

# Estimating the Net Benefit of Improvements in Hospital Performance

## *G-Computation With Hierarchical Regression Models*

Peter C. Austin, PhD\*†‡ and Douglas S. Lee, MD, PhD\*†§||

**Background:** It is important to be able to estimate the anticipated net population benefit if the performance of hospitals is improved to specific standards.

**Objective:** The objective of this study was to show how G-computation can be used with random effects logistic regression models to estimate the absolute reduction in the number of adverse events if the performance of some hospitals within a region was improved to meet specific standards.

**Research Design:** A retrospective cohort study using health care administrative data.

**Subjects:** Patients hospitalized with acute myocardial infarction in the province of Ontario in 2015.

**Results:** Of 18,067 patients hospitalized at 97 hospitals, 1441 (8.0%) died within 30 days of hospital admission. If the performance of the 25% of hospitals with the worst performance had their performance changed to equal that of the 75th percentile of hospital performance, 3.5 deaths within 30 days would be avoided [95% confidence interval (CI): 0.4–26.5]. If the performance of those hospitals whose performance was worse than that of an average

hospital had their performance changed to that of an average hospital, 6.0 deaths would be avoided (95% CI: 0.7–47.0). If the performance of the 75% of hospitals with the worst performance had their performance changed to equal that of the 25th percentile of hospital performance, 11.0 deaths would be avoided (95% CI: 1.2–79.0).

**Conclusion:** G-computation can be used to estimate the net population reduction in the number of adverse events if the performance of hospitals was improved to specific standards.

**Key Words:** hospital performance, multilevel models, hierarchical regression models, G-computation, health services research

(*Med Care* 2020;58: 651–657)

Hospital report cards are reports in which outcomes or processes of care are compared across hospitals for patients treated for the same medical condition or undergoing the same surgical procedure. Hospital report cards are used to monitor the quality of care across hospitals. The American states of New York, Pennsylvania, Massachusetts, New Jersey, as well as the Canadian province of Ontario have reported publicly on hospital performance for patients undergoing coronary artery bypass graft (CABG) surgery.<sup>1–5</sup> Similarly, Pennsylvania, California, and Ontario have publicly reported on hospital performance for patients hospitalized with acute myocardial infarction (AMI).<sup>6–8</sup> The HospitalCompare website produced by Medicare.gov reports on hospital-specific risk-adjusted 30-day mortality rates for patients hospitalized with AMI, heart failure, and pneumonia and for those undergoing CABG surgery ([www.medicare.gov/HospitalCompare/Data/Death-rates.html](http://www.medicare.gov/HospitalCompare/Data/Death-rates.html)). Hospital report cards permit the identification of health care providers that provide quality of care that is significantly above or below average. Quality improvement interventions can be targeted at providers with worse-than-expected outcomes to improve the quality of care provided to their patients and to improve the outcomes of their patients. Similarly, hospital report cards permit the identification of health care providers that provide excellent quality of care. The reasons for their excellent performance can be investigated, so that information on best practices can be disseminated to all health care providers.

Publication of hospital report cards can lead to interventions tailored to improve hospital performance. Following the publication of Ontario's first public report card on the

From the \*ICES; †Institute of Health Management, Policy, and Evaluation, University of Toronto; ‡Sunnybrook Research Institute; §Department of Medicine, University of Toronto; and ||Peter Munk Cardiac Centre and University Health Network, Toronto, ON, Canada.

Supported by 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 paper 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. This research was supported by an operating grant from the Canadian Institutes of Health Research (MOP 86508). P.C.A. and D.S.L. are supported in part by Mid-Career Investigator awards from the Heart and Stroke Foundation. The datasets used for this study were held securely in a linked, deidentified form and analyzed at ICES. Parts of this material are based on data and/or information compiled and provided by CIHI. However, the analyses, conclusions, opinions, and statements expressed in the material are those of the authors, and not necessarily those of CIHI.

The authors declare no conflict of interest.

Correspondence to: Peter C. Austin, PhD, Institute for Clinical Evaluative Sciences (ICES), G106, 2075 Bayview Avenue, Toronto, ON, Canada M4N 3M5. E-mail: peter.austin@ices.on.ca.

Copyright © 2020 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 0025-7079/20/5807-0651

outcomes of patients hospitalized with AMI, individual hospitals implemented initiatives to improve the quality of care provided to these patients.<sup>9</sup> The EFFECT trial was a cluster-randomized trial that examined the effect of public reporting of hospital performance on an array of quality indicators for patients hospitalized with AMI or heart failure.<sup>10</sup> Following the public reporting of hospital performance, individual hospitals implemented quality improvement initiatives to improve their performance on specific indicators.<sup>10</sup>

An important quantity to know when designing and evaluating interventions to improve hospital performance is the absolute reduction in the number of adverse events (eg, deaths) across the entire population or jurisdiction if the performance of individual hospitals were to improve to specific normative standards. In proposing this metric, we assume that there is variation in performance across hospitals, with some hospitals having acceptable performance, and other hospitals having unacceptable performance. We want to estimate the absolute reduction in the number of adverse events if all hospitals with unacceptable quality of care were to improve their performance to be equal to the threshold defining acceptable quality of care. This would permit an estimate of the net population benefit if the performance of underperforming hospitals was improved to equal that of their high-performing peers. Knowledge of this quantity would provide information as to anticipated benefits if the performance of hospitals with inferior performance improved to that of a typical hospital. This quantity is not intended to inform on the need for quality improvement initiatives at specific hospitals. Instead, this quantity is important for the prioritization of system-wide quality improvement efforts. It can inform health planners about the potential absolute reduction in the number of deaths, across all hospitals, if the performance of all hospitals were changed to meet certain normative standards.

The paper is structured as follows: In the Statistical methods for estimating absolute reduction in the number of adverse events section, we describe statistical methods to estimate the absolute reduction in the number of adverse events due to improvements in hospital performance. In the Case study section, we provide a case study illustrating the application of these methods using data on patients hospitalized with AMI in Ontario. In the Discussion section, we summarize our findings and place them in the context of the existing literature.

### STATISTICAL METHODS FOR ESTIMATING ABSOLUTE REDUCTION IN THE NUMBER OF ADVERSE EVENTS

In this section, we describe statistical methods to estimate the absolute reduction in the number of adverse events (eg, death) at the population level due to improvements in hospital performance. The method is based on the use of random effects logistic regression models (also known as multilevel logistic regression models or hierarchical logistic regression models) and is motivated by the G-computation method for estimating the effects of interventions and exposures.<sup>11</sup> We describe bootstrap-based methods for use

with multilevel data to construct confidence intervals (CIs) around this estimated reduction in the number of adverse events.

### The Potential Outcomes Framework and G-Computation for Binary Exposures

The potential outcomes framework allows one to formalize the definition of the effects of exposures or interventions.<sup>12</sup> We describe this framework in a setting in which one active treatment is compared with one control treatment; however, the framework can be easily extended for settings with nonbinary categorical exposures or continuous exposures.

The 2 potential outcomes,  $Y(1)$  and  $Y(0)$  are the outcomes under the active and control treatments, respectively.  $Y(1)$  denotes a subject's outcome after receiving the active treatment, while  $Y(0)$  denotes the same subject's outcome when receiving the control treatment, under identical circumstances and at the identical time. Let  $Z$  denote an indicator variable denoting the actual treatment received ( $Z=1$  denoting receipt of the active treatment and  $Z=0$  denoting receipt of the control treatment). For a given subject, the effect of treatment is defined as  $Y(1)-Y(0)$ . The average treatment effect is defined as  $E[Y(1)-Y(0)]$ , the average effect of treatment in the population. However, for a given subject, only 1 of the 2 potential outcomes can be observed.

Parametric G-computation is a regression-based method for estimating the effects of interventions or exposures. A multivariable parametric regression model is used to regress the outcome on treatment status and baseline covariates.<sup>11</sup> In conventional parametric G-computation, assuming outcomes are binary, the following logistic regression model can be fit:

$$\text{logit}(p) = \beta_0 + \beta_1 X + \beta_2 Z. \quad (1)$$

where  $p$  denotes the probability of the occurrence of the binary outcome  $Y$ , and  $X$  denotes a vector of baseline covariates. In equation (1),  $\beta_1 X$  denotes the effect of patient characteristics on the log-odds scale, while  $\beta_2 Z$  denotes the effect of the binary treatment or exposure on the log-odds scale. Unlike a linear model, random variation is not induced through the inclusion of a subject-level error term, but through the distribution,  $Y \sim \text{Be}(p_i)$ . Using the fitted logistic regression model, the predicted probability of the outcome is estimated for each subject as if that subject had not been treated:  $\hat{P}(0) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$ . Second, the predicted probability of the outcome is estimated for each subject as if that subject had been treated:  $\hat{P}(1) = \frac{\exp(\beta_0 + \beta_1 X + \beta_2)}{1 + \exp(\beta_0 + \beta_1 X + \beta_2)}$ . For a given subject, the effect of the treatment can be estimated as the difference between the 2 imputed potential outcomes:  $\hat{P}(1) - \hat{P}(0)$ . Finally, the average treatment effect of interest can be estimated by averaging the subject-specific treatment effects over the entire sample.

If one had a categorical exposure with a few levels (eg, hospital A vs. hospital B vs. hospital C), one can easily modify the above approach. One would modify formula (1), so that, rather than a binary treatment/exposure variable, one would have a categorical variable with 3 levels (possibly represented using indicator variables for 2 of the 3 hospitals). One would then use the fitted model to estimate the probability of the

outcome if, possibly contrary to what was observed, all patients were treated at hospital A. One would then determine the mean predicted probability of the outcome if all patients were treated at hospital A. This procedure would then be repeated for hospitals B and C. Then, the risk difference comparing outcomes if all hospitals were treated at hospital A versus hospital B can be computed (along with the 2 other pair-wise comparisons). If the number of hospitals was large, this approach would be statistically inefficient due to the inclusion of a large number of indicator variables for representing the hospitals. In the Estimation of the Absolute Reduction in the Number of Adverse Events Due to Hospital Improvement in Performance section, we describe how this approach can be modified for use with a large number of hospitals.

### Estimation of the Absolute Reduction in the Number of Adverse Events Due to Hospital Improvement in Performance

In this section, we describe statistical methods to estimate the absolute reduction in the number of adverse events due to improvements in hospital performance. The method is based on the use of G-computation with hierarchical logistic regression models. Let  $Y_{ij}$  denote the binary outcome for the  $i$ th subject treated at the  $j$ th hospital [eg,  $Y_{ij} = 1$  denotes that the patient experienced the outcome (eg, death), while  $Y_{ij} = 0$  denotes that the patient did not experience the outcome] and let  $X_{ij}$  denote a vector of risk factors measured on this subject that will be used for risk adjustment.

The following random intercept logistic regression model can be fit to the data:

$$\text{logit}(p_{ij}) = \alpha_0 + \alpha_{0j} + \beta X_{ij}. \tag{2}$$

where  $p_{ij} = \Pr(Y_{ij} = 1)$ ,  $\alpha_0$  denotes the average intercept,  $\alpha_{0j}$  denotes the hospital-specific random effects for the  $j$ th hospital, and  $\beta$  denotes the vector of regression coefficients associated with the patient-level risk factors or covariates. We make the distributional assumption that  $\alpha_{0j} \sim N(0, \sigma^2)$ . Hospitals whose random effects are positive ( $\alpha_{0j} > 0$ ) have adverse events that occur with a higher probability than at an average hospital, while hospitals whose random effects are negative ( $\alpha_{0j} < 0$ ) have adverse events occur with a lower probability than at an average hospital. Let  $\hat{\alpha}_0, \hat{\alpha}_j, \hat{\beta}$ , and  $\hat{\sigma}^2$  denote the estimated average intercept, the predicted hospital-specific random effect for the  $j$ th hospital, the estimated vector of regression coefficients, and the estimated variance of the random effects distribution, respectively.

The predicted probability of the outcome for the  $i$ th patient at the  $j$ th hospital can be estimated as:

$$p_{ij} = \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_{0j} + \hat{\beta} X_{ij})}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_{0j} + \hat{\beta} X_{ij})}. \tag{3}$$

This denotes the predicted probability of the outcome for this subject conditional on the predicted hospital-specific random effect, which represents an estimate of the hospital's current performance. The predicted or expected number of adverse events at all hospitals can be determined as:

$$\sum_{j=1}^K \sum_{i=1}^{N_j} p_{ij}, \text{ where } N_j \text{ denotes the number of subjects at the } j\text{th}$$

hospital and  $K$  denotes the number of hospitals. This represents an estimate of the expected number of adverse events across all hospitals conditional on the current performance of the hospitals.

One can estimate the predicted or expected number of adverse events if hospital performance improved at some or all of the hospitals. For example, those hospitals whose relative performance was worse than that of an average hospital (ie,  $\alpha_{0j} > 0$ ) have their performance set equal to that of an average hospital (ie,  $\alpha_{0j} = 0$ ), while those hospitals whose relative performance was better than average (ie,  $\alpha_{0j} < 0$ ) have their performance left unchanged. One can then estimate each patient's predicted probability of an adverse event under this modification of hospital performance:

$$p_{ij}^{\text{modified}} = \begin{cases} \frac{\exp(\hat{\alpha}_0 + \hat{\beta} X_{ij})}{1 + \exp(\hat{\alpha}_0 + \hat{\beta} X_{ij})} & \text{if } \hat{\alpha}_{0j} > 0 \\ \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_{0j} + \hat{\beta} X_{ij})}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_{0j} + \hat{\beta} X_{ij})} & \text{if } \hat{\alpha}_{0j} \leq 0 \end{cases}. \tag{4}$$

Note that in the first component of this formula, the random effect has been omitted, implying that the random effect has been set equal to zero, which is the random effect for an average hospital. This modified probability can then be summed over all patients at all hospitals to determine the predicted or expected number of adverse events if those hospitals whose performance was worse than average had their performance improved to equal that of an average hospital. The difference between the expected number of adverse events under current hospital performance and that under the modified performance denotes the expected change in the number of adverse if those hospitals that had worse performance than that of an average hospital had their performance improved to be the same as that of an average hospital.

### The Multilevel Bootstrap for Estimating Confidence Intervals

Bootstrap-based methods can be used to estimate CIs for the absolute reduction in the number of adverse events.<sup>13</sup> We describe the use of the nonparametric residuals bootstrap adapted for use with 2-level multilevel data structures.<sup>14–16</sup>

First, the predicted hospital-specific random effects are centered so as to have mean zero. Let  $S^2$  denote the maximum likelihood estimate of the variance of the predicted hospital-specific random effects [note that  $S^2$  differs from  $\sigma^2$  following formula (2); the former is the sample variance of the predicted hospital-specific random effects, while the latter is the population variance of the true hospital-specific random effects]. Let  $R^2$  denote the estimated variance of the distribution of hospital-specific random effects obtained from the fitted multilevel logistic regression model (thus  $R^2$  is an estimate of  $\sigma^2$ ). We define an inflation factor by  $R/S$ , and each of the empirical or predicted hospital-specific random effects is multiplied by this inflation factor.<sup>16</sup> This accounts for the shrinkage in the predicted cluster-specific random effects. The inflated random effects will have the same sample variance as the variance of the distribution of the random effects from the fitted model. The nonparametric residuals bootstrap draws a bootstrap sample of hospital-specific random effects from the set of inflated hospital-specific predicted random effects.

Given that there are  $K$  hospitals, one draws  $K$  random effects ( $\alpha_{0j}^{bs}$ ) from this empirical distribution. For each subject, one then modifies formula (3) by replacing the predicted hospital-specific random effects with those drawn from the set of inflated empirical hospital-specific random effects:

$$p_{ij}^{bs} = \frac{\exp(\hat{\alpha}_0 + \alpha_{0j}^{bs} + \hat{\beta}X_{ij})}{1 + \exp(\hat{\alpha}_0 + \alpha_{0j}^{bs} + \hat{\beta}X_{ij})} \quad (5)$$

One then simulates a binary outcome for each subject as:  $Y_{ij}^{bs} \sim \text{Bernoulli}(p_{ij}^{bs})$ . Using the simulated data ( $Y^{bs}, X$ ), one then applies the methods described in the Estimation of the absolute reduction in the number of adverse events due to hospital improvement in performance section to estimate the absolute decrease in the expected number of adverse events under the given modification of hospital performance. This reduction in the number of adverse events is denoted by  $\Delta_i^{bs}$ . This procedure is followed  $B$  times, resulting in  $\Delta_i^{bs}, i=1, \dots, B$ . Percentile-based bootstrap CIs can be constructed by using the 2.5th and 97.5th percentiles of the empirical distribution of  $\Delta_i^{bs}$ .

### CASE STUDY

We provide a case study to illustrate the application of the methods described in Statistical methods for estimating absolute reduction in the number of adverse events section. We use data on patients hospitalized with AMI in the province of Ontario. We consider 3 different scenarios for improvements in hospital performance.

### Data Sources

We used data from the Ontario Myocardial Infarction Database (OMID), which contains data on patients hospitalized with an AMI at Ontario hospitals between 1992 and 2016.<sup>17</sup> For the case study, we used hospitalizations that occurred in the 12-month period between April 1, 2015, and March 31, 2016. Because of the study inclusion and exclusion criteria, no patient had > 1 hospitalization during the 1-year time frame of the study.<sup>17</sup> The data have a multilevel structure, with patients nested within hospitals. Similar to the HospitalCompare project, we excluded hospitals that treated <25 AMI patients during the 1-year period ([www.medicare.gov/HospitalCompare/Data/Death-rates.html](http://www.medicare.gov/HospitalCompare/Data/Death-rates.html)). The study sample consisted of 18,067 patients treated at 97 hospitals. The number of patients treated per hospital ranged from 25 to 979, with a median of 123 (25th–75th percentiles: 55–206).

Eleven patient-level variables, consisting of the variables in the Ontario AMI Mortality Prediction model (age, sex, congestive heart failure, cardiogenic shock, arrhythmia, pulmonary edema, diabetes mellitus with complications, stroke, acute renal disease, chronic renal disease, and malignancy), were used for risk-adjustment in the subsequent analyses.<sup>18</sup> The one continuous explanatory variables (age) was centered around the sample average. Information on the presence of the 9 comorbidities was extracted from the 24 secondary diagnosis fields from the discharge abstract database for the given hospitalization.

We considered 2 binary outcomes: death within 30 days of hospital admission and death within 1 year of hospital admission. These outcomes included both in-hospital deaths

and out-of-hospital deaths. A total of 1441 (8.0%) patients died within 30 days of hospital admission, while 2881 (15.9%) died within 1 year of admission.

### Statistical Analyses

For each of the 2 binary outcomes, we fit a random effects logistic regression model to regress the binary outcome on the 11 variables in the Ontario AMI Mortality Prediction model. The model incorporated random hospital-specific intercepts that were assumed to follow a normal distribution. We used the methods described above to estimate the population impact of the following 3 levels of hospital improvement: (i) those hospitals whose random effects were in the highest quartile of the empirical random effects distribution had their random effect decreased to equal that the 75th percentile of the empirical random effects distribution; (ii) those hospitals whose estimated random effects were positive (ie,  $\alpha_{0j} > 0$ ) had their random effects set equal to zero (ie,  $\alpha_{0j} = 0$ ); (iii) those hospitals whose random effects exceeded the 25th percentile of the empirical random effects distribution had their random effect decreased to equal the 25th percentile of the empirical random effects distribution. The first modification is the most modest, with only 25% of hospitals experiencing an improvement in performance, with the remaining 75% of hospitals having their performance unchanged. The improved performance at those hospitals that were in the top quartile was still worse than that of 75% of hospitals. The second modification is moderate, with 50% of hospitals experiencing an improvement in performance. The third modification was the most comprehensive, with 75% of hospitals experiencing an improvement in performance. When estimating bootstrap-based CIs, we used  $B = 5000$  bootstrap replicates.

The random effects logistic regression models were fit using PROC GLIMMIX in SAS (version 9.4), while G-computation was conducted using a series of data steps.

### Results

The estimated odds ratios and associated 95% CIs for the hierarchical logistic regression models predicting 30-day and 1-year mortality are reported in Table 1. The estimated

**TABLE 1.** Estimated Odds Ratios From the 2 Logistic Regression Models

Variables	30-Day Mortality	1-Year Mortality
Intercept	-3.38	-2.67
Age (per 10 y increase)	1.80 (1.70–1.89)	2.04 (1.95–2.13)
Female	1.12 (0.99–1.27)	1.03 (0.93–1.13)
Congestive heart failure	1.73 (1.50–1.99)	2.28 (2.04–2.55)
Stroke	2.77 (1.93–4.00)	1.91 (1.36–2.67)
Pulmonary edema	3.32 (1.84–6.01)	2.74 (1.57–4.78)
Diabetes with complications	1.23 (1.09–1.40)	1.55 (1.41–1.71)
Cancer	2.78 (2.17–3.57)	6.68 (5.42–8.25)
Chronic renal failure	1.19 (0.94–1.51)	1.85 (1.54–2.23)
Acute renal failure	2.14 (1.80–2.55)	2.02 (1.74–2.35)
Cardiogenic shock	18.30 (14.30–23.40)	12.69 (9.79–16.46)
Arrhythmia	1.60 (1.39–1.85)	1.57 (1.39–1.76)

The cells contain the estimated odds ratios and associated 95% confidence intervals. The intercept denotes the average intercept of the random effects logistic regression model.

variances of the distributions of the random effects were 0.00863 and 0.02206 for the models for 30-day and 1-year mortality, respectively. These are equivalent to standard deviations of 0.093 and 0.149. Thus, the hospital-specific random intercepts would come from an  $N(-3.38, \sigma = 0.093)$  and  $N(-2.67, \sigma = 0.149)$ , respectively. Increasing patient age and 8 of the 9 comorbid conditions were associated with an increased odds of 30-day mortality. The 95% CIs for female sex and chronic renal failure both contained the null value. Increasing patient age and all 9 comorbid conditions were associated with an increased odds of 1-year mortality. The 95% CI for female sex contained the null value.

For 30-day mortality, the median predicted hospital-specific random effect was zero, while the first and third quartiles were  $-0.008$  and  $0.006$ , respectively. The fifth and 95th percentiles were  $-0.021$  and  $0.023$ , respectively. For 1-year mortality, the median predicted hospital-specific random effect was  $-0.002$ , while the first and third quartiles were  $-0.033$  and  $0.047$ , respectively. The fifth and 95th percentiles were  $-0.135$  and  $0.109$ , respectively.

The absolute reduction in the number of deaths avoided within 30 days and 1 year under the 3 different scenarios about improvements in hospital performance are reported in Table 2. In the analytic sample, 1441 patients died within 30 days of hospital admission. If the performance of the 25% of hospitals with the highest adjusted mortality had similar performance to the hospital at the 75th percentile of performance, the estimated reduction in the number of deaths within 30 days would be 3.5. If the performance of those hospitals whose performance was worse than that of an average hospital had their performance modified to be equal to that of an average hospital, the estimated reduction in the number of deaths within 30 days would be 6.0. If the performance of the 75% of hospitals with the highest adjusted mortality had similar performance to that of the hospital at the 25th percentile, the estimated reduction in the number of deaths within 30 days would be 11.0. On the basis of the observed number of deaths of 1441, the percentage of deaths that could be avoided by these 3 modifications of hospital performance were 0.2%, 0.4%, and 0.8%, respectively. For all 3 scenarios of hospital improvement, an estimated 95% CIs excluded the null value of zero. The CIs provide an indication of the precision with which the absolute reduction in mortality is estimated. Narrower intervals imply that this quantity is estimated with greater precision. If the CI contains

the null value of zero, then the reduction in the number of deaths is not statistically significantly different from zero (ie, no net reduction in the number of deaths). Note the relatively wide CIs, indicating moderate uncertainty in the absolute reduction in the number of deaths due to hospital improvement. Furthermore, the width of the CIs increases as the number of hospitals at which improvements occur increases. Similar results, with amplification in the absolute reduction in the number of deaths, were observed for 1-year mortality.

### Complementary Analyses

We conducted a set of complementary analyses to estimate hospital-specific risk-adjusted mortality rates. We fit 2 conventional logistic regression models in which each of the 2 binary outcomes (30-day and 1-year mortality) was regressed on the 11 patient characteristics described above. From each of the fitted models, we computed the expected number of deaths at each hospital as the sum of the predicted probabilities of the outcome across all patients at that hospital. Risk-adjusted mortality rates were computed for each hospital as the ratio of the observed number of deaths to the expected number of deaths, multiplied by the overall cohort-wide mortality rate.<sup>19</sup> Ninety-five percent CIs were constructed using a method described by Hosmer and Lemeshow.<sup>20</sup> Hospitals whose 95% CI lay entirely above the overall cohort-wide mortality rate were classified as high-mortality outliers, while those whose 95% CIs lay entirely below the overall cohort-wide mortality rate were classified as low-mortality outliers. For 30-day mortality, 5 hospitals were classified as high-outliers while 3 hospitals were classified as low-outliers. For 1-year mortality, 6 hospitals were classified as high-outliers, while 3 hospitals were classified as low-outliers.

For each hospital that was identified as a high-mortality outlier, we computed the number of excess deaths as the difference between the observed number of deaths and the expected number of deaths and summed this quantity across the high-mortality outliers. For 30-day mortality, the excess number of deaths at the 5 high-mortality outliers was 34.4. For 1-year mortality, the excess number of deaths at the 6 high-mortality outliers was 43.8.

Finally, for each hospital, we computed the observed number of deaths. We also computed the ratio of observed-to-expected number of deaths at each hospital based on the fitted logistic regression model. For those hospitals whose observed-to-expected ratio exceeded the 75th percentile (ie, were in the top 25% of hospitals), we set the observed-to-expected ratio to equal the 75th percentile of this ratio (ie, the performance of these hospitals was improved). We then determined the anticipated number of observed deaths (under this improvement) by multiplying the expected number of deaths by this modified ratio of observed-to-expected deaths. We then determined the difference between the actual observed number of deaths and the anticipated number of observed deaths under this improvement and summed this quantity across those hospitals whose observed-to-expected ratio exceeded the 75th percentile. The nonparametric or cases bootstrap, in which 5000 bootstrap samples of hospitals

**TABLE 2.** Estimated Absolute Decrease in the Number of Deaths Within 30 Days and 1 Year With Associated 95% Confidence Intervals

Improvement Scenario	30-Day Mortality	1-Year Mortality
Top 25% of hospitals have performance changed to that of the 75th percentile	3.5 (0.4–26.5)	17.9 (6.9–66.4)
Hospitals whose performance is worse than average have performance changed to that of an average hospital	6.0 (0.7–47.0)	44.5 (13.8–121.8)
Top 75% of hospitals have performance changed to that of the 25th percentile	11.0 (1.2–79.0)	73.3 (23.0–206.2)

were drawn, was used to construct 95% bootstrap percentile CIs for this quantity.<sup>14,15</sup> For 30-day mortality, we estimated that 45.6 deaths (95% CI: 26.8–65.1) would be avoided by improving the performance of 25% of hospitals with the highest observed-to-expected ratios. For 1-year mortality, the corresponding figure was 45.4 deaths (95% CI: 30.3–67.5).

## DISCUSSION

We described a method based on G-computation in conjunction with a fitted random effects logistic regression model to estimate the absolute reduction in the number of adverse events if the performance of some hospitals was improved to specified normative standards. We illustrated the utility of this method by applying it to data consisting of patients hospitalized with an AMI in Ontario, Canada. We found that in the given year, 6.0 deaths would be avoided within 30 days of hospital admission if the performance of those hospitals whose performance was worse than that of an average hospital had their performance improved to equal that of an average hospital. Providing this information would not be intended to inform quality improvement initiatives at specific hospitals. Instead, this information would provide an estimate of the expected reduction in the number of deaths (or other adverse events) across the health care system, if the performance of all hospitals was improved to meet specific normative standards.

We compared the use of G-computation using hierarchical models with methods based on fitting conventional logistic regression models and computing ratios of observed-to-expected mortality. We found that estimates of the expected reduction in the number of deaths, if the top 25% of hospitals were to have their performance improved, was substantially attenuated towards zero when using G-computation with hierarchical models compared with when using conventional model-based indirect standardization. We hypothesize that at least part of this attenuation is due to the shrunken estimates of the predicted random effects that are obtained when using hierarchical regression models.

Hospital report cards are expensive and time-consuming to produce. The methods described in this study permit estimation of the expected reduction in the number of adverse events if the performance of a subset of hospitals was to improve to equal specific standards. These methods do not permit estimation of the impact of a particular intervention such as implementing standing orders. However, the described methods do provide an estimate of the anticipated population benefit if performance at some hospitals was improved so as to equal the performance that is already being achieved by other hospitals. On the basis of the assumption that the risk-adjusted performance observed at some hospitals is a realistic goal for all hospitals, these methods provide an estimate of the anticipated population benefit if this achievable performance was attained by all hospitals. By applying the proposed methods to different conditions (eg, patients undergoing CABG surgery, patients hospitalized for heart failure, and patients hospitalized for AMI), conditions and procedures could be ranked according to the anticipated absolute population benefit if the performance of poorly performing hospitals was improved to that of an average hospital. This would allow health system administrators and

policymakers to prioritize which conditions or procedures should be the focus of a hospital report card.

The use of G-computation to address the effects of policies or of questions around the structure of the health care system appears to be rare. A search of PubMed ([www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed)) using the search term “G-computation” identified 90 articles (date of search: August 22, 2019). Almost all applications of G-computation involved examining the effects of conventional treatments and exposures, and not to address health systems questions. One of the few exceptions was a study that used conventional G-computation with a binary exposure to estimate the reduction in surgical deaths due to the regionalization of higher risk surgical procedures.<sup>21</sup> It was estimated that the regionalization of colorectal surgery, esophagectomy, and pancreaticoduodenectomy in Ontario would reduce the average annual number of perioperative deaths by 20.2, 2.0, and 3.6, for the 3 procedures, respectively. Another novel application of G-computation was a study that examined transportation planning policies on the number of bicyclist fatalities.<sup>22</sup> The novelty of the current application is the combining G-computation with random effects logistic regression models. While not using the term G-computation, 2 further studies were identified that used similar approach to address questions around the effect of changing hospital or regional performance. Simpson et al<sup>23</sup> examined between-hospital variation in rates of severe intraventricular hemorrhage in preterm babies in Australia and New Zealand. They found that if all neonatal intensive care units could achieve a rate equal to the 20th percentile, then 60 cases of severe intraventricular hemorrhage could be prevented over 3 years. Similarly, Yu et al<sup>24</sup> studied between-health areas survival from colorectal cancer in New South Wales, Australia. They estimated that 784 patients who died within 5 years due to colon cancer could have had their survival increased to >5 years if the excess risk of death in all health areas was reduced to the 20th percentile.

There are certain limitations of the current study. First, we relied on administrative health care data. While these data provide population-based coverage, they do not contain information on risk factors such as blood pressure and smoking status as well as on factors such as coronary disease anatomy. It is possible that the estimated reduction in the number of deaths could change were we able to account for additional risk factors. However, the primary objective of the current study was to describe a methodological framework for estimating the absolute reduction in deaths were hospital performance to improve. Second, our estimate of the absolute reduction in the number of deaths relies on the assumption that the regression model has been correctly specified. If the model were misspecified, then it is possible that the estimated reduction in deaths is subject to bias. However, the model was developed in Ontario and was then subsequently validated in both Manitoba and California.<sup>18</sup> Thus, it is likely that the model accurately predicts AMI mortality.

In summary, G-computation in conjunction with random effects logistic regression models can be used to estimate the absolute reduction in the number of adverse events if the performance of some health care providers was improved so as to meet specified normative standards. This method allows

for an estimation of the possible net benefit of campaigns to improve hospital performance.

## REFERENCES

1. New York State Department of Health. *Coronary Artery Bypass Graft Surgery in New York State 1989–1991*. Albany, NY: New York State Department of Health; 1992.
2. Jacobs FM. *Cardiac Surgery in New Jersey in 2002: A Consumer Report*. Trenton, NJ: Department of Health and Senior Services; 2005.
3. Massachusetts Data Analysis Center. *Adult Coronary Artery Bypass Graft Surgery in the Commonwealth of Massachusetts: Fiscal Year 2010 Report*. Boston, MA: Department of Health Care Policy, Harvard Medical School; 2012.
4. Naylor CD, Rothwell DM, Tu JV, et al. Outcomes of coronary artery bypass surgery in Ontario. In: Naylor CD, Slaughter PM, eds. *Cardiovascular Health and Services in Ontario: An ICES Atlas*. Toronto, ON, Canada: Institute for Clinical Evaluative Sciences; 1999:189–198.
5. Pennsylvania Health Care Cost Containment Council. *Consumer Guide to Coronary Artery Bypass Graft Surgery*. Harrisburg, PA: Pennsylvania Health Care Cost Containment Council; 1995.
6. Luft HS, Romano PS, Remy LL, et al. *Annual Report of the California Hospital Outcomes Project*. Sacramento, CA: California Office of Statewide Health Planning and Development; 1993.
7. Pennsylvania Health Care Cost Containment Council. *Focus on Heart Attack in Pennsylvania: Research Methods and Results*. Harrisburg, PA: Pennsylvania Health Care Cost Containment Council; 1996.
8. Tu JV, Austin PC, Naylor CD, et al. Acute myocardial infarction outcomes in Ontario. In: Naylor CD, Slaughter PM, eds. *Cardiovascular Health and Services in Ontario: An ICES Atlas*. Toronto, ON, Canada: Institute for Clinical Evaluative Sciences; 1999:83–110.
9. Tu JV, Cameron C. Impact of an acute myocardial infarction report card in Ontario, Canada. *Int J Qual Health Care*. 2003;15:131–137.
10. Tu JV, Donovan LR, Lee DS, et al. Effectiveness of public report cards for improving the quality of cardiac care: the EFFECT study: a randomized trial. *J Am Med Assoc*. 2009;302:2330–2337.
11. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol*. 2011;173:731–738.
12. Rubin DB. Statistical inference for causal effects, with emphasis on applications in epidemiology and medical statistics. In: Rao CR, Miller JP, Rao DC, eds. *Handbook of Statistics Volume 27 Epidemiology and Medical Statistics*. Amsterdam, The Netherlands: North-Holland; 2008:28–58.
13. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall; 1993.
14. Goldstein H. Bootstrapping in multilevel models. In: Hox JJ, Roberts JK, eds. *Handbook of Advanced Multilevel Analysis*. New York, NY: Routledge; 2011:163–171.
15. van der Leeden R, Meijer E, Busing FMTA. Resampling multilevel models. In: de Leeuw J, Meijer E, eds. *Handbook of Multilevel Analysis*. New York, NY: Springer; 2008:401–433.
16. Carpenter JR, Goldstein H, Rasbash J. A novel bootstrap procedure for assessing the relationship between class size and achievement. *J R Stat Soc Ser C Appl Stat*. 2003;52:431–443.
17. Tu JV, Austin P, Naylor CD. Temporal changes in the outcomes of acute myocardial infarction in Ontario, 1992–96. *Can Med Assoc J*. 1999;161:1257–1261.
18. Tu JV, Austin PC, Walld R, et al. Development and validation of the Ontario acute myocardial infarction mortality prediction rules. *J Am Coll Cardiol*. 2001;37:992–997.
19. Iezzoni LI. *Risk Adjustment for Measuring Health Outcomes*. Chicago, IL: Health Administration Press; 1997.
20. Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance based on logistic regression models. *Stat Med*. 1995;14:2161–2172.
21. Austin PC, Urbach DR. Using G-computation to estimate the effect of regionalization of surgical services on the absolute reduction in the occurrence of adverse patient outcomes. *Med Care*. 2013;51:797–805.
22. Mooney SJ, Magee C, Dang K, et al. “Complete Streets” and adult bicyclist fatalities: applying G-computation to evaluate an intervention that affects the size of a population at risk. *Am J Epidemiol*. 2018;187:2038–2045.
23. Simpson JM, Evans N, Gibberd RW, et al. Analysing differences in clinical outcomes between hospitals. *Qual Saf Health Care*. 2003;12:257–262.
24. Yu XQ, O’Connell DL, Gibberd RW, et al. A population-based study from New South Wales, Australia 1996–2001: area variation in survival from colorectal cancer. *Eur J Cancer*. 2005;41:2715–2721.