Developing risk-adjusted 30-day hospital mortality rates

Ann Tourangeau[1,2], Jack V.Tu[2,3,4]

[1]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, Ontario, Canada
[2]Institute for Clinical Evaluative Sciences in Ontario, Toronto, Canada
[3]Faculty of Medicine, University of Toronto
[4]Division of General Internal Medicine, Sunnybrook and Women's College Hospital

Published 2003 in *Research in Nursing & Health*, 26, 483-496
http://onlinelibrary.wiley.com/doi/10.1002/nur.10110/abstract

ABSTRACT

This article describes one risk-adjustment method useful for minimizing threats to internal validity that stem from the impact on the outcomes under investigation from patients' own characteristics and their associated risks. Mortality is the outcome used to illustrate the risk-adjustment approach. A two-step approach to outcomes research is suggested. The first step includes risk-adjusting outcomes for patient characteristics by developing standard mortality rates. In the second step these risk-adjusted standard rates can be used as dependent variables in outcomes analytic models. This article focuses on the first step. In the current study risk adjustment resulted in changes in both absolute values and rank ordering of hospital mortality rates compared to crude rates. This procedure is useful for developing risk-adjusted.

Key Words: risk-adjustment methodology; mortality; administrative data; acute care hospitals; outcomes research

Outcomes are end results of care (Hegyvary, the health status of individuals, groups, or 1991; Jeffs, 1998; Jennings & Staggers, 1998) and societies and attributing those changes to the one group of indicators of health care quality structures and processes of health care. Conclu(Burns, Moores, & Breslin, 1996; Mark, 1995). sions about the quality of hospital care often are Outcomes research involves examining changes in made based on hospital performance in such patient outcomes as mortality rates. There are a number of issues related to the feasibility, usefulness, and validity of outcomes research in nursing. These issues include which outcomes should be measured (Brooten & Naylor, 1995; Griffiths, 1995; Hegyvary; Mark & Burleson, 1995), lack of theory linking a set of outcomes with their hypothesized determinants (Griffiths; Hegyvary; Mark, 1995; Sidani & Braden, 1998; Wood & Brink, 1998), lack of accessible data and databases (French, 1997; Griffiths; O'Brien-Pallas, Irvine, Peereboom, & Murray, 1997), data reliability (Barnsley, Lemieux-Charles, & Baker, 1996; Mark & Burleson), and threats to the internal validity of findings because of differences in patient characteristics (Iezzoni, 1997c; Lamb, 1997; Tourangeau, Giovannetti, Tu, & Wood, 2002). It is this last issue that is examined in this article.

The nursing research literature has remained somewhat silent about appropriate and effective methods for calculating risk-adjusted outcomes in order to compare performance across groups or organizations. Effective risk adjustment should account for the patient characteristics and associated risks that affect outcomes under investigation. In this article we describe one risk adjustment method used to calculate hospital mortality rates. The patient data used in this study were extracted from a large administrative database containing uniform data and limited clinical details about patients discharged from hospitals in Ontario, Canada. Risk-adjustment procedures were implemented in the study in order to minimize the threats to internal validity from the impact of differences in patient characteristics within and across hospitals on hospital mortality rates.

The three major sources of variation in patient outcomes are individual patient characteristics, quality of patient care (e.g., the structures and processes of delivered health care), and random variation (Jencks et al., 1988; Silber, Rosenbaum, & Ross, 1995; Smith, 1994). By far, the greatest determinant of patient mortality is patient characteristics such as age and preexisting health conditions (Knaus, Wagner, Zimmerman, & Draper, 1993; Silber & Rosenbaum, 1997). In fact, Silber and Rosenbaum found that patients' own characteristics were 315 times more important than hospital care delivery characteristics in predicting mortality.

Conclusions drawn from studies comparing outcomes across any group, including hospitals, are only as valid or credible as the effectiveness of the measures used to adjust these outcomes for patient characteristics and their associated risks. If a study purpose is to understand or explain health care delivery determinants of an outcome, it is important that the effects of patients' own charateristics on that outcome are identified and controlled or accounted for before the effects of health care services on the outcome are isolated and measured.

A good outcome indicator should be sensitive to differences in quality of care provided in the hospital yet insensitive to differences in the health of patients in that hospital (Silber et al., 1995). Risk-adjustment procedures are conducted to account for those patient factors that existed before a health care intervention was delivered (e.g., hospitalization) and that could affect or produce variations in the outcomes under investigation (Daley & Shwartz, 1997; Iezzoni, 1997c, 1997d).

Researchers may choose a variety of ways to account

for the effects of patients' own characteristics on the outcomes being studied. One method is to add indicators of patient characteristics into analytic models such as a linear regression, along with other health care delivery predictors (Daley & Shwartz, 1997). The major drawback of this approach is that patient characteristics have a very large effect on outcomes, far overshadowing the impact of indicators of health care delivery. Using this method, it may be very difficult to interpret the impact of health care delivery indicators on crude or unadjusted outcomes.

In a second category of methods a two-step approach is used in which in the first step the effects of patient characteristics on the outcome are measured and used to develop standard outcome rates (Daley & Shwartz, 1997), followed by the second phase of analysis, in which these standard outcome rates are used as outcome indicators to determine the impact of health care delivery variables on the outcome. In this approach the ''noise'' of patients' own characteristics is already quieted before implementing the second-phase analysis with such statistical tools as regression models. In a two-step approach the effects of health care delivery (as well as random variation) are more visible and interpretable.

We illustrate the second risk-adjustment methodological approach by its use in the current study to determined risk-adjusted 30-day mortality rates for a sample of 75 Ontario, Canada, acute care hospitals over the 1998–1999 reporting period. Thirty-day mortality refers to patient death occurring within 30 days of admission, whether in the hospital or after discharge. Thirty-day post-admission mortality was chosen as the mortality outcome as it is believed to measure the full impact of hospitalization without introducing too many competing risks (Jencks, Williams, & Kay, 1988; Silber, Williams, Krakauer, & Schwartz, 1992).

In this article we first describe the sources of data used to develop the risk-adjusted mortality rates and then explain how and why the sample hospitals and the patients in those hospitals were selected for inclusion in this study. We identify the indicators used in the logistic regression models, chosen because of evidence found suggesting these indicators alter patients' risk of death. We describe how and why mortality rates were weighted to reflect patient case mix within hospitals before developing risk-adjusted standard mortality rates for each hospital. And we report the evaluation of the results of the refined logistic regression models for logistic regression model performance.

Method

Data sources

Three sources of data were used to calculate risk-adjusted hospital mortality rates. First, the administrative health database, the Discharge Abstract Database (DAD) from the Ontario Ministry of Health and Long-Term Care for the reporting period 1998–1999, was used to identify patients discharged from Ontario acute care hospitals for potential inclusion in the study. This database is produced for each fiscal reporting period beginning April 1 and ending March 31 of the following year. The DAD has a record for every patient discharged from every Ontario hospital for each fiscal year. Clinical variables in this database include up to 16 diagnosis codes, up to 10 procedures, and physician codes identifying the physician most responsible for care related to each of the diagnosis and procedure codes. At the time of this study, patient diagnosis codes in the DAD were classified according to the International Classification of Diseases, ninth revision (International Classification of Diseases, 1999). The DAD contains patient demographic variables (e.g., sex, date of birth, postal code), administrative hospital information (e.g., institution number, length of stay, dates of admission and discharge), and resource consumption information such as resource intensity weight and case mix group (Richards, Brown, & Homan, 2001).

The DAD is a Canadian national health administrative database. At least 75% of Canadian hospitals are mandated to abstract discharged patient records and report this information to the Canadian Institute for Health Information (CIHI), which is a national not-for-profit organization mandated to coordinate the development and maintenance of a comprehensive and integrated approach to health information in Canada (Richards et al., 2001). This database has similar data fields to those used by the U.S. Uniform Hospital Discharge Data Set (Iezzoni, 1997a). Major differences between the Canadian and U.S. databases are: race and ethnicity patient data are not collected in the Canadian DAD but are collected in the U.S. Uniform Hospital Discharge Data Set (UHDDS); the only source of hospitalization payment for Canadian residents is the resident's provincial government, whereas in the U.S. UHDDS there are multiple sources of payment; and definitions of some diagnostic categories and diagnoses may differ between databases.

Patient records from the DAD were linked with the second data source, a subset of the Statistics Canada 1996 Population Data file that contains postal codes and average incomes within each postal code. These data were used to estimate the socioeconomic status (SES) of each patient.

The third source of data was the Ontario Registered Persons Database, a subset of the Ontario Vital Statistics file, which includes death dates for Ontario residents. The Registered Persons Database was linked to cases from the DAD to identify dates of death within 30 days of sample patient admission to a hospital. All statistical analyses were completed in SAS[1], version 8. Ethical approval for this study was obtained from the Capital Health Region Ethics Review Panel, Edmonton, Alberta, in January 2000 and was renewed in February 2001.

Sample hospitals and patients

Ontario acute care hospitals discharged 1,187,414 patients between April 1, 1998, and March 31, 1999. To be in the sample, patients had to be current Ontario residents with valid Ontario Health Insurance Plan numbers, so that their hospitalizations could be linked with dates of death in the Ontario Registered Persons Database. Patients had to have one of nine most responsible diagnosis (MRD) codes in the four selected general diagnosis categories: acute myocardial infarction (AMI), stroke, pneumonia, and septicemia. The MRD is the diagnosis that describes the most significant patient condition causing hospitalization and contributing most to patient length of stay (Canadian Institute for Health Information, 1995). These four

medical diagnosis categories were included because they are considered acute rather than chronic conditions and because they have both high patient volumes and high crude death rates. A number of inclusion criteria were invoked to select the sample hospitals and the sample patients in those hospitals. First, only Ontario teaching and community(medium-sized) hospitalswere included. In Ontario, acute care hospitals are categorized as teaching, community, or small. Teaching hospitals are those hospitals designated as teaching by the Ontario Council of Teaching Hospitals because of their direct affiliation with an Ontario university that provides medical education. Small hospitals are those hospitals that discharge fewer than 3,500 weighted cases (weighted cases being the sum of the resource intensity weights for discharged patients over the fiscal-year period), have referral populations of fewer than 20,000 people, and are the only hospitals in their communities. All other Ontario acute care hospitals are categorized as community or medium-sized hospitals (Baker et al., 1999). Only teaching and community hospitals were included because of the low volume of patients discharged from small hospitals over the year. To develop risk-adjusted hospital mortality rates, the event of death must have occurred. In small hospitals, the outcome of death may not occur at all or may occur without sufficient frequency to develop risk-adjusted mortality rates. For the same reason, community hospitals that discharged fewer than 100 patients in each of the four diagnosis categories were excluded from the study, a criterion that resulted in four community hospitals being excluded. The final hospital sample consisted of 75 hospitals: 10 teaching and 65 community hospitals.

To be included, discharged patients from sample hospitals had to be older than 19 years of age. Only patients for whom the MRD was the initial reason for admission were included. This was determined using the DAD when a patient had an identical MRD diagnosis code and complication or secondary diagnosis code. This indicates that the initial reason for hospitalization was different from the MRD. These patients were excluded from the study. If a patient had more than one admission with the same MRD during the year, only the patient's first admission was included in the study. Two exclusion criteria were employed. First, patients who had transferred in from another acute care hospital were excluded in order to avoid double counting of patients. Second, patients who had a preadmission diagnosis of cancer, HIV, or palliative care were excluded as the trajectory of disease and health goals of these patients often differ from those of patients with acute disease. The final sample included 46,941 discharged patients from 75 hospitals, 4% of all patients discharged from Ontario hospitals that year. For each of the four diagnosis categories, Table 1 describes the ICD-9 codes used to select sample patients, the number of patients, the average patient age, its proportion of patients in the whole sample, and the crude 30-day mortality rate.

Risk-adjustment models and risk factors

We used an indirect standardization risk-adjustment method (Bland, 1999). The general form of the risk-adjusted 30-day standard mortality rate for each hospital was the ratio of observed deaths divided by the expected number of deaths. We calculated the numerator (observed number of deaths in each hospital) by linking the DAD with the Ontario Registered Persons Database, but logistic regression models were needed to calculate the denominator (expected number of deaths for each hospital). Four logistic regression models were implemented to estimate the expected value or probability of death within 30 days for each patient. Separate logistic regression models were developed for each of the four diagnosis categories because patient characteristics and their associated risk of death are not necessarily the same in each disease category. Two sets of logistic regression models were used. In the first set all candidate predictors were entered into the logistic regression models, and in the second set models were refined to include only those predictors found to be statistically significant at the .05 level (Daley & Shwartz, 1997).

Risk factors were defined as patient characteristics or medical conditions that existed on admission to hospital (preadmission diagnoses). In the first set of logistic regression models, the candidate predictors, all 21 of which were entered in each of the four logistic regression models, were age, sex, 14 indicators for the Deyo modification of the Charlson comorbidity index (Charlson, Pompei, Ales, & MacKenzie, 1987; Deyo, Cherkin, & Ciol, 1992), four SES indicators, and a newly developed chronicity variable. Chronicity was used as a proxy indicator for each patient's general health status and functioning. It was a dichotomous variable that identified patients admitted to acute care hospital from any category of health care facility that is nonacute but provides long-term supportive care (e.g., long-term care, chronic care). Only four of the five SES indicators were entered as predictors in each model to allow the regression model to converge. The indicator for very high income was left out of logistic regression models and acted as the reference group for SES. Within the stroke logistic regression model, one additional indicator predictor variable for the diagnosis of hemorrhagic stroke was added. This predictor was added because the high mortality associated with this type of stroke meant patients were at higher risk of death. Table 2 shows each of the model predictor mortality risk factors, the ICD-9 codes used to identify their occurrence (where appropriate), and the prevalence of predictors in each of the four logistic regression models.

Age was entered as a continuous variable in all four models. Deyo-modified Charlson indices were developed as independent dichotomous indicators by identifying the presence or absence of any of the comorbid diagnosis codes if coded as a preadmission diagnosis. The Charlson comorbidity index was first developed as a weighted prognostic taxonomy for comorbid conditions that might increase the risk of death in the short term (Charlson et al., 1987). Deyo and colleagues later adapted the Charlson index by translating the index into a set of ICD-9-CM codes (Deyo et al., 1992). Three of the 17 Charlson variables indicate cancer and HIV disease. As this study excluded patients with cancer and HIV preadmission conditions, only the following 14 Charlson indicators were used in logistic regression models to predict death: AMI, heart failure, peripheral vascular disease, cerebrovascular disease, presenile psychosis, respiratory disease, rheumatic-like diseases, ulcers of the digestive system, liver cirrhosis, diabetes mellitus, diabetes mellitus with complications, para/hemiplegia, kidney disease, and liver or esophageal disease. Because of evidence that SES is inversely related to mortality

Table 1 Characteristics of Sample Patients

Diagnosis Category	ICD-9 Codes for MRD	Number of Cases	Average Age	Proportion of Cases in sample	Crude Mortality Rate
AMI	410	17,617	67.8	.38	11
Stroke	431	11,445	73.8	.24	18
	434				
	436				
Pneumonia	481	15,643	71.1	.33	12
	482				
	485				
	486				
Septicemia	038	2,236	69.8	.05	20
Overall	—	46,941	70.4	1.00	14

Note: ICD-9 = International Classification of Diseases 9th revision.
MRD = most responsible diagnosis. AMI= acute myocardial infarction.

(Gregario, Walsh, & Paturzo, 1997; Pappas, Queen, Hadden, & Fisher, 1993), we used measures of SES as potential 30-day mortality predictors. SES was estimated by linking the domicile postal code of each patient with the average income for that postal code. Because we were interested in understanding exactly which categories of SES were risk factors for death, five indicator variables representing five average-income quintiles were derived from the Statistics Canada 1996 population file to measure SES. The range of incomes represented by the quintiles ranged from very low income to very high income. The cutoffs for each Ontario average yearly income category in Canadian dollars were:

- $7,680–$39,851 (very low)
- $39,852–$49,734 (low)
- $49,735–$57,852 (middle)
- $57,853–$70,494 (high)
- $70,495–$304,454 (very high).

Four of the five indicator SES variables were entered as predictors into logistic regression models, with the very high income indicator left out of models as it acted as the reference group.

The DAD does not have a field that identifies each patient's general health status and functional ability, either at admission or at discharge. However, there is evidence of higher mortality rates among patients admitted to acute care hospitals from other long-term supportive health care environments (Green, Passman, & Wintfeld, 1991). For each discharged patient, the DAD has a field that identifies the setting—such as home, long-term care facility, rehabilitation hospital, and so on—from which a patient is admitted to an acute care hospital. It is reasonable to assume that patients admitted from nonacute health care settings such as chronic hospitals, nursing homes, long-term care settings, and homes for the aged generally have decreased health status and functioning abilities, which necessitated their being cared for in such a long-term supportive care environment. The dichotomous indicator variable chronicity, a proxy indicator for general health status and functioning, was developed to identify patients admitted to acute care hospital from any of these non-acute and long-term supportive care settings and was entered as a predictor, or risk factor, for death in all four of the logistic regression models.

The crude death rate for patients diagnosed with hemorrhagic stroke (ICD-9 code 431) was found to be significantly higher than that for other stroke code diagnoses (41% versus 15% in this sample). Because of the increased risk of death for patients with this specific stroke type, a hemorrhagic stroke indicator was created and included only within the stroke logistic regression model.

To determine whether there was colinearity among candidate predictors, we examined Pear-son correlation coefficients among all possible pairs of predictors separately within each of the four logistic regression models. No evidence of colinearity was found in any model. The highest correlation among predictors in the four logistic regression models was .29, found between age and chronicity in the pneumonia and septicemia models. After the same candidate predictors were entered into the first set of four logistic regression models, the prediction models were further refined to include only those statistically significant predictors with probabilities less than .05. If any of the SES predictor variables in a model were statistically significant, the four SES variables were maintained in the model.

A common measure of logistic regression model performance is the c statistic, or area under the receiver operating characteristic (ROC) curve. The predictive validity of a model can be assessed through examination of the resulting c statistic. The value of the c statistic ranges from 0 to 1.0 and reflects the model's ability to accurately discriminate between those who die and those who do not by comparing predicted and actual values of 30-day death (Ash & Shwartz, 1997). A value of .50 indicates the model is no better than chance in predicting 30-day death, and a value of 1.0 indicates perfect prediction. In studies predicting death, areas under the ROC curve or c statistics greater than .70 are considered to have adequate or good model discrimination (Ash & Shwartz, Green, Wintfeld, Sharkey, & Passman, 1990; Hosmer & Lemeshow, 2000). A second common indicator of model performance is the Hosmer and Lemeshow goodness-of-fit test. This test evaluates whether the proposed model fits the observed data and, in this analysis, assesses for discrepancies between observed and predicted

Table 2 Prevalence of potential Risk Factors for 30-Day Mortality in Each Diagnostic Category

Risk Factors	ICD-9 Code	AMI (%)	Stroke (%)	Pneumonia (%)	Septicemia (%)
Sex (percentage male)		64.1	48.8	49.9	49.2
Acute myocardial infarction	410	10.4	6.7	5.0	5.1
Heart failure	428	22.1	5.8	16.2	9.5
Peripheral vascular disease	441, v43.4, 443.9, 785.4	3.3	2.7	2.0	2.9
Cerebrovascular disease	430, 431, 432, 433, 434, 435, 436, 437, 438	4.1	16.1	5.7	6.5
Pre/Senile psychosis	290	1.4	3.8	4.7	5.8
Respiratory disease	490, 491, 492, 493, 494, 495, 496, 500, 501, 502, 503, 504, 505, 506.4	9.4	6.3	30.4	8.5
Rheumatic-like disease	710.0, 710.1, 710.4, 714.0, 714.1, 714.2, 714.8, 725	1.0	0.9	2.0	2.2
Ulcers of digestive system	531, 532, 533, 534	0.4	0.6	0.5	0.6
Liver cirrhosis	571.2, 571.5, 571.6, 571.4	0.1	0.3	0.6	1.6
Diabetes mellitus	250.0, 250.1, 250.2, 250.3, 250.7	20.6	21.1	14.4	20.8
Diabetes mellitus with complications	250.4, 250.5, 250.6	1.7	1.7	0.9	2.6
Para/hemiplegia	344.1, 342	0.3	13.0	0.5	1.3
Kidney disease	582, 583.0, 583.1, 583.2, 583.3, 583.4, 583.5, 583.6, 583.7, 585, 586, 588	3.5	1.9	3.5	9.6
Liver/esophageal disease	572.2, 572.3, 572.4, 572.5, 572.6, 572.7, 572.8, 456.0, 456.1, 456.2	0.1	0.2	0.3	1.4
Socioeconomic status:					
Very low income		24.0	24.8	25.4	23.6
Low income		21.3	20.7	20.3	18.7
Middle income		22.7	24.5	28.1	29.9
High income		17.1	15.9	14.1	15.4
Very High income		14.9	14.1	12.1	12.4
Chronicity		3.7	6.7	14.4	15.6
Hemorrhagic stroke	431		11.8		

Note: ICD-9 = International Classification of Diseases 9th revision.
AMI = acute myocardial infarction.

deaths (Hosmer & Lemeshow, 2000). Non–statistically significant probability values in Hosmer and Lemeshow goodness-of-fit tests indicate a good fit between the model and the observed data. Unfortunately, the Hosmer and Lemeshow test is affected by sample size. The large size of the samples in this study including in the models meant the Hosmer and Lemeshow goodness-of-fit tests were likely to show results with small probability values, but those values, even if reflecting minute differences between observed and predicted deaths, were also likely to be statistically significant. For this reason, we evaluated logistic regression goodness-of-fit primarily with the resulting c statistic.

Development of weighted 30-day hospital mortality rates

Patient mix within the four diagnosis categories varied across hospitals. When using indirect standardization methods employing logistic reression models to adjust for patient characteristics, such as those used in this study, differences in patient mix may have a strong confounding effect on outcomes (Shwartz, Ash, & Iezzoni, 1997). Therefore, to adjust for differences in diagnosis mix across hospitals, a weighted standard mortality rate (SMR) was calculated for each study hospital.

First, the number of expected 30-day deaths per hospital diagnosis category was calculated by summing the predicted probability of death for each patient (from the logistic regression models) within each diagnosis category for that hospital. Similarly, the number of actual or observed 30-day deaths was summed within each hospital diagnosis category. Next, a nonweighted 30-day SMR for each hospital's diagnosis category was constructed as the ratio of the sum of actual deaths in that hospital diagnosis category divided by the sum of predicted number of deaths in that same category.

A weighted SMR was next calculated within each hospital diagnosis category, followed by the calculation of a whole-hospital weighted medical SMR, which included patients in all four categories. To accomplish this, the nonweighted specific-diagnosis category SMR for each hospital was multiplied by the proportion of whole-sample cases in that diagnosis category. The values of these proportions are shown in the second-to-last column of Table 1. Finally, a weighted SMR for each hospital was calculated by summing the weighted SMRs in each of the four hospital diagnosis categories. For ease of interpretation, a risk-adjusted hospital mortality rate was calculated by multiplying each hospital's weighted SMR by the crude whole-sample mortality rate of 14.7%.

Results

Regression coefficients

Table 3 lists the 30-day expected mortality regression coefficient, the probability associated with that coefficient, the odds ratio (OR), and the 95% confidence interval associated with the OR for each of the predictors used in the final AMI model. The OR indicates the relative risk of death for patients who have that predictor as a characteristic compared to those patients without that characteristic. The strongest predictor of 30-day AMI mortality was liver cirrhosis, with an OR of 4.47, followed by liver or esophageal disease (OR¼3.53), chronicity (OR¼2.11), and heart failure (OR¼1.50). The area under the ROC curve, or c statistic, in the AMI model was 0.75.

Table 4 shows the 30-day expected mortality regression coefficient, the probability associated with that coefficient, the odds ratio, and the 95% confidence interval associated with the OR for each of the statistically significant predictors used in the final stroke model. The strongest predictor of 30-day stroke mortality was a history of liver or esophageal disease, with an OR of 6.16, followed by hemorrhagic stroke (OR¼5.66), liver cirrhosis (OR¼3.42), and chronicity (OR¼3.18). The area under the ROC curve, or c statistic, in the stroke model was .75.

In Table 5 the 30-day expected mortality regression coefficient, the probability associated with that coefficient, the odds ratio, and the 95% confidence interval associated with the OR for each of the statistically significant predictors used in the final pneumonia model are shown. The strongest predictor of 30-day pneumonia mortality was liver or esophageal disease, with an OR of 12.73, followed by kidney disease (OR¼2.78), liver cirrhosis (OR¼2.47), and chronicity (OR¼2.18). The area under the ROC curve, or c statistic, in the pneumonia model was .76.

The 30-day expected mortality regression coefficient, the probability associated with that coefficient, the odds ratio, and the 95% confidence interval associated with the OR for each of the statistically significant predictors used in the final model, the septicemia model, are shown in Table 6. The strongest predictor of 30-day septicaemia mortality was liver or esophageal disease, with an OR of 6.47, followed by kidney disease (OR¼2.23), liver cirrhosis (OR¼2.21), and heart failure (OR¼2.04). None of the socioeconomic status variables were statistically significant predictors of 30-day death for septicemia patients. The area under the ROC curve, or c statistic, in the septicemia model was .71.

In summary, these six variables were statistically significant predictors of 30-day mortality in all four models: age, heart failure, liver cirrhosis, kidney disease, liver or esophageal disease, and chronicity. This last predictor, chronicity, was a new indicator used as a proxy for patient health status and functioning at the time of admission and was found to have strong predictive capability in all four logistic regression models. Three of the Deyo-adapted Charlson predictors were not statistically significant predictors in any of the models: rheumaticlike diseases, ulcers of the digestive system, and diabetes mellitus with complications. Interestingly, sex was a statistically significant predictor of 30-day mortality only in the pneumonia logistic regression model. Each of the four logistic regression models performed adequately in predicting 30-day postadmission mortality. The model with the greatest area under the ROC curve was the pneumonia model, and the model with the smallest area was the septicemia model. Patient characteristics that put patients at increased risk of death were not identical across the four disease categories.

Risk-adjusted 30-day hospital mortality rates

The mean crude and risk-adjusted 30-day mortality rates for hospitals were 14.95% and 15.03%, respectively. The mean risk-adjusted 30-day mortality rate for the 10 teaching hospitals was 14.0% and for the 65 community hospitals was 15.2%. Table 7 shows the mean, standard deviation, lowest value, and highest value for each of the crude and risk-adjusted mean 30-day mortality rates for the whole sample and separately for the teaching and community hospitals. The mean riskadjusted 30-day mortality rate for teaching hospitals was lower than the mean crude rate. For community hospitals, however, risk adjustment, on average, went in the opposite direction: the mean risk-adjusted 30-day mortality rate for community hospitals was higher than the mean crude rate. The difference in mean risk-adjusted 30-day mortality rates for teaching and community hospitals was not statistically significant at the .05 level. Risk adjustment resulted in changing both the absolute value and rank ordering of hospital mortality rates as compared to crude hospital rates. Pearson's correlation between riskadjusted and crude 30-day hospital mortality rates was .80 ($p < .0001$). Spearman's rank correlation between risk-adjusted and crude 30-day hospital mortality rates was lower, .74 ($p < .0001$).

Discussion

An ideal measure is one that is valid, reliable, and efficient. The quality of risk-adjustment methods is also judged by these characteristics. Although little consensus exists about how to demonstrate the validity of a risk-adjustment method, there is agreement that validity assessment relates to answering the question: How well does the risk-adjustment method account for patients' true risk of 30-day death or for the outcome in question? (Daley, 1997). In efficiency and sensibility, the predictors used in these models were easily generated from the

Table 3 Acute Myocardial Infarction Logistic Regression Model

Predictors	30-Day Regression Coefficient (b weight)	Odds Ratio	95% Confidence Interval	Probability
Age	0.0619	1.06	1.06–1.07	<.0001
AMI	-0.1822	0.83	0.72–0.97	.0179
Heart failure	0.4062	1.50	1.36–1.67	<.0001
Cerebrovascular disease	0.3570	1.43	1.19–1.72	.0002
Respiratory disease	0.1639	1.18	1.02–1.36	.0219
Liver cirrhosis	1.4984	4.47	1.91–10.47	.0005
Diabetes mellitus	0.1336	1.14	1.02–1.28	.0172
Kidney disease	0.5846	1.79	1.48–2.17	<.0001
Liver/esophageal disease	1.2610	3.53	1.13–11.06	.0304
Socioeconomic status*:				
Very low income	0.1629	1.18	1.00–1.38	.0468
Low income	0.2029	1.23	1.04–1.44	.0152
Middle income	0.1340	1.14	0.97–1.35	.1076
High income	0.1865	1.21	1.01–1.44	.0365
Chronicity	0.7454	2.11	1.76–2.52	<.0001

Note: AMI = acute myocardial infarction.
*Reference group for odds ratio is very high income.
Area under the receiving operator characteristic curve is .75. Hosmer and Lemeshow goodness-of-fit test results: Chisquare =23.11, DF=8, p =.0032. The following seven candidate risk factor predictors were not statistically significant predictors of 30-day mortality and were omitted from the final model: sex, peripheral vascular disease, pre/senile psychosis, rheumatic-like diseases, ulcers of digestive system, diabetes mellitus with complications, and para/ hemiplegia.

DAD, from the Ontario Registered Persons Database, and from linkage with the Statistics Canada 1996 population files. In addition, the predictor variables made clinical and practical sense and were consistent with predictors previously discussed in the literature.

Content validity refers to the extent to which model predictors include the universe of risk factors that should be included (Daley, 1997). In this study specific clinical data that may have been strong predictors of mortality were not available. For example, the Acute Physiology and Chronic Health Evaluation (APACHE) is a clinically based severity-of-disease classification system designed to estimate pretreatment risk of death in severely ill patients (Knaus, Draper, Wagner, & Zimmerman, 1986; Knaus et al., 1993). APACHE or similar severity-of-illness evaluation scales generally are not available in large administrative data sets such as the Canadian Discharge Abstract Database or the American Uniform Hospital Discharge Data Set (Iezzoni, 1997a). These administrative databases do not include patient-specific information about potentially important dimensions of risk such as acute clinical status, psychological or cognitive functioning, general health status, and patient attitudes or preferences. Any of these dimensions of risk might have affected 30-day mortality but were not included in the study prediction models (Iezzoni, 1997b). Such clinical data are

Table 4 Stroke Logistic Regression Model

Predictors	30-Day Regression Coefficient (b weight)	Odds Ratio	95% Confidence Interval	Probability
Age	0.0445	1.05	1.04–1.05	<.0001
AMI	0.2016	1.22	1.00–1.49	.0456
Heart failure	0.7760	2.17	1.81–2.61	<.0001
Liver cirrhosis	1.2295	3.42	1.65–7.08	.0009
Para/hemiplegia	-0.3608	0.70	0.59–0.82	<.0001
Kidney disease	0.6869	1.99	1.44–2.74	<.0001
Liver/esophageal disease	1.8176	6.16	2.47–15.34	<.0001
Socioeconomic status*:				
Very low income	0.1633	1.18	0.98–1.41	.0754
Low income	0.1815	1.20	1.00–1.44	.0554
Middle income	0.2094	1.23	1.03–1.48	.0223
High income	0.2493	1.28	1.06–1.56	.0119
Chronicity	1.1570	3.18	2.69–3.76	<.0001
Hemorrhagic stroke	1.7330	5.66	4.96–6.45	<.0001

Note: AMI=acute myocardial infarction.
*Reference group for odds ratio is very high income.
Area under the receiving operator characteristic curve is .75. Hosmer and Lemeshow goodness-of-fit test results: Chisquare = 35.77, DF=8, p < .0001. The following nine candidate risk factor predictors were not statistically significant predictors of 30-day mortality and were omitted fromthe final model: sex, peripheral vascular disease, cerebrovascular disease, pre/senile psychosis, respiratory disease, rheumatic-like diseases, ulcers of digestive system, diabetes mellitus, and diabetes mellitus with complications.

Table 5 Pneumonia Logistic Regression Model

Predictors	30-Day Regression Coefficient (b Weight)	Odds Ratio	95% Confidence Interval	Probability
Age	0.0535	1.06	1.05–1.06	<.0001
Sex (male)	0.2268	1.26	1.14–1.39	<.0001
Heart failure	0.3034	1.36	1.21–1.52	<.0001
Cerebrovascular disease	0.3765	1.46	1.23–1.73	<.0001
Pre/senile psychosis	0.4038	1.50	1.26–1.78	<.0001
Liver cirrhosis	0.9033	2.47	1.37–4.45	.0027
Kidney disease	1.0239	2.78	2.27–3.41	<.0001
Liver/esophageal disease	2.5437	12.73	6.69–24.22	<.0001
Socioeconomic status*:				
Very low income	0.1229	1.13	0.94–1.37	.2023
Low income	0.2199	1.25	1.03–1.51	.0259
Middle income	0.2932	1.34	1.12–1.61	.0017
High income	0.1591	1.17	0.95–1.45	.1381
Chronicity	0.7774	2.18	1.93–2.46	<.0001

Note: *Reference group for odds ratio is very-high-income socioeconomic status.
Area under the receiving operator characteristic curve is .76. Hosmer and Lemeshow goodness-of-fit test results: Chisquare = 35.30, DF = 8, p < .0001. The following eight candidate risk factor predictors were not statistically significant predictors of 30-day mortality and were omitted from the final model: acute myocardial infarction, peripheral vascular disease, respiratory disease, rheumatic-like diseases, ulcers of digestive system, diabetes mellitus, diabetes mellitus with complications, and para/hemiplegia.

generally collected for specific studies or projects.

Average income for postal code areas, the indicator used to estimate socioeconomic status, was not a precise measure of actual income. Using a more precise measure such as actual income might improve the validity of the method. However, this level of data is not accessible unless collected directly from each patient. In the American context, source of pay or race and ethnicity might also be used as a proxy indicator for socioeconomic status. However, indices of socioeconomic status such as those developed in this study may also be developed in the U.S. context by linking patient zip codes with the most recent population statistics files. In this study predictive validity referred to how well the risk-adjustment method used predicted 30-day death. It is preferable to assess predictive validity of a method by comparing results between the data set used to develop the model and the data set used for model validity because risk-adjustment models typically perform better in the data set used to develop the model (Daley, 1997). Our logistic regression models were developed a priori and applied directly to the study data set. The models were then refined using the same data set but only including those predictors that were statistically significant. In this study c-statistic results for the four models ranged from .71 to .76, indicating that the models were effective, but not perfect, in predicting 30-day death. In the future we will test these logistic regression models with patients discharged from Ontario hospitals during the 2002–2003 fiscal year, thereby further evaluating the models' predictive ability. A strategy that might have been undertaken to evaluate predictive validity of the models would have been to develop and test the logistic regression models on a sample of patients who met the same criteria but were hospitalized during a previous year and then apply the models to patients in sample hospitals for the study year. This way we could have evaluated the performance of the

Table 6 Septicemia Logistic Regression Model

Predictors	30-Day Regression Coefficient (b-Weight)	Odds Ratio	95% Confidence Interval	Probability
Age	0.0295	1.03	1.02–1.04	<.0001
Heart failure	0.7120	2.04	1.49–2.80	<.0001
Peripheral vascular disease	0.7445	2.11	1.23–3.62	.0069
Pre/senile psychosis	0.5206	1.68	1.13–2.51	.0105
Liver cirrhosis	0.7949	2.21	1.00–4.89	.0495
Kidney disease	0.8033	2.23	1.62–3.09	<.0001
Liver/esophageal disease	1.8666	6.47	3.05–13.71	<.0001
Chronicity	0.6306	1.88	1.43–2.47	<.0001

Note: Area under the receiving operator characteristic curve is .71.
Hosmer and Lemeshow goodness-of-fit test results: Chi-square=12.88, DF=8, p =.1161. The following 13 candidate risk factor predictors were not statistically significant predictors of 30-day mortality and were omitted from the final model: sex, acute myocardial infarction, cerebrovascular disease, respiratory disease, rheumatic-like diseases, ulcers of digestive system, diabetes mellitus, diabetes mellitus with complications, para/hemiplegia, and all socioeconomic indicators.

Table 7 Mean Crude and Risk-adjusted 30-Day Mortality Rates (as percentages) for Entire Sample of Hospitals and for Teaching and Community Hospital

Hospital Grouping	Mean Rate	Standard Deviation	Lowest Value (%)	Highest Value (%)
All hospitals: crude (n =75)	14.95	2.43	10.26	20.76
All hospitals: risk-adjusted	15.03	2.28	10.53	21.53
Teaching: crude (n =10)	14.68	1.52	12.54	16.38
Teaching: risk-adjusted	14.02	1.29	11.75	15.48
Community: crude (n =65)	14.99	2.55	10.26	20.76
Community: risk-adjusted	15.18	2.36	10.53	21.53

models by comparing c-statistic results between the test and study years.

The reliability of the risk-adjustment method used in this study was largely dependent on the accuracy, consistency, and completeness of information gathered from patient medical records and coded in the DAD (Hughes & Ash, 1997). In Canada the Canadian Institute for Health Information (CIHI) collects abstracted data on patients discharged from the majority of Canadian hospitals. A record for each individual patient encounter within an Ontario hospital is abstracted in the DAD. Canadian hospitals have uniform data fields to complete for the DAD, and they submit these data to CIHI. After editing by CIHI, these data are returned to the hospitals and provincial health departments for their use. As a result, abstracting and coding standards exist across Canadian hospitals. These standards promote the reliability and validity of the data and strengthen investigations of comparisons among hospitals (Canadian Institute for Health Information, 1995). Despite data standards, there are potential sources of variation that could affect the overall reliability of data and, consequently, the risk-adjustment models used. Sources of variation may involve differences in coding practices among hospitals, including decisions about which ICD-9 codes to use, differences in strategies used among hospitals to manage data, and differences in the training and motivation of raters or coders.

Direct measurement of the reliability of the DAD used in this study was not completed but was inferred from other studies addressing the reliability of these data. Background patient information such as admission and discharge dates, sex, death codes, birth dates, postal codes, and so on have been shown to be abstracted with high accuracy (The Doctor's Hospital, 1992; Hawker, Coyte, Wright, Paul, & Bombardier, 1997; Ontario Hospital Association et al., 1991; Richards et al., 2001). Lower levels of agreement, ranging from 74% to 95%, have been reported for accuracy in identification of the most responsible diagnosis (Delfino, Becklake, & Hanley, 1993; Newfoundland Department of Health, 1995; Ontario Hospital Association et al., 1991; Richards et al., 2001).

The most recent national reabstraction study of the reliability of coding of the most responsible diagnosis in the DAD for the two fiscal years 1999–2000 and 2000–2001 found 87% agreement between initial hospital coding and study reabstracting coding (Canadian Institute for Health Information, 2002). In a study involving the accuracy of coding for acute myocardial infarction, 100% accuracy at some hospitals was reported (Tu, Naylor, & Austin, 1999). Accuracy in coding stroke ICD-9 codes was reported to range from 70% to 80% of reviewed patient medical records, with varying levels of discrepancies in the use of the correct stroke code (Mayo, Danys, Carlton, & Scott, 1993). Studies reporting the accuracy and completeness in abstracting and coding complications and comorbid conditions have reported a wider spread of agreement scores, ranging between 37% and 95% (Malenka, McLerran, Roos, Fisher, & Wennberg, 1994; Newfoundland Department of Health, 1995; Ontario Hospital Association et al., 1991; Richards et al., 2001; Victoria Hospital, 1992). The most recent national reabstraction study of the reliability of coding of comorbid conditions in the DAD for the two fiscal years 1999–2000 and 2000–2001 found 77% agreement between initial hospital coding and study reabstracting coding (Canadian Institute for Health Information, 2002).

In Canada, health care is under the jurisdiction of provincial governments, and most provinces have mandated uniform reporting on DAD fields to CIHI for their hospitals. Each provincial jurisdiction also keeps vital statistic files that contain information similar to the Ontario Registered Persons Database such as birth and death dates of its residents. Finally, Statistics Canada population files, including average income data, exist for virtually every Canadian community, facilitating the development of socioeconomic indicators such as those developed in this study. As similar data is available in the U.S. UHDDS, this process also may be implemented to calculate risk-adjusted mortality outcomes. These risk-adjusted 30-day mortality rates can be further used in prediction models such as those investigating the nursing-related determinants of 30-day mortality (Tourangeau et al., 2002).

This risk-adjustment procedure may be used to calculate risk-adjusted mortality rates at different levels of analyses such as nursing units and regions. However, it is important to realize that it is the process or procedure that is transportable rather than the actual patient characteristics that put one at risk for death or whatever outcome is being studied (Daley & Shwartz, 1997). There is no known perfect set of risks for death for all hospitalized patients, although risk factors are generally similar for those patients with a specific presenting condition. The same argument can be made when developing risk-adjustment procedures to control for patient characteristics for other outcomes of interest such as unplanned readmission and functional status. Both the outcome being examined and the condition of the patient determine what patient characteristics affect that patient outcome. Risk factors are identified through evidence found in research literature and through clinical expertise (Iezzoni, 1997b).

The use of administrative data has limitations, but it is likely

that future outcomes researchers will continue to use administrative data in prediction models. More research and scholarly debate is needed in the area of risk adjustment in outcomes research. For example, a better understanding is needed of the effects that differences in the number of patients between hospitals used in regression models have on estimates of standard hospital mortality rates. Further testing and dissemination of risk-adjustment models are needed for a variety of patient outcomes. Although we have suggested a two-step approach to outcomes research that involves distinct risk-adjustment procedures before other analyses such as regressing the risk-adjusted outcomes on hospital characteristic predictors, there are alternate strategies. One alternate modeling strategy incorporates random effects or hierarchical modeling. We advocate further research to compare results of various strategies.

The results of this risk-adjustment method suggest that adequate predictions of 30-day mortality can be achieved through use of routinely collected data in administrative health databases. When effective risk-adjustment methods are used to account for patient characteristics within and across a particular study's units of analysis (e.g., hospitals), study internal validity is strengthened, and alternate explanations related to the effects of patient characteristics on the outcome become less plausible. This risk-adjustment procedure can be generalized to other Canadian and American jurisdictions.

References

Ash, A.S., & Shwartz, M. (1997). Evaluating the performance of risk-adjusted methods: Dichotomous outcomes. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes (pp. 427–469). Chicago: Health Administration Press.

Baker, R., Anderson, G.M., Brown, A.D., McKillop, I., Montgomery, C., Murray, M.A., et al. (1999). The hospital report 99: A balanced report card for Ontario acute care hospitals. Toronto, Ontario, Canada: Ontario Hospital Association. 494 RESEARCH IN NURSING & HEALTH

Barnsley, J., Lemieux-Charles, L., & Baker, R. (1996). Selecting clinical outcome indicators for monitoring quality of care. Healthcare Management Forum, 9, 5–12.

Bland, M. (1999). An introduction to medical statistics. Oxford, Great Britain: Oxford University Press. Brooten, D., & Naylor, M.D. (1995). Nurses' effect on changing patient outcomes. Image: Journal of Nursing Scholarship, 27, 95–99.

Burns, M., Moores, P., & Breslin, E. (1996). Outcomes research: Contemporary issues and historical significance for nurse practitioners. Journal of the American Academy of Nurse Practitioners, 8, 107–112.

Canadian Institute for Health Information. (1995). CIHI abstracting manual. Ottawa, Ontario, Canada: Canadian Institute for Health Information. Canadian Institute for Health Information. (2002). Discharge abstract database data quality re-abstraction study: Combined findings for fiscal years 1999/ 2000 and 2000/2001. Ottawa, Ontario, Canada: Canadian Institute for Health Information.

Charlson, M.E., Pompei, P., Ales, K.L., & MacKenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Disease, 40, 373–383.

Daley, J. (1997). Validity of risk-adjusted methods. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes (pp. 331–363). Chicago: Health Administration Press.

Daley, J., & Shwartz, M. (1997). Developing riskadjustment methods. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes (pp. 279–329). Chicago: Health Administration Press.

Delfino, R.J., Becklake, M.R., & Hanley, J.A. (1993). Reliability of hospital data for population-based studies of air pollution. Archives of Environmental Health, 48, 140–146.

Deyo, R.A., Cherkin, D.C., & Ciol, M.A. (1992). Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. Journal of Clinical Epidemiology, 45, 613–619.

The Doctors' Hospital. (1992). Data quality and utilization study. Toronto, Ontario, Canada: The Doctor's Hospital.

French, B. (1997). British studies which measure patient outcome, 1990–1994. Journal of Advanced Nursing, 26, 320–328.

Green, J., Passman, L.J., & Wintfeld, N. (1991). Analyzing hospital mortality: The consequences of diversity in patient mix. JAMA, 265, 1849–1853.

Green, J., Wintfeld, N., Sharkey, P., & Passman, L.J. (1990). The importance of severity of illness in assessing hospital mortality. JAMA, 263, 241–246.

Gregario, D.I., Walsh, S.J., & Paturzo, D. (1997). The effects of occupation-based social position on mortality in a large American cohort. American Journal of Public Health, 87, 1472–1475.

Griffiths, P. (1995). Progress in measuring outcomes. Journal of Advanced Nursing, 21, 1092–1100.

Hawker, G.A., Coyte, P.C., Wright, J.G., Paul, J.E., & Bombardier, C. (1997). Accuracy of administrative data for assessing outcomes after knee replacement surgery. Journal of Clinical Epidemiology, 50, 265– 273.

Hegyvary, S.T. (1991). Issues in outcome research. Journal of Nursing Quality Assurance, 5(2), 1–6.

Hosmer, D.W., & Lemeshow, S. (2000). Applied logistic regression (2nd ed.) New York: John Wiley & Sons.

Hughes, J.S., & Ash, A.S. (1997). Reliability of riskadjustment methods. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes. (pp. 365–390) Chicago: Health Administration Press.

International classification of diseases (9th revision). (1999). Los Angeles: Practice Management Information Corporation.

Iezzoni, L.I. (1997a). Data sources and implications: Administrative databases. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes (pp. 169–242). Chicago: Health Administration Press.

Iezzoni, L.I. (1997b). Dimensions of risk. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes (pp. 43–167). Chicago: Health Administration Press.

Iezzoni, L.I. (1997c). Risk and outcomes. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes (pp. 1–41). Chicago: Health Administration Press.

Iezzoni, L.I. (1997d). The risks of risk adjustment. JAMA, 278, 1600–1607.

Jeffs, L. (1998). Proving what you practice: The role of outcomes research. Registered Nurse Journal: Registered Nurses Association of Ontario, 10(5), 5–6.

Jencks, S.F., Daley, J., Draper, D., Thomas, N., Lenhart, G., & Walker, J. (1988). Interpreting hospital mortality: The role of clinical risk adjustment. JAMA, 260, 3611–3616.

Jencks, S.F., Williams, D.K., & Kay, T.L. (1988). Assessing hospital-associated deaths from discharge data: The role of length of stay and comorbidities. JAMA, 260, 2240–2246.

Jennings, B.M., &Staggers, N. (1998). The language of outcomes. ANS: Advances in Nursing Science, 20, 72–80.

Knaus, W.A., Draper, E.A., Wagner, D.P., & Zimmerman, J.E. (1986). An evaluation of outcome from intensive care in major medical centers. Annals of Internal Medicine, 104, 410–418.

Knaus, W.A., Wagner, D.P., Zimmerman, J.E., & Draper, E.A. (1993). Variations in mortality and length of stay in intensive care units. Annals of Internal Medicine, 118, 753–761.

Lamb, G.S. (1997). Outcomes across the care continuum. Medical Care, 35, NS106–NS114.

Malenka, D.J., McLerran, D., Roos, N., Fisher, E.S., & Wennberg, J.E. (1994). Using administrative data to

describe casemix: A comparison with the medical record. Journal of Clinical Epidemiology, 47, 1027– 1032.

Mark, B.A. (1995). The black box of patient outcomes research. Image: Journal of Nursing Scholarship, 27, 42.

RISK-ADJUSTMENT METHODOLOGY / TOURANGEAU ANDTU 495

Mark, B.A., & Burleson, D.L. (1995). Measurement of patient outcomes. Journal of Nursing Administration, 25(4), 52–59.

Mayo, N.E., Danys, I., Carlton, J., & Scott, S.C. (1993). Accuracy of hospital discharge coding for stroke. Canadian Journal of Cardiology, 9, 121D–124D.

Newfoundland Department of Health. (1995). Report of the Newfoundland reabstracting project. St. John's, Newfoundland, Canada: Newfoundland Department of Health.

O'Brien-Pallas, L.L., Irvine, D., Peereboom, E., & Murray, M. (1997). Measuring nursing workload: Understanding the variability. Nursing Economic$, 15(4), 171–182.

Ontario Hospital Association, Ontario Ministry of Health, & Hospital Medical Records Institute. (1991). Report of the Ontario data quality reabstracting study. Toronto, Ontario, Canada: Ontario Hospital Association.

Pappas, G., Queen, S., Hadden,W., & Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1966 and 1986. TheNewEngland Journal ofMedicine, 329, 103–109.

Richards, J., Brown, A., & Homan, C. (2001). The data quality study of the Canadian discharge abstract

database. In Proceedings of Statistics Canada Symposium 2001: Achieving Data Quality in a Statistical Agency. Ottawa, Ontario, Canada: Statistics Canada.

Shwartz, M., Ash, A.S., & Iezzoni, L.I. (1997). Comparing outcomes across providers. In L.I. Iezzoni (Ed.), Risk adjustment for measuring healthcare outcomes (pp. 471–516). Chicago: Health Administration Press.

Sidani, S., & Braden, C.J. (1998). Evaluating nursing interventions: A theory driven approach. Thousand Oaks, CA: Sage.

Silber, J.H., & Rosenbaum, P.R. (1997). A spurious correlation between hospital mortality and complication rates. Medical Care, 35(10), OS77–OS92. Silber, J.H., Rosenbaum, P.R., & Ross, R.N. (1995). Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? Journal of the American Statistical Association, 90(429), 7–18.

Silber, J.H., Williams, S.V., Krakauer, H., & Schwartz, J.S. (1992). Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure to rescue. Medical Care, 30, 615–628.

Smith, D.W. (1994). Evaluating risk adjustment by partitioning variation in hospital mortality rates. Statistics in Medicine, 13, 1001–1013.

Tourangeau, A.E., Giovannetti, P., Tu, J.V.,&Wood, M. (2002). Nursing-related determinants of 30-day

mortality for hospitalized patients. Canadian Journal of Nursing Research, 33(4), 71–88.

Tu, J.V., Naylor, C.D., & Austin, P. (1999). Temporal changes in the outcomes of acute myocardial

infarction in Ontario, 1992–1996. Canadian Medical Association Journal, 161, 1257–1261.

Victoria Hospital. (1992). CMG coding validation study. London, Ontario, Canada: Victoria Hospital.

Wood, M.J., & Brink, P.J. (1998). Comparative designs. In P.J. Brink&M.J.Wood (Eds.), Advanced design in

nursing research (pp. 143–159). Thousand Oaks, CA: Sage.