk adjustment

Numerous cardiovascular report cards have been released comparing patient outcomes between different healthcare providers for cardiovascular conditions (eg, acute myocardial infarction or AMI) and for cardiovascular procedures (eg, coronary artery bypass graft or CABG surgery). Hospital-specific reports for mortality after admission for AMI have been released in California,1 Pennsylvania,2 Ontario, Canada,3 and in Scotland.4 Hospital and surgeon-specific mortality rates after CABG surgery have been published in Massachusetts,5 New Jersey,6 New York,7 and in Pennsylvania,8 while hospital-specific mortality rates have been published in Ontario.9

Depending on the medical condition or surgical procedure, the number of patients treated by a given physician or at a given hospital may range from a few dozen to several hundred. Frequently, only providers whose patient volume (number of patients treated or cared for) exceeds some specified threshold are included in published reports of provider outcomes. However, there are no specific or agreed upon guidelines for selecting this threshold. Furthermore, the choice involves inherent tradeoffs. A higher provider volume allows one to make more precise conclusions about the performance of a given healthcare provider, and it would be reasonable to assume that this would be accompanied by a commensurate increase in the accuracy of the resultant report card. However, setting a high volume threshold results in the exclusion low-volume providers, who tend to have worse outcomes compared with patients treated by high-volume providers.3-12 Consequently, there is a strong rationale for including low-volume providers in hospital report cards, and it is obviously important to be able to identify which low-volume providers have worse than expected outcomes. Regardless of the number of patients treated by a given healthcare provider, there remains an element of stochastic or random variability to a provider's observed performance because observed performance for a finite number of patients may differ, because of chance alone, from the provider's true long-term performance. Earlier simulation modeling work performed in the 1990s illustrated the pernicious effects of small case volumes and low event rates on the accuracy of hospital report cards.13,14 Even in the face of perfect risk-adjustment, which is necessary to account for differences in case-mix between hospitals, risk-adjusted mortality rates did a poor job of identifying outlier hospitals when case volumes and event rates were low; simulation studies showed that more false-positive results were generated than true positive results among...
WHAT IS KNOWN

• Hospital report cards are frequently produced.
• Some readers are concerned about the accuracy of hospital report cards.
• Previous studies, restricted to small case volumes, have demonstrated that hospital report cards can be subject to misclassification even when perfect risk-adjustment is possible.

WHAT THIS ARTICLE ADDS

• The accuracy of hospital report cards increases as the hospital’s case volume increases and as the mortality rate increases.
• To ensure a high degree of hospital outlier classification accuracy, hospital case volumes must be larger than typically observed for treatment of heart attacks.

Small hospitals. More recent studies have also compared the performance of provider profiling between low- and medium-volume settings, again illustrating the substantial problem of misclassification among small providers. Small case volumes are even more of a concern when profiling is undertaken at the individual physician level, especially for surgical procedures. However, despite these findings, many profiling studies continue to include providers with small case volumes; in fact, the Centers of Medicare and Medicaid Services-endorsed hospital profiling models for AMI and heart failure do not exclude hospitals based on case volumes, and so include providers with few cases (ie, provider volumes <10).

The objective of the current article was to examine the relationship between provider volume and the accuracy of hospital report cards, with the goal of illustrating the relationship between accuracy and hospital volume across the continuum of case volumes. An extensive series of Monte Carlo simulations were used to examine the issue. Monte Carlo simulations are necessary to ensure complete control of 2 critical components: (1) adequate risk adjustment to account for between-hospital differences in case-mix; and (2) that true hospital performance is known with certainty. To increase the face-validity of these simulations, the parameters of the Monte Carlo simulations were obtained from empirical analyses of health services data on hospitalized patients with AMI in Ontario, Canada.

Methods

We performed an extensive set of Monte Carlo simulations to examine the effect of hospital case volume on the accuracy of hospital report cards based on 30-day all-cause mortality. Parameters for the data-generating process were obtained from empirical analyses of data on patients hospitalized with AMI in Ontario, Canada. As mentioned above, Monte Carlo simulations were used because it is only in simulated data that perfect risk-adjustment can be performed to account for differences in case-mix and that the true performance of each hospital can be known. The simulations in the current study are similar to those used in a recently published study that examined the relationship between the c-statistic of the risk-adjustment model and the accuracy of hospital report cards.

Research ethics approval was received from the Research Ethics Board of Sunnybrook Health Sciences Center. Informed consent was waived, as permitted under Ontario privacy laws for research using administrative data.

Data Sources

Parameters for the Monte Carlo simulations were obtained from analyses of an administrative healthcare database containing information on all patients hospitalized in Ontario with a diagnosis of AMI between April 1, 1992 and March 31, 2012. These analyses were based on 31,186 admissions that occurred between April 1, 2008, and March 31, 2010, at 159 acute care hospitals. The 30-day all-cause mortality rate was 11.1% (N=34,722 patients died).

Parameters for the Monte Carlo Simulations

We generated 1 variable (denoting a patient risk-score) and estimated several parameters needed for the subsequent Monte Carlo simulations: (1) a risk-score determining a patient’s risk of death; (2) the distribution of this risk-score across the population (including both between- and within-hospital variation, thereby allowing for systematic differences in case-mix across hospitals); (3) the variation in risk-adjusted mortality across Ontario hospitals. Details of the analyses follow:

1. Individual patient-risk score for 30-day mortality: Using the previously derived Ontario AMI mortality prediction model that incorporates 11 variables (age, sex, congestive heart failure, cardiogenic shock, arrhythmia, pulmonary edema, diabetes mellitus with complications, stroke, acute renal disease, chronic renal disease, and malignancy), we estimated the predicted probability of 30-day mortality for each patient. The resulting linear predictor (ie, the log-odds of death) was then standardized to have mean zero and unit variance. This standardized linear predictor served as a risk-score in our subsequent empirical analyses.

2. Between-hospital and within-hospital variation in risk scores (ie, variation in case-mix): To determine the magnitude of between-hospital variation in case-mix, we used a variance components model to decompose the total variability in the standardized risk score (determined above) into between-hospital and within-hospital components: $X_i = \mu + \epsilon_i$, where $X_i =$ risk score of the patient at the $i$th hospital, $\mu =$ hospital-specific component for the $j$th hospital (ie, the mean risk score at the $j$th hospital because the risk score was standardized to have mean zero), and $\epsilon_i =$ patient-specific deviation from the mean risk-score for the $j$th hospital. This model assumes that the hospital-specific random effects and the patient-specific random effects are independent of one another. The estimated variance components were $\mu_j \sim N(0, \sigma^2 = 0.03696)$ and $\epsilon_i \sim N(0, \sigma^2 = 0.96358)$, indicating that only 3.7% of the variation in the patient-specific risk score was because of systematic variation in the mean risk-score between hospitals (ie, the intraclass correlation coefficient = 0.037).

3. Variation in risk-adjusted 30-day mortality across hospitals: We used a hierarchical logistic regression model to model the log-odds of death using the patient-specific risk score: logit($p_{ij} = \alpha_j + \beta X_{ij}$), where $p_{ij} =$ probability of death for the $i$th patient at the $j$th hospital, $\alpha_j =$ hospital-specific random effect for the $j$th hospital, and $X_{ij} =$ the standardized risk score for the $i$th hospital. This model allowed the log-odds of death to vary across hospitals; the following estimates were obtained: $\alpha_j \sim N(-2.5129, \sigma^2 = 0.05250)$ and $\beta = 1.1604$.

We then used these parameters in the subsequent Monte Carlo simulations to simulate synthetic data sets in which the distribution of patient risk was similar to that observed in Ontario. Furthermore, systematic differences in case-mix between hospitals and variation in hospital performance were also similar to that observed in Ontario.

Monte Carlo Simulations

We used the parameters obtained above in Monte Carlo simulations to generate artificial data sets, each consisting of $N_{patients}$ patients...
admitted to each of 100 hospitals, in which the distribution of case-mix across hospitals resembled that observed in Ontario (ie, intraclass correlation coefficient=0.037), and hospital performance was known with certainty ($\alpha_j$). For all 100 hospitals we generated the following:

1. Simulated patient-risk scores for patient $i$ at hospital $j$: To simulate patient-risk scores for the $i$th patient at the $j$th hospital for each of the N patients, we used the variance components model described above. We simulated a hospital-specific component: $\mu_j \sim N(0, \sigma^2 = 0.037)$ for each of the 100 hospitals, and a patient-specific component for each of the N patients: $\epsilon_{ij} \sim N(0, \sigma^2 = 0.963)$. The 2 simulated components were added together to produce a patient risk-score for the $i$th patient at the $j$th hospital.

2. Generated hospital-specific random effects for 30-day mortality: We then generated a hospital-specific random effect for 30-day mortality for each of the 100 hospitals using the hospital random effect parameters determined in the section Parameters for the Monte Carlo Simulations: $\alpha_{0j} \sim N(0, \sigma = 0.0525)$. This random effect served as the gold standard to define each hospital’s true performance. Because the $\alpha_j$ term for each hospital is known, the true ranking of each hospital is also known (in contrast, when using actual data, the true performance of each hospital is never known with certainty).

3. Generated 30-day mortality outcomes for each simulated subject: In the third and last step, we generated an outcome (death within 30-days of admission) for each subject in the simulated data set using the equation $logit(p_{ij}) = -2.513 + \alpha_{0j} + 1.1604x_i$, where $\alpha_{0j}$ is the randomly generated hospital-specific random effect, and $x_i$ is the randomly generated patient risk-score. The intercept ($-2.513$) was determined in the section Parameters for the Monte Carlo Simulations, and it represents the mean of the distribution of the hospital-specific random intercepts; similarly the regression coefficient 1.1604 was also obtained from the analysis described above. Finally, we randomly generated a dichotomous outcome for each subject in the synthetic data set from a Bernoulli distribution with subject-specific parameter $p_{ij}$.

To summarize, we simulated a risk-score and an outcome for each subject. The within- and between-hospital variations in this risk-score were similar to that observed in Ontario, as was the between-hospital variation in outcomes.

**Statistical Methods for Hospital Profiling**

Within each of the simulated data sets, we used 2 different statistical methods to assess hospital performance: model-based indirect standardization to compute observed-to-expected mortality ratios, and hierarchical logistic regression models to compute predicted-to-expected mortality ratios. These methods are described briefly below.

Model-based indirect standardization is commonly used in cardiovascular hospital report cards.23 A conventional logistic regression model was fit to each simulated data set, in which the dichotomous mortality outcome for each subject was regressed on each subject’s risk-score to generate the predicted probability of 30-day mortality. Note that the form of this model is different from that used in the data-generating process because it does not incorporate hospital-specific random effects. These predicted probabilities were summed up within each hospital to determine the expected number of deaths within that hospital based on its case-mix. The observed number of deaths at each hospital was divided by the expected number of deaths to generate the O-E (observed-to-expected) ratio for each hospital.

The predicted-expected (P-E) ratio is a modification of the above approach.23,24 A random intercept logistic regression model is used to model the relationship between the patient-specific risk-score and patient outcomes. The model includes a hospital-specific random effect, and it is therefore similar in form to that used in the data-generating process because it incorporates hospital-specific random effects. Using the fitted logistic regression model (including the hospital-specific random effects), the predicted probability of death is estimated for each patient. These predicted probabilities are summed up to obtain the predicted number of deaths at each hospital based on its case-mix. The hospital-specific random effects (or deviations from the average intercept) were then set to zero (this step is used so that outcomes at each hospital are assumed to be the same as those at an average hospital). Using this modified model, predicted probabilities of death are obtained for each subject, which are then summed up within each hospital to obtain the expected number of deaths, based on the case-mix of the hospital’s patients, if the hospital had the same performance as an average hospital. The ratio of these 2 quantities is the P-E ratio.

We highlight that the design of our Monte Carlo simulations allowed us to perform perfect risk-adjustment. Outcomes were simulated for each patient from a single, measured risk score. This same risk score was then used in the risk-adjustment model. Thus, there were no unmeasured differences in case-mix between hospitals.

Furthermore, the risk-adjustment model was, by design, correctly specified (ie, all variables related to the outcome were included in the risk-adjustment model). Thus, any observed inaccuracies in estimates of hospital performance were not because of inaccurate or incomplete risk-adjustment.

**Determining the Effect of Hospital Case Volumes on Accuracy of Hospital Profiling**

To understand the relationship between hospital case volume and the accuracy of hospital profiling, we performed a series of simulations in which we allowed the number of patients at each hospital ($N_{hospital}$) to take on values from 100 to 2000 in increments of 100. In each of these 20 scenarios, we simulated 500 random data sets. In each of these 500 simulated data sets for a given scenario, we determined the Spearman rank correlation coefficients between the true hospital-specific random effects ($\alpha_{0j}$) that were used as the gold standard of hospital performance (which were simulated during the data-generating process as described in the section Monte Carlo Simulations) and each of the O-E ratio, the P-E ratio, and the estimated hospital-specific random effects ($\hat{\alpha}_{0j}$). For each correlation, the mean correlation coefficient was determined across the 500 simulated data sets. A higher correlation indicates greater concordance (or accuracy) between the estimate of hospital performance (ie, O-E, P-E ratio, or estimated random effect) and the true hospital performance (ie, $\alpha_{0j}$).

We also classified hospitals into performance categories. In each of the 500 simulated data sets for each increment of hospital volume, we used the true hospital-specific random effects ($\alpha_{0j}$), which were simulated during the data-generating process, to categorize hospitals into 3 groups of known mortality performance: the top 20% of hospitals, the middle 60% of hospitals, and the bottom 20% of hospitals (the top performers have low mortality, whereas the bottom performers have high mortality). This classification is similar to one used in comparing risk-adjustment models for profiling hospitals’ performance in acute ischemic stroke.25 In each simulated data set, we also classified hospitals into categories according to their estimated performance using both the O-E ratio and the P-E ratio. As with the gold standard, we divided hospitals into the top 20%, the middle 60%, and the bottom 20% of hospitals based on the hospitals’ O-E and P-E ratios (the top performers will have low ratios; whereas the bottom performers will have high ratios). In each of the 500 simulated data sets, we determined the number of hospitals in each performance category (ie, that truly were low-outliers, high-outliers, or in the middle) that were classified as being low-outliers, high-outliers, or in the middle. These numbers were averaged across the 500 simulated data sets for each of the 20 scenarios.

**Sensitivity Analysis to Assess the Impact of a Lower Rate of All-Cause Mortality**

The Monte Carlo simulations described in the section Monte Carlo Simulations simulated synthetic data sets that resembled the observed population of patients with AMI hospitalized in Ontario, where the overall 30-day mortality rate was 11.1%. To assess the effect of having a lower mortality rate (typical of procedures such as CABG surgery), we undertook a sensitivity analysis, in which we repeated
the simulations after lowering the 30-day mortality rate for an average patient at an average hospital to 2% (a typical rate observed after CABG surgery). To do so, we changed the intercept in the random-effects outcome model from −2.513 to −3.89. An advantage to changing only the intercept in the outcome regression model is that all other factors (eg, between-hospital variation in case-mix and the relationship between the patient risk-score and outcomes) remain unchanged. Thus, we can examine the independent effect of the incidence rate of the outcome on the accuracy of hospital report cards. Had we used empirical data on patients with CABG, it is likely that all the factors (event rate, between-hospital variation in case mix, and between-hospital variation in outcomes) would differ, making it difficult to isolate the independent effect of event rate.

The statistical simulations were performed using the R statistical programming language (version 2.11.1; R Foundation for Statistical Computing, Vienna, Austria). The conventional logistic regression model was fit using the glm function, while the random effects logistic regression model was fit using the glmer function in the lme4 package version 0.999999.0, as this has been shown to perform well for estimating random effects models.26

Results

The relationship between hospital volume and the correlation between the true hospital-performance ($\alpha_j$; the known hospital-specific random effects obtained during the data-generating process) and each of the 3 different estimates of hospital performance (estimated hospital-specific random effect, the O-E ratio and the P-E ratio) is described in Figure 1. Results for the primary scenario (11% mortality rate) are depicted in the left image, while results for the sensitivity analysis, with a lower event rate, are depicted in the right image. The relationship between hospital volume and a given correlation coefficient was virtually identical for each of the 3 different measures of hospital performance. Correlation increased with increasing hospital volume in an approximately logarithmic relationship: increases in correlation diminished with increasing hospital volume. In the primary scenario, hospital volume had to be ≥500 for the correlation to exceed 0.8. We observed that, for hospital volumes representative of those seen in many provider-profiling studies (ie, ≤200), the correlation between observed and estimated random effects was moderate at best (ie, <0.7). A similar pattern, albeit with a lower magnitude of correlation, was observed in the secondary scenario with a lower event rate. For a given level of hospital volume, the correlation between true hospital performance and each measure of hospital performance was lower when the event rate was low compared with when it was high. For example, the hospital volume now had to be ≥1300 for the correlations to exceed 0.8.

We also note that the empirical SEs of the estimated correlation coefficients were related to hospital volume. In the primary analysis, when hospital volume was 100, the empirical SEs ranged from ≈0.075 to ≈0.082. When hospital volume was 2000, the empirical SEs ranged from ≈0.011 to ≈0.014.

The relationship between hospital volume and the mean number of hospitals that were classified in different categories is described in Figure 2. The top 2 images describe results in the setting with a high event rate (the primary scenario), whereas the bottom 2 images describe results in the setting with a low event rate (the sensitivity analysis). For each of the 2 scenarios, there are 2 images: one for the use of P-E ratios and one for the use of O-E ratios. As hospital volume increased, the number of hospitals that were correctly classified increased. However, even when hospital volume was equal to 2000, in the primary scenario, on average, only 85 of the 100 hospitals were correctly classified, regardless whether P-E or O-E ratios were used. Again, similar to the results shown in Figure 1, the number of hospitals classified correctly tended to be higher when the outcome was more common compared with when the outcome was rare. Even when

Figure 1. Relationship between hospital volume and the correlation between the true hospital–specific random effect and estimates of hospital performance.
hospital volume was high, a small number of hospitals were misclassified by one level (eg, a hospital that was truly in the middle 60% of hospitals was classified as being in the bottom 20% of hospitals). The number of such misclassifications was slightly higher when the outcome was rare compared with when it was more common.

Discussion

We used an extensive series of Monte Carlo simulations to examine the relationship between hospital volume and the accuracy of hospital report cards. The parameters of our Monte Carlo simulations were based on an analysis of hospitalized patients with AMI in Ontario, Canada. As a consequence of this, the synthetic data sets created in the Monte Carlo simulations were similar to those of a population-based sample of patients with AMI in a jurisdiction in which AMI report cards have been publicly released. It is important to note that our primary results were based on an overall case fatality rate of 11%, and the variability in mortality between hospitals was similar to that observed in the Ontario between 2008 and 2010. We found that the accuracy of hospital report cards increased with increasing provider volume. However, the correlation between the true performance of the hospital and different estimates of performance was high (ie, >0.80) only when hospital volumes were large (ie, >500). Our overall findings and conclusions were similar even when we reduced the outcome event rate substantially, to approximate that observed for cardiovascular surgical procedures such as CABG.

Criticism of provider profiling has often focused on whether risk-adjustment has adequately accounted for differences in case-mix between hospitals. However, it is important to note that our simulations were not subject to inadequate risk-adjustment. Outcomes were simulated so that they were related to a single risk score. The same risk score that was used to generate patient outcomes was then used in the risk-adjustment process in the simulated data sets. Thus, errors in hospital classification in our simulated data sets cannot be ascribed to residual confounding because of unmeasured case-mix differences. The observed lack of agreement between the different measures of hospital performance and the gold standard (each true random effect of the hospital) is therefore solely because of stochastic or random variability and not inadequate risk-adjustment. Our findings are consistent with previous studies that have demonstrated that even if perfect risk-adjustment was possible, random error will result in a substantial proportion of hospitals being misclassified.13,14,17,27 The current study expands on these previous studies by examining a much wider range of provider volumes, demonstrating that random variation can have a strong negative impact on the accuracy of provider profiling even at hospitals with large patient volumes (ie, >500), and demonstrating that the accuracy of profiling is reduced even further when the mortality rates are reduced to levels associated with common cardiovascular surgical procedures.

This current study was motivated by a previous study, which focused on determining the effect of the c-statistic on the accuracy of hospital report cards. In this earlier study, we considered 3 factors (number of patients per hospital, number of hospitals, and the c-statistic of the risk-adjustment model), and found that only the case volume had a strong influence on the accuracy of the resultant report...
card. However, we only modeled this effect across 3 different levels of hospital volume: 50, 100, and 200 patients per hospital. Motivated by these findings, we expanded our study of the effect of provider volume to have a more in-depth examination of the effect of provider volume on the accuracy of hospital report cards across a much larger range (100–2000). There are several factors that we did not vary in the current Monte Carlo simulations, and we were thus unable to examine their effect on the accuracy of hospital report cards. The decision to fix the c-statistic of the risk-adjustment model and the number hospitals, while varying the case load, was based on results from our previous.15 By changing the value of β in the risk-adjustment model, we could have varied the c-statistic of the risk-adjustment model.28 However, our earlier study found that neither the c-statistic of the risk-adjustment model nor the number of hospitals included in the simulation studies had a meaningful impact on the accuracy of hospital report cards in simulations based on Ontario AMI data.15 Thus, we elected to use the value of β that was observed in the Ontario AMI data, so that the risk-adjustment model would have a fixed c-statistic (0.787) comparable to that of the Ontario AMI mortality prediction model. Similarly, we also decided to fix the number of hospitals because this factor was shown to have a minimal impact on the accuracy of hospital report cards.15

There are a few limitations and caveats to our analysis to consider. First, the volume effects reported here in the primary analysis are based on using 30-day all-cause mortality after AMI admission. The 30-day mortality rate in our empirical Ontario data set, on which our simulations were based, was 11.1%. In a sensitivity analysis, we modified the data-generating process to simulate data in which the outcome was rarer (a mortality rate of 2% for an average patient at an average hospital). In these secondary analyses, we used an outcome mortality rate that was comparable with that observed after CABG surgery. We found minor decreases in accuracy as the outcome event rate decreased. Although we anticipate that comparable results will be observed, our simulations merit replication in other disease contexts and using parameters informed by empirical analyses in other jurisdictions. The current study provides a methodological framework in which these subsequent analyses can be performed.

Some authors have argued against excluding hospitals with small volumes from studies of provider profiling when using hierarchical models because the estimation of the hospital random effect from the hierarchical model takes the sample size of each unit into account.19 In general, hospitals that provide little information (have small case volumes) will have predicted risk-standardized mortality rates that are near the overall sample average because these institutions do not provide sufficient information for an informed estimate of their performance (this is because, the degree of shrinkage of the hospital-specific random effects toward the mean of the distribution is inversely related to hospital volume). Moreover, no matter what analytic approach is adopted, it is difficult to calculate precise estimates of risk-adjusted outcome rates for hospitals or other healthcare units with small volumes. The current study confirms that it is difficult to accurately estimate provider-specific performance when hospital volumes are small. A novel contribution of the current study is the demonstration that hospital volumes must exceed those typically seen for many conditions and procedures in order for hospital report cards to have a high degree of accuracy. Given the volume–outcome relationship that has been demonstrated across a wide range of medical conditions and surgical procedures, small providers are more likely to be those that require identification and remedial action.9–12 The current study highlights the difficulty in identifying those hospitals that are most likely to provide suboptimal quality of care.

We have demonstrated that even when provider volumes are high, it can be expected that a number of hospitals will be incorrectly classified when hospitals are stratified into performance categories. These findings should not be taken to support the argument that provider profiling should be abandoned. Instead, investigators and policy-makers need to interpret the findings of hospital report cards with caution, with the understanding that what is being reported are estimates, with an associated degree of uncertainty, about hospital performance. This issue will be of particular concern when profiling small to medium-sized institutions. Reports of risk-adjusted mortality can be seen as the first step in a quality-improvement process. In depth investigations can subsequently be performed at individual hospitals to identify hospital-specific explanations for either their exemplary performance or for their poor performance. These examinations could focus on process-of-care and structural measures of performance (eg, standardized discharge orders). Furthermore, methods for assessing the longitudinal performance of hospitals over time could also be implemented to examine the trajectory of performance at each hospital over time.29 In the current study, we focused on estimation and categorizing performers based on estimates of performance. We did not incorporate statistical significance testing into hospital classifications. Finally, Bayesian methods can be used in provider profiling to assess the magnitude of the evidence that the performance of a given hospital exceeds specified thresholds for defining quality of care.7,30–33

In conclusion, the accuracy of healthcare provider report cards increases with increasing provider volume. High levels of accuracy were only observed when provider volume exceeded 300 cases. Moderate correlation between true and estimated performance was observed for hospital volumes reflective of those observed in many clinical contexts.

Acknowledgments
This study was supported by the Institute for Clinical Evaluative Sciences (ICES), which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The opinions, results, and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOHLTC is intended or should be inferred. The data sets used were held securely in a linked, deidentified form and analyzed at the Institute for Clinical Evaluative Sciences.

Sources of Funding
This research was supported by operating grant from the Canadian Institutes of Health Research (CIHR; MOP 86508). Dr Austin is supported in part by a Career Investigator award from the Heart and Stroke Foundation.
Disclosures

None.

References


Effect of Provider Volume on the Accuracy of Hospital Report Cards: A Monte Carlo Study
Peter C. Austin and Mathew J. Reeves

Circ Cardiovasc Qual Outcomes. 2014;7:299-305; originally published online March 11, 2014;
doi: 10.1161/CIRCOUTCOMES.113.000685
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272
Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circoutcomes.ahajournals.org/content/7/2/299

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes 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: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/