Multiple Approaches to Assessing the Effects of Delays for Hip Fracture Patients in the United States and Canada

Vivian Ho, Barton H. Hamilton, and Leslie L. Roos

Objective. To examine the determinants of postsurgery length of stay (LOS) and inpatient mortality in the United States (California and Massachusetts) and Canada (Manitoba and Quebec).

Data Sources/Study Setting. Patient discharge abstracts from the Agency for Health Care Policy and Research Nationwide Inpatient Sample and from provincial health ministries.

Study Design. Descriptive statistics by state or province, pooled competing risks hazards models (which control for censoring of LOS and inpatient mortality data), and instrumental variables (which control for confounding in observational data) were used to analyze the effect of wait time for hip fracture surgery on postsurgery outcomes.

Data Extractions. Data were extracted for patients admitted to an acute care hospital with a primary diagnosis of hip fracture who received hip fracture surgery, were admitted from home or the emergency room, were age 45 or older, stayed in the hospital 365 days or less, and were not trauma patients.

Principal Findings. The descriptive data indicate that wait times for surgery are longer in the two Canadian provinces than in the two U.S. states. Canadians also have longer postsurgery LOS and higher inpatient mortality. Yet the competing risks hazards model indicates that the effect of wait time on postsurgery LOS is small in magnitude. Instrumental variables analysis reveals that wait time for surgery is not a significant predictor of postsurgery length of stay. The hazards model reveals significant differences in mortality across regions. However, both the regressions and the instrumental variables indicate that these differences are not attributable to wait time for surgery.

Conclusions. Statistical models that account for censoring and confounding yield conclusions that differ from those implied by descriptive statistics in administrative data. Longer wait time for hip fracture surgery does not explain the difference in postsurgery outcomes across countries.

Key Words. Surgery delay, hip fracture, competing risks hazards, instrumental variables

Policymakers and researchers have long been debating the relative merits of the U.S. and Canadian healthcare systems. Numerous comparisons of the quality of healthcare services and health outcomes have been based on administrative data that describe patient care in both of these countries. However, the conclusions that can be drawn from such studies are limited by the observational nature of the data. For instance, comparisons of inpatient mortality are confounded by longer lengths of stay in Canada versus the United States, which leads to unequal probabilities of observing in-hospital death across countries. In addition, controls for inpatient case mix are limited to information on the number and types of comorbidities listed in the medical abstract. Thus, it is difficult to conclude whether differences in outcomes between the United States and Canada are truly due to differences in healthcare systems or to undocumented differences in case mix between the two countries.

This article proposes methods for comparing treatment effects in the United States and Canada when analyzing inpatient administrative data from these two countries. We specifically address the issues of right-censored inpatient mortality data as well as that of incomplete information on patient case mix. Although we seek to assess the effectiveness of healthcare in Canada versus the United States, the methods are readily generalizable to other cases in which one seeks to identify the determinants of inpatient mortality and hospital length of stay.

Our case study is a comparison of surgical queues and outcomes between the United States and Canada after a patient has fractured his or her hip. Critics of the Canadian healthcare system claim that universal health insurance has led to unacceptable delays in obtaining healthcare in Canada. Although the existence of queues has been widely documented (Globerman 1991; U.S. Government Accounting Office 1991; Katz, Mizgala, and Welch

---

This work was supported by HEALNet (Health Evidence Application and Linkage Network—A Canadian Network of Centres of Excellence) and by a National Health Scientist Award (6607-1314-48) from the National Health Research and Development Program.

Address correspondence to Vivian Ho, Ph.D., Assistant Professor, Economics and Management, John M. Olin School of Business, Washington University in St. Louis, Campus Box 1133, One Brookings Drive, St. Louis MO 63130-4899. Barton H. Hamilton, Ph.D. is Assistant Professor, Economics and Management, the John M. Olin School of Business, Washington University in St. Louis; and Leslie L. Roos, Ph.D. is Professor, the Manitoba Centre for Health Policy and Evaluation, Dept. of Community Health Sciences, Faculty of Medicine, University of Manitoba.
This article, submitted to *Health Services Research* on March 5, 1998, was revised and accepted for publication on March 1, 1999.

1991), there is little evidence that the queues observed in Canada detrimentally affect clinical outcomes relative to those in the United States.

We focus on queues for hip fracture patients for three reasons. First, hip fractures are common and are, in fact, the leading cause of hospitalization for injuries among the elderly (Baker, O'Neill, and Karpf 1984). Second, hip fracture is easy to diagnose, with little ambiguity about coding or the need for hospitalization. Thus, patient populations presenting for hip fracture surgery will be similar across sites (Roos et al. 1996). Third, there is a strong a priori presumption that surgical queues for hip fracture will affect outcomes. Unlike queuing for many other surgical procedures, treatment for hip fracture is urgent. Patients who fracture their hips are immobile and must remain in the hospital, in traction and on medication, until they receive surgery. Prolonged immobility resulting from surgery delay can lead to complications that are potentially fatal (Sabiston 1991). In fact, several studies in the orthopedic literature find a detrimental effect of surgery delays on postsurgery outcomes (Bredahl, Nyholm, Hindsholm, et al. 1992; Davis, Woolner, Frampton, et al. 1987; Hoerer, Volpin, and Stein 1993; Sexson and Lehner 1988). However, these studies tend to have relatively small sample sizes, or they control for only a limited number of covariates. Thus, hip fractures represent an important case study for comparing queues and clinical outcomes between Canada and the United States.

We analyzed data from administrative databases from California, Massachusetts, Quebec, and Manitoba to obtain information on large samples of patients in both the United States and Canada. Competing risks hazards regression models are used to analyze the effect of surgery wait times on postsurgery length of stay and inpatient mortality. We introduce parameters in the hazards model that account for unobserved differences in health status across patients which are not captured in the administrative data. We then apply instrumental variables analysis to examine the robustness of the conclusions regarding surgery delays and postsurgery outcomes.

METHODS

Data Sources

Data for the United States are drawn from the Agency for Health Care Policy and Research (AHCPR) Healthcare Cost and Utilization Project (HCUP-3) Nationwide Inpatient Sample (NIS). The NIS is designed to approximate a 20 percent sample of U.S. community hospitals as defined by the American

Hospital Association (AHA).[1] For each general acute care hospital sampled in California and Massachusetts, we have discharge abstract information on 100 percent of patients discharged during the relevant calendar year.[2] Data for Quebec are drawn from the Quebec Ministry of Health and Social Services MED-ÉCHO database, which contains discharge abstracts for all inpatients in acute care hospitals in the province. Information for Manitoba is drawn from the Manitoba Health file, which also contains discharge abstracts for all inpatients in acute care hospitals in the province.

Study Population

The study population consists of patients discharged between 1990 and 1992 in California (n = 20,025), Massachusetts (n = 11,692), and Quebec (n = 13,555), and between 1990 and 1994 in Manitoba (n = 4,257). All patients admitted to acute care hospitals with a primary diagnosis of hip fracture (International Classification of Diseases, 9th Revision [ICD-9-CM], diagnosis codes 820.0–820.9) who were age 45 or older and whose length of stay in the hospital was 365 days or less were initially included in the sample. Patients who fractured their hip due to multiple trauma (e.g., motor vehicle accident) are more likely to receive expedited surgery (i.e., hip fracture is not the primary health problem). Because these cases are typically handled in a special manner, we excluded them from the analysis. Thus, patients documented as having had head trauma (ICD-9-CM codes 860–869) were excluded. In addition, we chose to study patients admitted only from their home or the emergency room, because total wait times for surgery for transferring patients were not available. The only exception to this exclusion was Manitoba, where we had information on length of stay in the transferring hospital. In this province, patients with a prior length of stay greater than one day in a transferring hospital were excluded. These exclusions led to patient sample sizes of 16,948 in California, 8,971 in Massachusetts, 10,006 in Quebec, and 3,591 in Manitoba. In each state or province, over 90 percent of patients admitted with a primary diagnosis of hip fracture eventually received hip surgery (n = 15,711, 93 percent in California; n = 8,202, 91 percent in Massachusetts; n = 10,013, 91 percent in Quebec; n = 3,364, 94 percent in Manitoba). Only data on patients who receive surgery are used in the following analysis.

Covariates

The two dependent variables are postsurgery length of stay and inpatient mortality (whether the patient was discharged from the hospital alive or dead). The primary explanatory variable of interest is wait time until surgery. This variable is constructed by calculating the amount of time that elapsed between the date of admission to the hospital and the date of hip fracture surgery, measured in days. Age, gender, state or province of residence, and hip fracture type are also included as explanatory variables. These covariates are represented categorically, with age dummy variables for those ages 65 to 79 versus younger patients and for those patients age 80 and over versus younger patients. A dummy variable indicator distinguishes patients with a pertrochanteric fracture from other hip fracture patients. Indicator variables for the individual comorbidities used to construct the Charlson comorbidity index are included as regressors to control for observable differences in health status across patients (Romano, Roos, and Jollis 1993). The Charlson index derived from the weighted aggregation of these individual dummy variables is also presented in the descriptive statistics and instrumental variables analysis. Both the indicator variables and the Charlson index were constructed using a coding methodology developed specifically for administrative data (Romano, Roos, and Jollis 1993). The number of hip fracture surgeries performed in the patient's hospital during the year of admission is also included as a regressor. This variable is included to test for the positive relationship between surgical volume and more favorable postsurgical outcomes identified in previous studies (Luft et al. 1990; Hannan, Kilburn, Bernard, et al. 1991; Hamilton and Hamilton 1997).

Modeling Strategies

Multivariate regression analysis is required to assess the effect of wait time for surgery on postsurgery length of stay and inpatient mortality, controlling for variations in other factors that may also affect patient outcomes. The regression is estimated on pooled data, so that we can also test for significant differences in patient outcomes across the four regions.

The regression is specified using a survival analysis framework, because the dependent variables are time until discharge for patients discharged alive, and time until death for patients who die in the hospital (Kalbfleisch and Prentice 1980). Note that these two dependent variables censor observable data on each other. For example, if patients in the United States have shorter postsurgery lengths of stay than Canadian patients, then one will also be less likely to observe inpatient mortality among U.S. patients. We control for this censoring by estimating a competing risks hazards model (Lancaster 1990; Hamilton, Hamilton, and Mayo 1996; Hamilton and Hamilton 1997). This model estimates the conditional (on time in the hospital) probability of live discharge and the conditional probability of a dead discharge, controlling for right censoring of each variable by the other "competing" outcome. These conditional probabilities are referred to as "transition intensities" in the duration literature.

Although a number of covariates are included in the model to account for differences in patient characteristics, unobservable differences among patients that affect the probability of live discharge and inpatient mortality may still remain. For example, the administrative data may lack sufficient detail to identify frailer patients who are less likely to be discharged alive and are more likely to die in the hospital as well (Iezzoni, Foley, Daley, et al. 1992). An approach used in the duration literature to account for unobserved (in the data) patient heterogeneity is to specify the transition intensities as dependent on a univariate random variable, as well as on observable patient characteristics (Lancaster 1990; Pickles and Crouchley 1995; Hamilton, Hamilton, and Mayo 1996; Hamilton and Hamilton 1997). We assume that this random variable takes on two discrete values to be estimated in the data (Heckman and Singer 1984). Intuitively, this random variable captures systematic differences in the error term of the regression model that may remain after controlling for observable covariates. We allow this random variable to enter the death and live-discharge equations with different factor loadings so that the unobserved risk factors may affect the two outcomes in different ways.

The resulting competing risks model with unobserved patient heterogeneity may be estimated using maximum likelihood techniques. The likelihood function is estimated using a proportional hazards specification. That is, observable covariates and unobserved patient heterogeneity shift the transition intensity above or below its baseline. The baseline hazard is assumed to follow a log-logistic distribution (Hamilton and Hamilton 1997). A detailed description of the likelihood function appears in the appendix.

Instrumental Variables Analysis

In order to examine the robustness of our results, we also analyze the effect of wait time for surgery on postsurgery outcomes using instrumental variables analysis. This approach has been applied previously to examine the effect of intensive treatment of acute myocardial infarction when analyzing observational data (McClellan, McNeil, and Newhouse 1997). Observational data are problematic when one analyzes the impact of a treatment on outcomes, because differences in health status may determine which treatment patients receive. These variations in health status may also systematically affect outcomes, making it difficult to distinguish variations in outcomes that are due to treatment effects versus health status. Instrumental variables analysis attempts to isolate the treatment effect in observational data and to circumvent the potential confounding effect of health status.

This approach requires the identification of an "instrumental variable" (IV)—a variable that is correlated with wait time for surgery, but a priori is not believed to affect postsurgery outcomes directly. Patients are then grouped by their reported value of this IV. If the IV is valid, then the correlation between the IV and postsurgery outcomes reveals the relationship between wait time for surgery and outcomes, purged of confounding factors such as patient comorbidities. This correlation may differ both in magnitude and significance relative to basic descriptive statistics on the relationship between surgery delay and postsurgery outcomes.

In this case we test the hypothesis that the treatment effect—surgery delay—has no significance for postsurgery outcomes. Thus, we must compare differences in postsurgery outcomes across patients with different values of the IV. If the patient groups differ significantly in wait times for surgery but their postsurgery outcomes are not significantly different, then one may conclude that wait time for surgery does not have a significant effect on postsurgery outcomes.

We use day of the week of admission to the hospital as an instrument for wait time for surgery. Surgical staff may prefer to operate on weekdays rather than on weekends, which will systematically affect the distribution of wait times across days of the week. However, day of the week of admission is less likely to affect patient outcomes. We therefore test for significant differences in the fraction of patients delayed for surgery using a one-way analysis of variance test. For the IV analysis, patients hospitalized three or more days prior to surgery were classified as having their operations delayed (Roos et al. 1996). We then test for significant differences in the fraction of patients who died in the hospital and their mean postsurgery length of stay by day of the week of admission using one-way analysis of variance tests.

One may be concerned that hospitals have lower levels of nursing staff or other ancillary services on weekends, so that day of the week of admission might also directly affect patient outcomes. However, we focus on the "extreme" hypothesis that day of the week of admission has no effect at all on patient outcomes. Thus, if we find that day of the week of admission is correlated with wait time for surgery, but not with postsurgery outcomes, then the IV is indeed valid.

RESULTS

Descriptive Findings

Table 1 presents descriptive statistics on wait times for surgery and postsurgery outcomes in each of the four regions being compared. Mean wait times are longer in the two Canadian provinces (delay = 3.32 days in Manitoba and 3.06 days in Quebec.) than in the two U.S. states (delay = 2.60 days in Massachusetts and 2.17 days in California). Moreover, inpatient mortality is over 5 percent in both Canadian provinces while it is under 5 percent in the United States; and postsurgery lengths of stay are close to 30 days in Manitoba and Quebec, and well under 20 days in the United States.

Within each region, longer surgery delays also appear to be associated with poorer postsurgery outcomes. For example, Manitoba patients who wait six or more days for surgery are twice as likely to die in the hospital relative to patients who undergo surgery within one day of admission. In addition, their postsurgery lengths of stay are over 15 days longer. Thus, comparisons of wait times and postsurgery outcomes both across regions and within regions give the impression that surgery delays harm patient outcomes.

However, in each region, hip fracture patients with longer surgery delays also record higher values of the Charlson comorbidity index. In each state or province the Charlson index is at least 30 percent higher for patients who wait six or more days for surgery relative to patients operated on within one day of admission. Thus, the poorer outcomes of patients who were delayed for surgery may be due to poorer health status at the time of admission to the hospital.

Note also that longer postsurgery lengths of stay in Canada imply that one is more likely to observe inpatient mortality in Manitoba and Quebec than in the U.S. states. This correlation between length of stay and inpatient mortality can be controlled for in a competing risks hazards model.

Also, the mean value of the Charlson index is higher in Massachusetts and California (.63 and .60, respectively) than in Manitoba and Quebec (.48 and .55, respectively). Given that the age and sex distributions and the percentage of pertrochanteric fractures are similar across regions, little reason exists for health status to differ. However, U.S. hospitals have an incentive to code as many comorbidities as possible in order to increase reimbursement under the Medicare DRG system; no such incentive exists in Canada. Thus, observable data on comorbidity status may not fully account for differences in health status that contribute to surgery delay.

Table 1: Relationship Between Wait Time Until Surgery, Case Mix, and Outcomes for Patients Undergoing Surgery

Surgery Delay (days)	*Percent*	*Charlson Index (Mean)*	*Inpatient Mortality (%)*	*Postsurgery Length of Stay*	
				Mean	*Median*
Manitoba ($N = 3{,}364$) mean delay = 3.32 days					
1	13	.31	4.7	28.11	16
2	41	.39	4.8	28.86	16
3	25	.54	6.0	30.67	17
4	9	.64	6.7	33.03	19
5	4	.70	6.4	33.34	23
6+	8	.68	9.9	46.15	26
Total	100	.48	5.7	31.10	17
Quebec ($N = 10{,}013$) mean delay = 3.06 days					
1	24	.48	7.0	26.30	16
2	42	.52	7.7	27.16	18
3	17	.58	9.3	28.35	19
4	7	.57	8.8	28.06	19
5	3	.71	13.9	35.60	21
6+	7	.80	12.2	37.89	23
Total	100	.55	8.5	28.22	18
Massachusetts ($N = 8{,}202$) mean delay = 2.60 days					
1	18	.56	2.9	13.02	10
2	52	.60	3.7	13.13	10
3	16	.66	6.2	15.02	10
4	5	.76	6.0	14.41	11
5	3	.84	6.6	14.64	11
6+	6	.74	7.9	17.77	12
Total	100	.63	4.4	13.72	10
California ($N = 15{,}711$) mean delay = 2.17 days					
1	30	.54	2.3	8.37	8
2	50	.58	2.7	8.47	8
3	12	.67	3.2	8.92	8
4	3	.72	4.4	9.23	8
5	2	.75	6.0	8.64	8
6+	3	.85	6.7	9.79	8
Total	100	.60	2.9	8.57	8

Competing Risks Model

Table 2 presents parameter estimates of the determinants of length of stay that result in a live discharge or an in-hospital death. The coefficients are presented in terms of hazard ratios (relative risks), so that a hazard ratio of 1.5, for example, implies that a one-unit increase in a variable leads to a 50 percent increase in the conditional probability of discharge on any given day.

The variables that capture differences in patient health status behave as hypothesized. The hazard ratios (and 95 percent confidence intervals)

Table 2: Competing Risks Proportional Hazards Estimates for Postsurgery Length of Stay

	Discharge Status					
Variable	*Alive*			*Dead*		
Hazard Ratios						
Age 65–69	0.725	(0.692,	0.758)	2.930	(1.926,	3.934)
Age 80+	0.607	(0.579,	0.634)	6.721	(4.435,	9.006)
Male	0.923	(0.896,	0.950)	1.943	(1.702,	2.184)
Pertrochanteric fracture	0.850	(0.829,	0.872)	1.286	(1.140,	1.432)
Myocardial infarction	1.033	(0.946,	1.119)	1.595	(1.064,	2.126)
Peripheral vascular disease	0.824	(0.766,	0.881)	1.424	(1.107,	1.741)
Dementia	0.985	(0.944,	1.026)	0.915	(0.750,	1.081)
Chronic pulmonary disease	0.866	(0.834,	0.898)	1.895	(1.616,	2.174)
Rheumatologic disease	0.992	(0.893,	1.091)	1.568	(0.777,	2.360)
Mild liver disease	0.544	(0.450,	0.638)	3.453	(1.672,	5.235)
Diabetes (mild to moderate)	0.920	(0.876,	0.963)	1.243	(1.001,	1.484)
Diabetes with chronic complications	0.755	(0.691,	0.819)	1.421	(0.885,	1.956)
Renal disease	0.524	(0.461,	0.587)	3.210	(2.322,	4.099)
Cancer	0.727	(0.671,	0.783)	1.566	(1.160,	1.971)
Moderate/Severe liver disease	0.549	(0.352,	0.747)	2.697	(0.010,	5.384)
Metastatic solid tumor	0.678	(0.600,	0.757)	3.013	(1.987,	4.039)
Volume	1.002	(1.002,	1.002)	0.999	(0.997,	1.000)
Delay	0.974	(0.970,	0.977)	1.011	(1.006,	1.017)
Massachusetts	0.405	(0.391,	0.418)	0.850	(0.715,	0.985)
Manitoba	0.151	(0.143,	0.158)	0.845	(0.669,	1.021)
Quebec	0.148	(0.142,	0.155)	1.432	(1.207,	1.657)
Constant				0.045	(0.023,	0.068)
Parameter Estimates						
ρ	0.0001	(0.0001,	0.0001)	0.014	(0.011,	0.016)
α	4.044	(3.980,	4.108)	1.452	(1.329,	1.576)
π	1			−2.129	(−2.729,	−1.528)

Note: 95% confidence intervals are in parentheses. Estimated coefficients for the remaining parameters capturing unobserved heterogeneity are listed in the appendix. $N = 37{,}290$. Log-likelihood $= -127{,}384$.

presented under the column labeled Alive show that, relative to patients ages 45–64, patients ages 65–69 who are discharged alive are only .73 times as likely to be discharged on any given day; thus 65–69 year old patients have longer postsurgery lengths of stay than younger patients. The hazard ratios presented in the column labeled Dead indicate that this relationship between age and length of stay is attenuated by the fact that older patients are more likely to die on any given day in the hospital than are younger patients. Presence of comorbidities tends to reduce the conditional probability of a live discharge on any given day and to increase the conditional probability of inpatient death. For example, for a hip fracture patient the presence of chronic pulmonary disease decreases the relative probability of live discharge by a factor of .87 on any given day, and it increases the relative probability of in-hospital death on any given day by 1.90.

After controlling for differences in patient health status, delay still appears to affect postsurgery length of stay for patients discharged alive. For these individuals, the relative probability of discharge is reduced by .97 for each additional day spent waiting for surgery. The relative probability of being discharged dead appears to increase slightly as wait time increases. However, the estimate of the hazard ratio (1.01) is small in magnitude. Thus, longer wait times for surgery contribute to longer postsurgery length of stay but they have little effect on in-hospital death.

Most noticeable are the substantial differences across provinces and states in postsurgery length of stay and inpatient mortality that persist after controlling for covariates. Using California as the base case, the relative probability of live discharge is .41 in Massachusetts, .15 in Manitoba, and .15 in Quebec. Thus, Massachusetts patients who are discharged alive have much longer postsurgery lengths of stay relative to comparable patients in California. Yet the postsurgery length of stay for Canadian patients discharged alive is even longer.

We also find that Quebec patients face a higher relative probability of in-hospital death (1.43) compared to California patients. In contrast, the relative probability of death in the hospital is lower for Massachusetts patients (0.85). The relative probability of in-hospital death is not significantly different for Manitoba versus California patients.

Past researchers hypothesized that practice-makes-perfect effects account for a significant proportion of the variation in outcomes across hospitals. Given that Canadian hospitals are on average substantially smaller than U.S. hospitals,[3] surgery volume was included as a regressor. However, the relative probability of discharge dead did not differ significantly by hospital volume, and the effect of volume on live discharge is positive but very small (1.002) in magnitude.

Finally, we find it important to allow for patient characteristics that are potentially unobservable in the administrative data (e.g., patient frailty) and that may affect both live discharge and death in-hospital. In particular, the estimated coefficient on the unobserved heterogeneity random variable has a value of −2.13 (95% CI: −2.729, −1.528), which implies that the two outcomes are significantly negatively correlated. That is, after controlling for observable covariates, individuals who tend to have a higher probability of live discharge, conditional on length of time in hospital, also have lower hazards of in-hospital death.

Instrumental Variables Analysis

Table 3 presents information on patient characteristics and outcomes by day of the week for each region. Column 2 lists the percentage of all patients admitted on each day of the week. Note that admission of hip fracture patients is fairly even across days of the week: between 13 and 15 percent per day in each region. Column 3 lists the percentage of patients admitted on each day of the week who faced a surgery delay of three or more days. For example, the first row of column 3 indicates that 52 percent of hip fracture patients admitted on a Monday in Manitoba waited three or more days for hip fracture surgery. Note that there is wider variation by day of the week in column 3 versus column 2. The p-values in column 3 indicate that in all regions except for California, average wait times differ significantly by day of the week. In California, wait times are on average substantially shorter. Therefore, in that state we also examined the proportion of patients who waited two or more days for hip surgery and found that the fraction of patients delayed two or more days differed significantly by day of the week ($p = .044$).

We had hypothesized that day of the week of admission would serve as a useful IV in an analysis of surgery delays, because surgical staff may prefer to operate on weekdays rather than on weekends. The figures in column 3 indicate in some instances that delays are less frequent near the end of the normal work week relative to the weekend. However, the evidence is not overwhelming. Yet weak evidence for this hypothesis does not undermine the validity of day of the week of admission as an IV. As long as wait times for surgery are statistically significantly different by day of the week of admission, then the IV can be used to assess the effect of wait times on patient outcomes.

Column 4 indicates that average values of the Charlson index do not vary by day of the week of admission. Thus, although delay varies by day of

Table 3: Relationship Between Admitting Day, Wait Time, Case Mix, and Outcomes for Patients Undergoing Surgery

Admitting Day	*Percent*	*Percent Delayed 3+ Days*	*Charlson Index (Mean)*	*Inpatient Mortality (%)*	*Postsurgery Length of Stay (Mean)*
Manitoba ($N = 3{,}364$)					
Monday	15	52	.52	6.3	33.3
Tuesday	15	49	.45	6.1	29.0
Wednesday	15	52	.50	4.8	32.4
Thursday	15	45	.42	4.7	30.6
Friday	14	43	.50	6.5	29.3
Saturday	13	37	.47	6.8	29.4
Sunday	13	45	.48	5.0	33.9
p-value	–	<.0001	.665	.647	.377
Quebec ($N = 10{,}013$)					
Monday	14	38	.56	7.4	28.4
Tuesday	14	36	.58	8.9	27.7
Wednesday	15	37	.60	9.0	27.6
Thursday	15	27	.52	9.1	28.6
Friday	15	30	.50	7.7	28.9
Saturday	13	34	.51	8.6	27.8
Sunday	13	34	.57	8.6	28.6
p-value	–	<.0001	.091	.561	.912
Massachusetts ($N = 8{,}202$)					
Monday	15	32	.67	2.9	14.57
Tuesday	15	31	.68	5.0	14.01
Wednesday	14	27	.64	4.7	13.93
Thursday	14	25	.60	4.8	13.18
Friday	15	32	.58	4.2	13.52
Saturday	14	34	.58	4.5	13.84
Sunday	13	31	.62	4.5	13.42
p-value	–	<.0001	.123	.188	.489
California ($N = 15{,}711$)					
Monday	15	21	.60	3.2	8.64
Tuesday	15	21	.61	3.0	8.76
Wednesday	14	20	.61	3.2	8.59
Thursday	14	19	.60	3.2	8.62
Friday	15	20	.60	2.8	8.38
Saturday	14	20	.57	2.0	8.43
Sunday	13	21	.58	2.6	8.54
p-value	–	.577	.809	.112	.182

the week of admission, these variations are not attributable to differences in health status. We can therefore use differences in postsurgery outcomes by day of the week as a measure of the effect of delay on outcomes that is not confounded by health status.

Columns 5 and 6 present postsurgery outcomes by day of the week of each region. In each region, the hypothesis tests reveal that neither post-surgery length of stay nor inpatient mortality differs significantly by day of the week of admission. Thus, instrumental variables analysis indicates that wait time for surgery does not lead to detrimental postsurgery outcomes.

If we had instead found that patient outcomes varied significantly by day of the week, the finding could have occurred for two reasons. First, significant variations in outcomes could have been due to tangible effects of delay on postsurgery outcomes. Second, such a finding could have occurred if hospital quality did truly differ by day of the week, for example, lower staffing on weekends. However, we found that in-hospital death and postsurgery length of stay did not vary by day of the week of admission. Therefore, it appears that hospital quality is not directly affected by day of the week of admission, and that day of the week can serve as a valid instrument for delay that does not directly affect patient outcomes.

DISCUSSION

Descriptive statistics drawn from both U.S. and Canadian data imply a causal relation between delay for surgery and postsurgery outcomes for hip fracture patients. However, more careful multivariate analysis that controls for censoring reveals that surgery delay has a relatively small effect on postsurgery length of stay and that it has little bearing on inpatient mortality.

Differences in 30-day mortality between Manitoba and Massachusetts have been identified in previous research that found that surgery delay did not explain this variation (Roos et al. 1996). We expand on this analysis, using competing risks hazards models to control for censoring, so that the determinants of postsurgery length of stay and inpatient mortality can also be analyzed. These variables are of interest because they specifically reflect the consequences of care provided within the hospital, whereas comparisons of care based on 30-day mortality may be confounded by care provided after hospital discharge. In addition, analyses of length of stay are important given the potential cost consequences of longer acute care hospital stays.

Our analysis reveals that postsurgery length of stay is significantly longer in the two Canadian provinces versus the two U.S. states but that it is also longer in Massachusetts than in California. In addition, inpatient mortality is significantly higher in Quebec versus the U.S. states; yet Massachusetts inpatient mortality is slightly lower than that in California. Thus, unexplained differences in outcomes within countries also exist and require further investigation.

The fact that delays and outcomes differ between Canada and the United States may be a function of their different reimbursement systems for healthcare services. Hospitals in the United States are reimbursed a fixed price for each admission based on the DRG system, which encourages prompt discharge. In fact, past research has noted a 42 percent decline in length of stay for patients treated in a large U.S. community hospital after the introduction of DRG reimbursement in the 1980s (Fitzgerald, Moore, and Dittus 1988). In contrast, Canadian hospitals receive a global budget each year, which is not directly related to patient length of stay. Examination of the impact of price incentives on the provision of healthcare warrants further attention.

The competing risks hazards model is readily generalizable to other cases in which a researcher seeks to analyze the determinants of patient outcomes, but it faces the problems of censoring and limited case-mix information found in administrative data. For example, these methods can also be used to compare treatment effects across different regions within the United States, to perform a before-and-after analysis of a particular intervention, or to determine the effect of a continuous variable such as surgical volume on patient outcomes (Hamilton and Hamilton 1997).

We have also demonstrated how we can check the robustness of our conclusions using instrumental variables analysis. Note that in this case, the IV approach provides less detailed information than the competing risks multivariate hazard model. The IV analysis provides a yes/no answer to the question: Does a wait time for surgery of three or more days significantly affect postsurgery outcomes?" In contrast, the multivariate hazards model yields quantitative estimates of the effects of delay and other covariates on patient outcomes. However, the advantage of the IV approach we use is that it is computationally much simpler than the competing risks hazards model. Thus, the IV analysis can be readily applied before more sophisticated models are attempted.

The application of IV analysis requires the presence of an appropriate instrument. Day of the week of admission is not readily generalizable as a valid IV in all examples. However, distance from a "high-tech" hospital has also proved to be a useful IV when the health benefits of more advanced medical technologies have been assessed (McClellan, McNeil, and Newhouse 1997).

Whether one is able to use IV analysis depends on the research question being posed and the data that are available.

The limitations of this study should be noted. We excluded from the analysis patients who were admitted from long-term care facilities. These patients were likely to be more frail than the general hip fracture population and therefore would have more clinical justification for surgery delays not documented in the medical record. Although our proposed methodology aims to account for unobservable differences in patient frailty, it seems reasonable to exclude a subpopulation for whom longer wait times for surgery may be clinically justified.[4]

The measure of delay in this study is the wait time between admission and hip fracture surgery, measured in days. Thus, no explicit distinction was made between medically necessary delays that may be required to stabilize patients with certain comorbidities (Kenzora et al. 1984) and medically unnecessary delays. We chose this approach for two reasons. First, given the associated pain and suffering involved, it would have been difficult to conduct a randomized controlled trial that purposely delayed patients. In addition, a more specific clinical measure of delay would have been difficult to apply because one cannot definitively state ex ante which comorbidities constitute a medically necessary delay. Therefore, this article takes a more general approach and controls for all differences in health status (observable and unobservable) that may be correlated with wait time and postsurgery outcomes.

Our approach for modeling potentially correlated competing risks assumes that a common unobserved factor, patient frailty, potentially influences both live discharges and discharges following death. Although this assumption appears plausible for hip fracture patients, other applications may require a more general model that allows for separate unobserved confounders for each competing risk. The more general model is particularly appropriate if the unobserved confounders are not correlated across risks. Discrete heterogeneity distributions associated with each competing risk would then need to be estimated.

The IV approach works most effectively when a great deal of variability exists in the instrument, which in turn generates a wider range of values for the explanatory variable of interest. In this case wait time by day of the week of admission varied relatively less for California, because all patients tended to be treated quickly. Thus, although the results remain statistically significant for California, validation with a second IV would be more convincing.

Healthcare providers, researchers, and policymakers are becoming increasingly interested in identifying the determinants of high-quality healthcare, but cost and time constraints often limit their analyses to administrative data. The methodologies outlined in this article demonstrate how such data can be analyzed while accounting for their inherent limitations. Further advances in statistical methodology will contribute to an understanding of healthcare outcomes that is both more informative and cost-effective.

APPENDIX

This appendix describes the construction of the likelihood function used to estimate the competing risks model. Denote the duration of a hospital stay by t and suppose that there exist 2 mutually exclusive and exhaustive destinations indexed by $r = a$ (discharged alive from hospital), d (died in hospital). Let $\delta_r = 1$ if the patient is discharged to destination r, and zero otherwise. The transition intensity, $\lambda_r(t)$, is defined as the probability that the patient is discharged to destination r after t days in hospital, conditional on survival in the hospital for at least t days. Suppose the transition intensities for patient i depend on a vector of characteristics, X_i. The probability of observing an exit to r after a hospital stay of length t is then

$$f_r\left(t_i|X_i\right) = \lambda_r\left(t_i|X_i\right) \prod_{j\in a,d} \exp\left(-\int_0^{t_i} \lambda_j\left(u|X_i\right) du\right), r = a, d. \qquad (1)$$

The first term on the right-hand side of Equation 1 is the transition intensity representing the probability that the patient is discharged after t days in the hospital to destination r given that his or her length of stay is greater than or equal to t. The second term, the survivor function, is the probability that the individual survives at least to time t in the hospital and hence does not exit either alive or dead prior to t.

Unobserved characteristics, such as patient frailty, are likely to affect both the live discharge and in-hospital mortality transition intensities. Suppose that the transition intensities depend on a univariate random variable ν in addition to observed characteristics. Let $G(\nu)$ be the distribution function of ν. Following Heckman and Singer, we assume that $G(\nu)$ is a discrete distribution with two points of support. One interpretation of this specification is that there are two types of patients in the population. The location of the points of support and their associated probability mass are estimated jointly with the other parameters of the model. With this specification of $G(\nu)$, the likelihood function may be written as

$$L = \prod_i \sum_{k=1}^{2} \omega_k f_a\left(t_i|X_i, \nu_k\right)^{\delta_{ia}} f_d\left(t_i|X_i, \nu_k\right)^{\delta_{id}}, \qquad (2)$$

where ν_k, $k = 1, 2$ are the points of support with associated probabilities ω_k, which sum to one. The first term of Equation 2 is the probability of observing a stay of t_i days that results in a live discharge, while the second term is the probability of observing a stay of t_i days ending in an in-hospital death. We experimented with more than two points of support, but this did not affect the results.

The final step in the construction of the empirical model involves the specification of the functional form of the transition intensities in Equation 2. We follow a common approach and adopt a proportional hazards specification. In addition, the unmeasured component is allowed to have different factor loadings in each transition intensity function, so that

$$\lambda_r\left(t_i|X_i,\nu\right)=\exp\left(X_i\beta_r+\pi_r\nu\right)\lambda_{0r}\left(t_i\right),\quad r=a,d, \tag{3}$$

where $\lambda_{0r}(t)$ represents the baseline transition intensity function. The factor loadings π_r allow the unobserved factors to influence the death discharge and live discharge transition intensities in different ways. For example, if ν affected live discharge but not mortality, then $\pi_d = 0$. Because we cannot separately identify all of the ν, π_r, and intercept terms in β_r, we normalize $\pi_\alpha = 1$. Alternative normalizations were considered, but these did not alter the results.

A specification of the baseline transition intensity that yields a reasonable fit of the data is the log-logistic distribution:

$$\lambda_{0r}(t)=\frac{\rho_r\alpha_r t^{\alpha_r-1}}{1+\rho_r t^{\alpha_r}},\quad \alpha_r>0,\,\rho_r>0. \tag{4}$$

When $\alpha_r > 1$, $\lambda_{0r}(t)$ has an inverted U shape reaching a maximum at $m=[(\alpha_r-1)/\rho_r]^{1/\alpha_r}$.

The points of support, their associated probabilities, and the coefficients of the log-logistic distribution were all precisely estimated (t-statistics $>$ 3.00).

Table A: Proportional Hazard Unobserved Heterogeneity Parameter Estimates

Variable	*Estimate*	t-*Statistic*
ν_1	−0.080	(−1.420)
ν_2	1.279	(30.632)
ω	0.113	(14.771)

NOTES

1. The AHA defines community hospitals as "all nonfederal, short-term general and other specialty hospitals, excluding hospital units of institutions." The documentation states that the sample for California and Massachusetts was drawn from all general acute care hospitals.
2. The NIS sample is stratified by geographic region, ownership, location, teaching status, and bed size. Thus, the resulting samples do not lead to sample sizes in

California and Massachusetts that are proportionate to their respective populations. Specifically, up to 20 percent of the total *number* of U.S. hospitals within each stratum are randomly selected for inclusion in the NIS. Most of the previous years' hospitals are also reselected for up to three years. Because the northeast tends to have fewer hospitals than the west, the NIS includes a higher proportion of all Massachusetts hospitals than California hospitals. For instance, in 1992 the NIS provides inpatient data for 42 percent of Massachusetts hospitals, but only 26 percent of California hospitals.

3. The average annual volume of hip fracture surgeries per hospital was 77.4 for Manitoba patients, 59.6 in Quebec, 101.9 in Massachusetts, and 96.2 in California.
4. Manitoba patients with a prior length of stay greater than one day in a transferring hospital were excluded for the same reason. There were only 95 such patients, and they were older and had more comorbidities than the general Manitoba sample.

REFERENCES

Baker, S. P., B. O'Neill, and R. S. Karpf. 1984. *The Injury Fact Book.* Lexington, MA: D. C. Heath and Company.

Bredahl, C., B. Nyholm, K. B. Hindsholm, J. S. Mortensen, and A. S. Oleson. 1992. "Mortality After Hip Fracture: Results of Operation Within 12h of Admission." *Injury* 23 (1): 83–86.

Davis, F. M., D. F. Wolner, C. Frampton, A. Wilkinson, A. Grant, R. T. Harrison, M. T. S. Roberts, and R. Thadaka. 1987. "Prospective, Multi-Centre Trial of Mortality Following General or Spinal Anaesthesia for Hip Fracture Surgery in the Elderly." *British Journal of Anaesthesia* 59 (9): 1080–88.

Fitzgerald, J. F., P. S. Moore, and R. S. Dittus. 1988. "The Care of Elderly Patients with Hip Fracture." *The New England Journal of Medicine* 319 (21): 1392–97.

Globerman, S. 1991. "A Policy Analysis of Hospital Waiting Lists." *Journal of Policy Analysis and Management* 10 (Spring): 247–62.

Hamilton, B., and V. Hamilton. 1997. "Estimating Surgical Volume-Outcome Relationships Applying Survival Models: Accounting for Frailty and Hospital Fixed Effects." *Health Economics* 6 (4): 383–95.

Hamilton, B. H., V. H. Hamilton, and N. E. Mayo. 1996. "What Are the Costs of Queuing for Hip Fracture Surgery in Canada?" *Journal of Health Economics* 15 (2): 161–85.

Hannan, E. L., H. Kilburn, Jr., H. Bernard, J. F. O'Donnell, G. Lukacik, and E. P. Shields. 1991. "Coronary Artery Bypass Surgery: The Relationship Between Inhospital Mortality Rate and Surgical Volume After Controlling for Clinical Risk Factors." *Medical Care* 29 (11): 1094–1107.

Heckman, J., and B. Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (2): 271–320.

Hoerer, D., G. Volpin, and H. Stein. 1993. "Results of Early and Delayed Surgical Fixation of Hip Fractures in the Elderly: A Comparative Retrospective Study." *Bulletin: Hospital for Joint Diseases* 53: 29–33.

Iezzoni, L. I., S. M. Foley, J. Daley, J. Hughes, E. S. Fisher, and T. Heeren. 1992. "Comorbidities, Complications, and Coding Bias: Does the Number of Diagnosis Codes Matter in Predicting In-Hospital Mortality?" *Journal of the American Medical Association* 267 (16): 2197–2203.

Kalbfleisch, J. D., and R. L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. New York: John Wiley and Sons.

Katz, S. J., H. F. Mizgala, and H. G. Welch. 1991. "British Columbia Sends Patients to Seattle for Coronary Artery Surgery." *Journal of the American Medical Association* 266 (8): 1108–11.

Kenzora, J. E., R. E. McCarthy, J. D. Lowell, and C. B. Sledge. 1984. "Hip Fracture Mortality." *Clinical Orthopaedics and Related Research* 186 (1): 45–55.

Lancaster, T. 1990. *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.

Luft, H. S., D. W. Garnick, D. H. Mark, and S. J. McPhee. 1990. *Hospital Volume, Physician Volume, and Patient Outcomes: Assessing the Evidence*. Chicago: Health Administration Press.

McClellan, M., B. J. McNeil, and J. P. Newhouse. 1997. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272 (11): 859–66.

Pickles, A., and R. Crouchley. 1995. "A Comparison of Frailty Models for Multivariate Survival Data." *Statistics in Medicine* 14 (13): 1447–61.

Romano, P. S., L. L. Roos, and J. G. Jollis. 1993. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Data: Differing Perspectives." *Journal of Clinical Epidemiology* 46 (10): 1075–79.

Roos, L. L., R. K. Walld, P. S. Romano, and S. Roberecki. 1996. "Short-term Mortality After Repair of Hip Fracture: Do Manitoba Elderly Do Worse?" *Medical Care* 34 (4): 310–26.

Sabiston, D. C. 1991. *Textbook of Surgery*. Philadelphia: W. B. Saunders Company.

Sexson, S. B., and J. T. Lehner. 1988. "Factors Affecting Hip Fracture Mortality." *Journal of Orthopaedic Trauma* 1 (4): 298–305.

U.S. Government Accounting Office. 1991. *Canadian Health Insurance: Lessons for the United States*. Pub. No. GAO/HRD-91-90. Washington, DC: Government Printing Office.