Estimating Surgical Volume–Outcome Relationships Applying Survival Models: Accounting for Frailty and Hospital Fixed Effects

Article *in* Health Economics · July 1997

Impact Factor: 2.23 · DOI: 10.1002/(SICI)1099-1050(199707)6:43.0.CO;2-L · Source: RePEc

2 authors:

Barton Hamilton
Washington University in St. Louis

Vivian Ho
Rice University

HEALTH ECONOMICS, VOL. **6**: 383–395 (1997)

ESTIMATING SURGICAL VOLUME—OUTCOME RELATIONSHIPS APPLYING SURVIVAL MODELS: ACCOUNTING FOR FRAILTY AND HOSPITAL FIXED EFFECTS

BARTON H. HAMILTON[1]* AND VIVIAN H. HAMILTON[1,2]
[1]*John M. Olin School of Business, Washington University in St Louis, MO, USA*
[2]*Centre for the Analysis of Cost-Effective Care, Montreal General Hospital, Canada*

SUMMARY

This paper investigates the surgical volume–outcome relationship for patients undergoing hip fracture surgery in Quebec between 1991 and 1993. Using a duration model with multiple destinations which accounts for observed and unobserved (by the researcher) patient characteristics, our initial estimates show that higher surgical volume is associated with a higher conditional probability of live discharge from the hospital. However, these results reflect differences between hospitals rather than differences within hospitals over time: when we also control for differences between hospitals that are fixed over time, hospitals performing more surgeries in period $t + 1$ than in period t experience no significant change in outcomes, as would be predicted by the 'practice makes perfect' hypothesis. The volume–outcome relationship for hip fracture patients thus appears to reflect quality differences between high and low volume hospitals. © 1997 by John Wiley & Sons, Ltd.

Health Econ. **6**: 383–395 (1997)

No. of Figures: 1. No. of Tables: 6. No. of References: 16.

KEY WORDS—surgical volume–outcome relationships; survival models; frailty; hospital fixed effects

INTRODUCTION

Past studies have found that patients receiving surgery in a hospital performing a large number of surgeries have better outcomes (shorter lengths of stay; lower probabilities of in-hospital mortality) than do those undergoing surgery in a low volume hospital (see Luft *et al.*[1] for a summary of these studies). These results are particularly pertinent in the context of the current health care debate in many countries. For example, many Canadian provinces are being faced with the decision to shut down hospitals and/or reduce the number of hospital beds due to declining health care budgets. To the extent that the positive relationship between volume and outcomes is valid, these governments may actually improve outcomes by regionalizing surgery and closing low volume providers.[2]

While the positive volume–outcome relationship has been strongly established, substantial debate exists in the literature as to the interpretation of this finding. Virtually all of the

*Correspondence to: B. H. Hamilton, John M. Olin School of Business, Washington University in St Louis, Campus Box 1133, One Brookings Drive, St Louis, MO 63130, USA. Tel.: (314) 935-8057. E-mail: hamiltonb@wuolin.wustl.edu.
Contract grant sponsor: Social Sciences and Humanities Research Council of Canada

CCC 1057–9230/97/040383–13 $17.50

Received 14 October 1996
Accepted 21 March 1997

empirical evidence is based on comparisons of outcomes between high and low volume hospitals at a point in time. That is, if hospital A performs more surgeries than hospital B, outcomes will be better for patients admitted to A than for those admitted to B. This relationship may reflect a 'practice makes perfect' effect in which high volume providers are able to gain expertise in performing the procedure, leading to improved outcomes. On the other hand, the relationship between higher volume and better outcomes may simply represent a 'selective referral effect': high quality hospitals which have better outcomes, *ceteris paribus*, are likely to get more referrals from primary care providers. Consequently, when regressing the outcome measure on surgical volume, the estimated coefficient on surgical volume is likely to be biased, since volume is a proxy for hospital quality as well as any practice makes perfect effect. Finally, some have argued that the positive relationship reflects case-mix differences between low and high volume hospitals that are not adequately accounted for in the empirical analysis.[3]

In the light of this uncertainty regarding interpretation, this paper re-examines the relationship between surgical volume and outcomes using longitudinal data on patients undergoing hip fracture surgery at acute care hospitals in Quebec between 1991 and 1993. Hip fractures are a particularly relevant case study, given that these fractures are the leading cause of hospitalization for injuries among the elderly,[4] and account for a disproportionately large number of hospital bed days owing to the relatively long recovery period associated with hip surgery. In addition, examination of the volume–outcome relationship at Canadian hospitals is particularly timely, since the Canadian health care system is currently facing substantial cutbacks and reorganization and most volume–outcome studies have examined hospitals in the USA. As a result, with few exceptions little is known about the relationship in Canada.[5–7]

Our empirical approach differs from that taken in previous studies, because we attempt to distinguish between the various explanations hypothesized in the literature concerning the volume–outcome relationship. First, because the same hospitals are observed over time in the sample, these longitudinal data may be exploited to account for systematic differences in quality between hospitals in a very general way using hospital-specific fixed effects.[8] The period to period fluctuation in the number of surgical procedures performed at each hospital then identifies the effect of volume on outcomes, purged of any difference in quality across hospitals that is fixed over time. Our estimates of the volume–outcome relationship thus rely on the variation in volume and outcomes *within* hospitals over time, while almost all previous studies have relied on variations in volume and outcomes *between* hospitals to estimate the relationship. The results presented in this paper thus relate more closely to one of the key questions faced by policy makers: what would happen to patient outcomes if the number of surgeries were increased (or decreased) at a given hospital?

A second feature of the empirical framework is that the outcome measures of interest, post-surgery length of stay and inpatient mortality, are allowed to be correlated by estimating a competing risk duration model in which the individual may be discharged alive or dead. Previous studies have assumed that these outcomes are independent. The primary drawback of this assumption is that if length of stay and in-hospital mortality are not independent, then one may be more likely to observe an in-hospital death for individuals with long hospital stays. To account for this possibility, the duration model allows for unobserved (to the econometrician) systematic differences in frailty among patients at the time of admission to the hospital using the non-parametric approach described in Heckman and Singer.[9] Finally, we include much more detailed controls for patient health status than have typically been used in the literature to account for case-mix differences across patients.

Using this empirical methodology, we first estimate a specification of the volume–outcome relationship which does not control for fixed differences in hospital quality. The estimation results show that higher volume is associated with an increased conditional (on time in hospital) probability of live discharge, although no significant effect is found on the conditional (on time in hospital) probability of in-hospital mortality. However, when we re-estimate the model accounting for fixed quality differences between hospitals by including hospital-specific dummy variables in the specification, the coefficient on volume shrinks in magnitude and becomes insignificantly different from zero. Consequently, better patient outcomes in larger hospitals do not appear to be based upon a volume/learning effect.

Rather, the results are more consistent with the hypothesis that higher quality hospitals are able to draw more patient referrals.

The next section describes the data and presents descriptive statistics on the volume–outcome relationship, and the subsequent section outlines the statistical framework. We present our results and concluding remarks in the last two sections.

DATA AND PRELIMINARY EVIDENCE

This paper analyses data from the MED-ICHO database, which contains standardized information from hospital discharge abstracts. All acute care hospitals in Quebec report details of each discharge to the provincial Ministry of Health and Social Services. All patients admitted to acute care hospitals with a primary diagnosis of hip fracture (ICD-9 codes 820.0–820.9, Fracture of neck of femur; transcervical, pertrochanteric or other unspecified) who were admitted during or after April 1990 and discharged before March 1993 are included in the sample. Because information on patients whose hospital stays were still in progress at the end of March 1993 was unavailable, we may undercount the volume of surgeries performed in February and March 1993. These patients were thus excluded from the analysis, although the results were virtually identical when they were included. Patients admitted for revision of prior hip fracture surgery were excluded from our sample (1% of original sample). In addition, 6% of patients admitted to the hospital with a hip fracture did not undergo surgery. It appears that many of these patients were admitted to hospitals that did not perform hip fracture surgery and then transferred to another hospital. Since our goal is to examine the impact of surgical volume on post-surgical length of stay, we excluded patients not undergoing surgery from the analysis.

For each patient, data were obtained on date of surgery and date and type of discharge (live or dead) and were used to construct post-surgery length of stay. Information was also obtained on other covariates hypothesized to affect post-surgery mortality and length of stay: age, sex, marital status, type of hip fracture (transcervical, pertrochanteric, other), whether or not the patient was admitted to a teaching hospital, year of admission, median male income in postal code of residence and the number and type of comorbidities at the time of admission. Comorbidities coded as complications were not included since these may be endogenous with respect to length of stay. Information on comorbidities was used to construct a Charlson comorbidity index[10] for each patient using a coding methodology developed specifically for administrative data.[11] This index has been validated as a predictor of mortality in longitudinal studies.

Measuring hospital surgical volume

Studies in the literature have typically measured surgical volume as the number of surgeries (of the particular type) performed in the hospital during the calendar year in which the patient is admitted.[1] If the number of surgeries performed in each hospital frequently fluctuates, this variable may not accurately measure the hospital's amount of cumulative experience at the time of surgery. A more appropriate volume variable designed to capture any practice makes perfect effect is a measure of the number of operations performed by hospital h in the time period *prior* to the current patient's surgery. Consequently, we construct HVOL_{ht} to be equal to the total number of surgeries performed in hospital h in the 12-month period prior to the date (t) of the current patient's surgery. Because we cannot construct the number of surgeries performed at each hospital prior to April 1990, the subsequent estimates are based on the sample of hip fracture patients undergoing surgery at Quebec hospitals between April 1991 and January 1993. Finally, like almost all studies in the literature, we are unable to identify the surgeon for all the patients in the sample and hence cannot estimate the volume–outcome relationship at the surgeon level.

Preliminary tabulations of between and within hospital differences

One of the primary goals of this paper is to distinguish between variations in outcomes associated with differences in volume *between* hospitals and variations in outcomes associated with fluctuations over time in volume *within* hospitals. To provide a first look at the volume–outcome relationship between hospitals, we calculated the average number of hip fracture surgeries performed per 12 month period at each hospital in the sample between April 1991 and March 1993. Mean surgical volume for the 68 hospitals was 53 surgeries per year. Using this average volume measure, we divided hospitals into three groups: low volume hospitals performing fewer than 34 surgeries on average in a 12 month period (25% of hospitals); average volume hospitals performing between 34 and 71 surgeries (50% of hospitals); and high volume hospitals performing more than 71 surgeries (25% of sample). Table 1 gives the average length of stay, fraction of patients dying in-hospital, the average number of comorbidities and Charlson index value by volume category. The first row indicates that patients at low volume hospitals had an average length of stay of 31 days, with 9.4% dying in-hospital. Moving down the rows of the table, increasing surgical volume is associated with lower average lengths of stay and mortality rates. For example, length of stay is almost 10 days shorter at a high volume hospital than at a low volume hospital. Column 3 suggests that at least part of this decline could reflect differences in case mix between hospitals: high volume hospitals appear to have patients with significantly fewer comorbidities. However, column 4 indicates no significant difference in the average Charlson index score across volume quartiles. Finally, the last column of Table 1 investigates the relationship between a commonly cited measure of hospital quality, whether the hospital is university affiliated and volume. Using this measure, high volume hospitals appear to be of higher quality. Consequently, some of the positive relationship found between volume and average hospital outcomes is likely to reflect differences between hospitals in quality and case-mix.

Given the large variations between hospitals in surgical volume and outcomes shown in Table 1, our next task is to describe the relationship between changes over time in volume and outcomes within hospitals. To do so, we construct the percentage difference between the number of surgeries performed at the hospital in the 12 months prior to date t and the average number performed per 12 month period at the hospital over the entire sample. Thus, for each hospital at each date, we are able to determine whether volume is above, equal to or below its long-term average. We then divide this variable into three categories: hospitals whose volume at date t was 10% or more below its sample average (20% of sample); hospitals whose volume at date t was within $\pm 10\%$ of its sample average (60% of sample); and hospitals whose volume was more than 10% above its sample average at date t (20% of sample). Once hospitals are classified into periods of high, average and low volume, we calculate the differences between mean outcomes (and case-mix) for patients admitted to a particular hospital at date t and the average outcome at that hospital over the sample period. Table 2 examines whether hospitals performing an above or below average (for the particular hospital) number of surgeries as of date t also have above or below average outcomes and case mix for that particular period.

The first row of Table 2 shows that in hospitals with surgical volumes at date t more than 10% below their sample mean, average length of stay and in-hospital mortality are both above the hospitals' sample average, but the difference is small and insignificant. The remainder of columns 1 and 2 indicate surprisingly that patients admit-

Table 1. Differences in outcomes and case-mix between hospitals, by volume

		Averages				
Average number of surgeries performed		Length of stay (days) (1)	Fraction died in hospital (2)	Number of comorbidities (3)	Charlson index (4)	Fraction of hospitals university affiliated
(Low)	<34	31.0	0.094	2.77	0.56	0.29
(Average)	34–71	25.9	0.083	2.37	0.55	0.47
(High)	>71	21.8	0.065	1.90	0.54	0.67
Significance test: p-value[a]		0.000	0.001	0.000	0.683	0.000

[a]p-value is from a test of the null hypothesis of equality of means across volume categories.

Table 2. Differences in outcomes and case-mix across periods within hospitals, by differences in volume

Percentage difference in period t volume from hospital sample average		Period t difference from hospital sample average: Length of stay (days) (1)	Fraction died in hospital (2)	Number of comorbidities (3)	Charlson index (4)	Fraction of hospitals university affiliated
(Below)	<–10	0.50	0.006	0.05	–0.02	0.51
(Average)	–10 to 10	–0.39	–0.002	–0.03	0.002	0.56
(Above)	>10	0.63	–0.002	0.03	0.01	0.48

Tests of the null hypothesis that the elements in columns 1–4 equal zero cannot be rejected in any case.

ted during high volume periods have increased hospital durations, although mortality is lower. Nevertheless, these differences are statistically insignificant. Columns 3 and 4 show very small differences in hospital case mix when hospitals perform above or below the average number of surgeries as of date t. The final column shows that hospitals experiencing relatively large fluctuations in volume are less likely to be university affiliated, but again the variation is small. Consequently, comparison of Tables 1 and 2 suggests that the relationship between outcomes and volume primarily reflects differences between hospitals, although it is not clear to what extent this results from case mix and quality differences. From these simple tabulations, it does not appear to be the case that a hospital performing more surgeries in the 12 months prior to period t than its sample average experiences significantly improved outcomes.

METHODOLOGY

While the summary statistics presented in Tables 1 and 2 are suggestive, they do not simultaneously control for the multiple factors which may affect outcomes. This section presents the empirical framework for examining the impact of hospital surgical volume on the duration of hospital stay after hip fracture and the probability of inpatient mortality. This framework is constructed to address two potential pitfalls in the estimation of the relationship between surgical volume and outcomes: (1) the possible correlation between length of stay and inpatient mortality; and (2) fixed differences between hospitals, such as quality, which may also be correlated with volume.

Turning to the first issue, length of stay and discharge destination are estimated jointly using a duration model with multiple destinations. Studies in the literature have estimated separate length of stay and inpatient mortality regressions, hence assuming that these events are independent. If this assumption is false, these regressions potentially yield incorrect inferences regarding the effect of surgical volume on outcomes. For example, suppose that higher volume leads to shorter lengths of stay, but has no effect on in-hospital mortality conditional upon length of stay. A separate regression of mortality on volume may still yield a significant effect, since high volume leads to shorter lengths of stay and in-hospital deaths are less likely to be observed for patients with shorter lengths of stay when the outcomes are positively correlated. Consequently, the empirical framework must allow for the potential non-independence between the unobservables affecting length of stay and mortality.

Denote the duration of a hospital stay by m and suppose that there exist two 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 building block of the analysis is the transition intensity, $\lambda_r(m)$, defined as:

$$\lambda_r(m) = \lim_{\Delta m \to 0^+} \frac{Pr[m < M \leq m + \Delta m, \delta_r = 1 | M \geq m]}{\Delta m} \tag{1}$$

which is the probability that the patient is discharged to destination r after m days in hospital, conditional upon surviving in the hospital for at least m days. Suppose the transition intensities depend upon a vector of individual and hospital characteristics recorded at the date t when patient i is admitted to hospital h, X_{iht}. Note that X_{iht} includes a measure of surgical volume, HVOL_{ht}. The probability of observing an exit to r after a hospital stay of length m is then

$$f_r(m_{iht}|X_{iht}) = \lambda_r(m_{iht}|X_{iht}) \prod_{j\in a,d} \exp\left[-\int_0^{m_{iht}} \lambda_j(u|X_{iht})du\right],\ r = a,d \qquad (2)$$

The first term on the right-hand side of equation (2) is the transition intensity representing the probability that the patient is discharged after m days in hospital to destination r given that his or her length of stay is $\geq m$. The second term, the survivor function, is the probability that the individual survives at least to time m in the hospital and hence did not exit either alive or dead prior to m. The product of the quantities defined in equation (2) across individuals provides the basis for the likelihood function.

Accounting for unobserved patient frailty

Unobserved patient characteristics are likely to impact both the live discharge and in-hospital mortality transition intensities. For example, frailer patients are less likely to be discharged alive and also more likely to die in hospital. The typical approach used in the duration literature to account for unmeasured individual heterogeneity is to suppose that the transition intensities depend on a scalar random variable ν in addition to observed characteristics.[12,13] In our case, the unmeasured characteristics ν could reflect the unobserved health status of the patient at the time of admission to the hospital, which affects outcomes. The estimation approach conditions on ν and integrates it out of the likelihood function.

We are now able to construct the likelihood of observing a post-surgery length of stay of m and a discharge to destination r, conditional upon both measured and unmeasured characteristics. Let $G(\nu)$ be the distribution function of ν. Using equation (2), if the length of stay transition intensities are allowed to depend upon ν, the likelihood function for the model is given by:

$$L = \prod_i \int f_a(m_{iht}|X_{iht}, \nu)^{\delta_{ia}} f_d(m_{iht}|X_{iht}, \nu)^{\delta_{id}}\, \mathrm{d}G(\nu) \qquad (3)$$

The first term of equation (3) is the probability of observing a stay of m_{iht} days that results in a live discharge from hospital h for individual i admitted at date t, while the second term is the probability of observing a stay of m_{iht} days ending in an in-hospital death. The integral in equation (3) reflects the fact that ν is not observed and must be integrated out.

Accounting for fixed differences between hospitals

As noted in the Introduction, much controversy exists in the literature as to whether the positive volume–outcome relationship reflects a practice effect or differences in hospital quality. Some studies have attempted to account for quality differences by including proxies for quality, such as whether the hospital is university affiliated or offers certain facilities, in cross-sectional regressions.[1] However, hospitals may differ in a wide variety of quality dimensions and it is unlikely that a set of three or four variables will fully capture variations in quality between hospitals.

The estimation strategy employed in the paper to address this issue is similar to that of Farley and Ozminkowski[8] and relies on longitudinal hospital data. If hospitals differ in quality and these hospital-specific differences persist over the T periods in the sample, quality differences across hospitals may be accounted for by including a dummy variable for each hospital in the specification. It does not seem unreasonable to assume that hospital quality is relatively constant over a few years. Variables typically used in the literature to measure hospital quality, such as university affiliation and whether the hospital offers particular services, generally remain unchanged during a time period of this length. The coefficients on the hospital dummy variables indicate which hospitals have above or below average outcomes after controlling for observed patient characteristics and the number of surgeries performed at the hospital in the past 12 months. Hence, the hospital fixed effects reflect variations in outcomes *between* hospitals during the entire sample period. The volume coefficient is then identified by the relationship between outcomes and surgical volumes *within* hospitals over time. For example, a positive coefficient estimate on volume in the live discharge transition intensity when hospital specific dummies are included in the model implies that a hospital performing more surgeries in period $t+1$ than in period t would also have improved outcomes (shorter lengths of stay) in period $t+1$, on average. When the hospital indicators are excluded from the model, the coefficient on volume will reflect both differences between high and low volume hospitals and differences within hospitals over time.

Specification of functional forms

The final step in the construction of the empirical model involves the specification of the functional form of the transition intensities in equation (3). 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

$$\lambda_r(m_{iht}|X_{iht},\nu) = \exp(X_{iht}\beta_r + \theta_{hr} + \pi_r\nu)\lambda_{0r}(m_{iht}) \quad r = a,d \tag{4}$$

where θ_{hr} denotes the hospital-specific fixed effect and $\lambda_{0r}(m)$ represents the baseline transition intensity function. Measured and unmeasured characteristics thus shift the transition intensity above or below its baseline. Not all of the factor loadings in equation (4) are identified, so π_a is normalized to 1. We also considered an alternative specification which allowed for separate heterogeneity components, ν_a and ν_d, for the live and dead discharge transition intensities, respectively (all the π_r are set to 1 in this case). However, the results were virtually identical with those presented below using this alternative specification, so we adopted the simpler one factor specification of the frailty distribution.

A variety of parametric and non-parametric methods are available to estimate the baseline transition intensity.[12] Some guidance as to the appropriate functional form may be gained by examining the empirical transition intensities, shown in Figure 1. A parsimonious specification of the baseline transition intensity which allows for the non-monotonic behaviour shown in the figure and which yields a reasonable fit of the data is the log–logistic distribution:

$$\lambda_{0r}(m) = \frac{\varrho_r\alpha_r m^{\alpha_r-1}}{1 + \varrho_r m^{\alpha_r}} \quad \alpha_r > 0, \varrho_r > 0. \tag{5}$$

When $\alpha_r > 1$, $\lambda_{0r}(m)$ has an inverted U shape reaching a maximum at $m = [(\alpha_r-1)/\varrho_r]^{1/\alpha r}$. Specification (5) allows the parameters of the baseline transition intensities to differ for each destination.

Estimation of the model requires that a functional form be chosen for $G(\nu)$. Pickles and Crouchley[13] describe a variety of specifications for $G(\nu)$. We adopt the non-parametric approach suggested by Heckman and Singer[9] and assume that $G(\nu)$ may be approximated by a discrete distribution with a finite number of points of support. 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

$$L = \prod_i \sum_{k=1}^{K} \omega_k f_a(m_{iht}|X_{iht},\theta_h,\nu_k)^{\delta_{ia}} f_d(m_{iht}|X_{iht},\theta_h,\nu_k)^{\delta_{id}} \tag{6}$$

where ν_k, $k = 1, \ldots, K$ are the points of support with associated probabilities ω_k which sum to one. Empirical applications have shown that the value of K required to non-parametrically represent $G(\nu)$ is usually small, generally $K = 3$ or 4.

EMPIRICAL RESULTS

This section presents the estimation results from the empirical model described above. Primary interest focuses on the impact of HVOL_{ht} on outcomes. To make our estimates comparable to those in the literature, we use the natural logarithm of HVOL_{ht} in the specifications, although the results are similar when HVOL_{ht} is used as a regressor. To account for differences across patients, indicators for gender and marital status are included in X_{iht}, as are patient age and income as measured by median male income in 1988 in the postal code of residence. Given the concern in the literature regarding the possible correlation between patient case-mix and surgical volume, we include substantially more variables describing the patient's health status at the time of hospital admission than is typically found in the literature: X_{iht} includes indicator variables for whether the patient has 0, 1, 2, 3 or 4–5 comorbidities (the

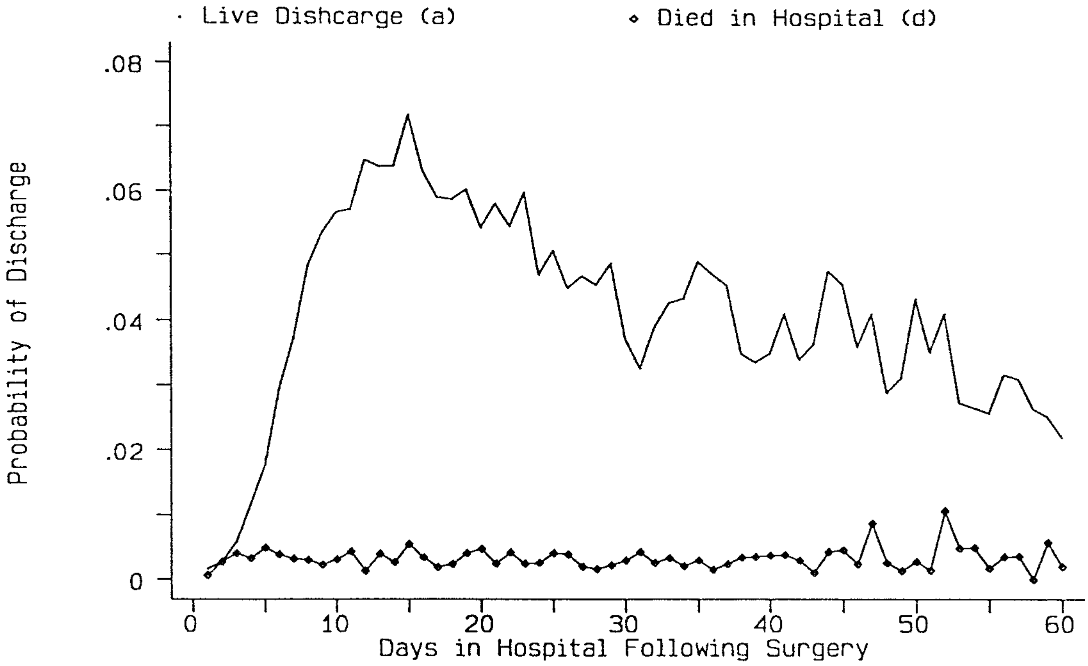

Figure 1. Empirical transition intensities, by destination.

omitted category is 6+ comorbidities), in addition to 10 dummy variables corresponding to the comorbid conditions which make up the Charlson index. Finally, we include indicators for the type of fracture and the year in which the surgery was performed. The year dummies capture any common trend across hospitals in outcomes over time.

Tables 3 and 4 present the parameter estimates of the length of stay transition intensities from the competing risk model. Positive coefficients indicate that an increase in the variable implies an increase in the transition intensity. Regarding the frailty distribution, the data indicated clustering around three points, so that $G(\nu)$ was approximated by a finite distribution with three points of support ($K = 3$).

Table 3 presents the parameter estimates of the determinants of length of stay resulting in a live discharge (columns 1 and 2) and an in-hospital death (column 3 and 4). The odd-numbered columns correspond to a specification which excludes the hospital dummies, while the even-numbered columns allow for fixed differences across hospitals. Both specifications allow for unmeasured (by the econometrician) differences across patients. Turning first to the live discharge transition intensities, the positive and precisely estimated coefficient on $\log(\mathrm{HVOL}_{ht})$ in column 1 implies that after controlling for other potential confounders, such as demographic characteristics and observed comorbidities, patients undergoing surgery at a high volume hospital have a significantly higher probability of leaving the hospital on day m, conditional upon having survived in hospital at least m days. This is consistent with studies in the literature (e.g. Hughes *et al.*[14]) showing that hospitals performing a large number of hip fracture surgeries have shorter lengths of stay in the USA. The remainder of the coefficient estimates indicate that older patients and those in poorer health at the time of admission have lower conditional probabilities of leaving the hospital on any particular day. Finally, the estimates of the baseline hazard parameters ϱ and α imply that the live discharge transition intensity initially increases with length of stay, peaks at approximately 20 days and then declines thereafter.

The estimates in column 2 of Table 3 indicate that the surgical volume-live discharge relationship found in the first column primarily reflects fixed differences between hospitals rather than a within-hospital effect. When hospital fixed effects are included in the specification, the coefficient on

Table 3. Proportional hazard estimates (baseline hazard specification: log–logistic)

	Exit destination			
	Live discharge (a)		Died in hospital (d)	
Variable	(1)	(2)	(3)	(4)
Log(HVOL$_{ht}$)	0.245	–0.085	–0.013	–0.329
	(8.306)	(–0.806)	(–0.124)	(–0.941)
Age	–0.017	–0.019	0.072	0.069
	(–16.281)	(–15.843)	(8.587)	(9.037)
Male	0.013	0.022	0.580	0.547
	(0.403)	(0.625)	(4.547)	(4.434)
Married	0.054	0.041	0.073	0.105
	(1.876)	(1.289)	(0.642)	(0.891)
Income	–0.020	–0.080	0.012	0.056
	(–0.762)	(–2.449)	(0.113)	(0.439)
Pertrochanteric	–0.221	–0.195	0.287	0.301
fracture	(–7.024)	(–5.812)	(2.318)	(2.437)
Other fracture	–0.100	–0.018	0.075	0.041
	(–2.568)	(–0.390)	(0.464)	(0.225)
0 Comorbidities	1.348	1.737	–1.820	–2.129
	(16.615)	(20.892)	(–6.393)	(–6.721)
1 Comorbidity	1.034	1.373	–0.641	–0.891
	(13.493)	(17.704)	(–3.092)	(–4.021)
2 Comorbidities	0.777	1.044	–0.407	–0.634
	(10.422)	(13.758)	(–2.157)	(–3.240)
3 Comorbidities	0.753	0.958	–0.378	–0.606
	(9.963)	(12.562)	(–1.989)	(–3.146)
4–5 Comorbidities	0.324	0.437	–0.132	–0.301
	(4.544)	(6.015)	(–0.845)	(–1.928)
Comorbidity types:				
p-value[a]	0.000	0.000	0.000	0.000
π	1	1	–2.664	–1.343
			(–3.877)	(–3.911)
ϱ	0.0004	0.0003	0.015	0.015
	(8.696)	(8.621)	(3.571)	(4.049)
α	2.699	2.612	1.454	1.513
	(49.910)	(55.556)	(10.989)	(9.901)
Hospital dummies?	No	Yes	No	Yes
p-value[b]	–	0.000	–	0.000

t-Statistics in parentheses. Each regression based on 7383 patient observations at 66 hospitals. Each regression also includes a constant, dummy variables for year in sample and indicators for 10 co-morbidity types. The age and income variables are deviations from their sample means.

[a] *p*-Value is from a test of the null hypothesis that the coefficients on the 10 comorbidity indicators are jointly zero.

[b] *p*-Value is from a test of the null hypothesis that the coefficients on the 65 hospital indicators are jointly zero.

volume in the live discharge transitions declines substantially and is insignificant. Consequently, if the average hospital performs 10 more surgeries in the previous 12 months than its sample 12 month average, the conditional probability of live discharge will not change significantly. On the other hand, the hospital indicator variables, which capture permanent differences between hospitals over the sample period, are jointly strongly statistically significant.

Columns 3 and 4 show that surgical volume does not have a significant effect on the conditional probability of dying in hospital, although the volume coefficient increases in absolute magnitude when the hospital fixed effects are included in the specification. While Hughes *et al.*[14] find a significant hip surgery volume-in-hospital mortality link (not controlling for fixed differences between hospitals), they include only a small set of patient comorbidity indicators; in contrast, our specification incorporates variables measuring the number and type of comorbidities, as well as unobserved patient frailty. The hospital fixed effects may capture some of the between-hospital differences in case-mix. When we adopt a specification similar to Hughes *et al.* which does not allow for unobserved patient differences and hospital fixed effects and only includes comorbidity indicators for diabetes and heart disease, we also find that higher volume is associated with significantly lower in-hospital mortality. Consequently, we suspect that the significant volume–mortality link previously found in the literature reflects patient case-mix differences across hospitals that are correlated with volume. Note that the joint significance of the hospital dummies in column 4 suggests that the conditional probability of in-hospital mortality differs across hospitals, but this does not appear to be related to volume after controlling for observed and unobserved patient heterogeneity.

The final notable finding in Table 3 is the significant and negative estimate of π_d in both columns 3 and 4, implying that unobserved patient characteristics leading to declines in the live discharge transition intensity are associated with higher conditional probabilities of in-hospital mortality. Therefore, treating length of stay and mortality as independent outcomes is overly restrictive.

Table 4 presents the estimates of the unobserved heterogeneity distribution for the specifications which exclude and include the hospital fixed effects. For convenience, denote a realization of ν_k as a 'type k' patient. One interpretation of the K = 3 support points is that there are three types of patients. In the specification which includes hospital dummies, approximately 0.5% of the sample are type 3 individuals who experience significantly longer hospital stays and a substantially higher probability of dying in hospital than do type 1 or type 2 patients.

In summary, surgical volume has a significant and positive effect on the conditional probability of a live discharge from the hospital and an insignificant effect on the conditional probability of in-hospital mortality, thus implying that higher volumes are associated with shorter lengths of stay. However, this relationship appears to primarily reflect differences between hospitals. After including hospital-specific fixed effects, period to period variation in volume within hospitals has no significant impact on the transition intensities. These results appear to be more consistent with explanations for the volume–outcome relationship that emphasize quality differences between hospitals and casts doubt on the practice makes perfect hypothesis. This conclusion is further strengthened if one believes that quality changes substantially over time and is positively correlated with volume as is usually assumed. In this case, the volume coefficients in the fixed effects models would be an upper bound (lower bound in the case of inpatient mortality) on the practice effect, but of course the coefficients are small and insignificant. However, given our short panel, the assumption of relatively constant quality appears reasonable.

Table 4. Proportional hazard model heterogeneity parameter estimates (baseline hazard specification: log–logistic)

	Specification	
Variable	No hospital dummies (Table 3, columns 1 and 3)	Hospital dummies (Table 3, columns 2 and 4)
ν_1	–1.704	0.336
ν_2	–2.360	–0.722
	(4.456)	(2.102)
ν_3	–2.862	–2.753
	(3.114)	(3.754)
ω_1	0.714	0.796
	(4.518)	(4.369)
ω_3	0.034	0.005
	(1.153)	(0.856)
Log-likelihood	–30711.9	–30223.1

t-Statistics in parentheses. *t*-Statistics on ν_2 (ν_3) is from a test of the hypothesis that $\nu_1=\nu_2$ ($\nu_1=\nu_3$).

Decomposing the live discharge outcome

The live discharge outcome encompasses a wide range of possible discharge destinations, including routine discharges to home, discharges to chronic care facilities and discharges to rehabilitation or other hospitals. While we found no effect of volume on the conditional probability of live discharge as a whole after accounting for hospital-specific effects, it may be the case that this result hides a significant effect of volume on a particular subset of live discharges. To address this possibility, we decompose the live discharge outcome into three subcategories: exits to home (h), comprising 45% of total discharges; exits to a chronic care facility (c), comprising 12% of discharges; and exits to a rehabilitation (or other) hospital (b), comprising 35% of discharges. The impact of volume on the transition intensities is then estimated using a likelihood function similar to equation (6), where the set of mutually exclusive and exhaustive discharge destinations is now $r = h, c, b, d$.

The model was re-estimated for the expanded set of exit destinations including the set of patient characteristics X_{iht} used previously, accounting for unobserved patient heterogeneity as above. The first row of Table 5 presents the coefficient estimate on the natural logarithm of surgical volume for each of the four transition intensities when the hospital indicators are excluded from the model. Since our focus is on the volume–outcome relationship, the parameter estimates for the other variables are not presented. The results indicate that the strong positive relationship between surgical volume and live discharge shown in column 1 of Table 3 primarily reflects the fact that increases in volume increase the conditional probability of discharge to a rehabilitation hospital. There is little evidence to suggest that higher volumes lead to speedier discharges to home, or have a significant impact on the conditional probability of discharge to a chronic care facility.

The second row of Table 5 presents the estimates of the volume variable when hospital fixed effects are included in the specification. The results are similar to those presented in Table 3. When the hospital indicators are included in the specification, the effect of surgical volume on the conditional probability of discharge to a rehabilitation hospital becomes small and insignificant, implying that fluctuations in volume within hospitals do not have a substantial effect on this or any of the other transition intensities. Consequently, these results suggest that low and high volume hospitals differ in their propensity to discharge patients to rehabilitation facilities. It may be the case that larger hospitals have developed working relationships with rehabilitation centers, which facilitate placement in these institutions. Overall, the main thrust of our results do not change when live discharges are decomposed into subsets of destinations. The volume–outcome relationship reflects differences between hospitals, rather than within hospitals, for hip fracture patients in Quebec.

Table 5. Proportional hazard estimates of log($HVOL_{ht}$) for expanded set of outcomes, excluding and including hospital indicators

Includes hospital dummies?	Exit destination			
	Home (h)	Chronic care facility (c)	Rehabilitation hospital[a] (b)	Died in hospital (d)
(1) No	0.010 (0.270)	–0.091 (–1.171)	0.750 (16.033)	0.011 (0.123)
(2) Yes	–0.140 (–1.061)	0.010 (0.075)	–0.013 (–0.043)	–0.333 (–1.178)
Fraction discharged to destination	0.45	0.12	0.35	0.08

t-Statistics in parentheses. Each regression based on 7383 patient observations at 66 hospitals. Each regression includes all of the variables shown in Table 3, and accounts for unobserved patient heterogeneity.
[a]Rehabilitation hospital includes other hospital.

CONCLUSION

This paper documents a significant relationship between surgical volume and length of stay among hip fracture patients at Quebec hospitals in the early 1990s. This result is similar to those found for American patients undergoing hip fracture surgery as well as surgery for a variety of other procedures. However, in contrast to these studies, which rely on cross-sectional samples of hospital volume and outcomes, we utilize longitudinal data to decompose the volume–outcome relationship into a 'within' hospital effect determined by period to period changes in a hospital's volume and a 'between' hospital effect reflecting differences among hospitals. In addition, we allow for potential correlation in live and dead discharges by incorporating unobserved (by the researcher) differences across patients at the time of hospital admission. Finally, we account for case mix differences between hospitals by including a substantial number of comorbidity variables. Our findings show that accounting for both measured and unmeasured patient characteristics, period to period fluctuations in a hospital's volume have no significant effect on length of stay or mortality. The significant volume–outcome relationship found in the data reflects differences between hospitals that are fixed over time. This finding persists when we decompose live discharges into exits to home, chronic care facilities or rehabilitation hospitals.

These results cast doubt on the practice makes perfect hypothesis in the case of hip fracture surgery. The results are more consistent with the hypothesis that higher quality hospitals attract more surgical volume, thus yielding the positive volume–outcome relationship observed in cross-sectional data. The results have important implications for health care providers and policy makers who must make decisions regarding resource allocation across hospitals. If declining budgets necessitate hospital closures or bed reductions, then closure of small, low quality hospitals and regionalization of care at large, high quality hospitals is likely to maintain or perhaps even improve overall patient outcomes. On the other hand, if hospital closures are not politically feasible, then reductions in surgical volume may be distributed amongst all hospitals, with no significant detrimental effect on overall patient outcomes. Of course, further information on the relationship between volume and costs is necessary before making such decisions.

Our results suggest that the volume–outcome relationship reflects fixed differences across hospital, such as quality. Further research is necessary to examine more deeply the determinants of between hospital variation. For example, these differences may reflect the quality of surgeons or the surgical team, treatment protocols or the scale of production at the hospital. Understanding of these issues may be important for making certain types of detailed policy recommendations, such as standardizing protocols. We intend to pursue these areas of investigation in future work.

Finally, our findings do not necessarily generalize to other types of surgery. For example, while hip fracture surgery is relatively routine, many government health authorities and oversight boards require that surgeons and hospitals performing operations such as PTCA must meet a minimum per year surgical volume in order to maintain competence and remain certified. For these types of procedures, practice makes perfect effects are clearly believed to be present.[15] However, it is the case that studies of the volume–outcome relationship for these procedures found in the literature rely on cross-sectional data and are unable to distinguish between the practice makes perfect and selective referral hypotheses.[16] Moreover, these studies have not controlled for the correlation between length of stay and inpatient mortality; nor have they accounted for potential differences in patient frailty. The empirical methodology outlined in this paper may be fruitfully applied to analysing the volume–outcome relationship for these other procedures.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada. We are grateful for helpful comments from two referees, Rob Manning and participants in the Fifth European Workshop on Econometrics and Health Economics held at Pompeu Fabra University, Barcelona, Spain.

APPENDIX

Summary statistics and patient characteristics are given in Table A1.

Table A1. Summary statistics, patient characteristics

Variable	Mean (S.D.)
Length of stay (*m*)	24.7 (29.7)
Died in hospital	0.077 (0.267)
Age	76.1 (13.9)
Male	0.27 (0.45)
Married	0.39 (0.49)
Income (median in postal code)	21691 (5260)
Pertrochanteric fracture	0.49 (0.50)
Other fracture	0.18 (0.38)
Number of surgeries (HVOL)	70.1 (31.6)
Admitted to university-affiliated hospital	0.54 (0.50)
0 Comorbidities	0.26 (0.44)
1 Comorbidity	0.21 (0.41)
2 Comorbidities	0.17 (0.38)
3 Comorbidities	0.13 (0.34)
4 or 5 Comorbidities	0.14 (0.35)
Charlson index	0.55 (1.1)
N	7383

REFERENCES

1. Luft, H. S., Garnick, D. W., Mark, D. H. and McPhee, S. J. *Hospital Volume, Physician Volume and Patient Outcomes: Assessing the Evidence.* Ann Arbor, MI: Health Administration Press Perspectives, 1990.
2. Luft, H. S., Bunker, J. and Enthoven, A. Should operations be regionalized? An empirical study of the relation between surgical volume and mortality. *New England Journal of Medicine* 1979; **301**: 1364–1369.
3. Hannan, E. L., Kilburn, H., Jr, Bernard, H., O'Donnell, J. F., Lukacik, G. and Shields, E. P. Coronary artery bypass surgery: the relationship between in hospital mortality rate and surgical volume after controlling for clinical risk factors. *Medical Care* 1991; **29**: 1094–1107.
4. Baker, S. P., O'Neill, B. and Karpf, R. S. *The Injury Fact Book*. Lexington, MA: D. C. Heath, 1984.
5. Roos, L. L., Cageorge, S. M., Roos, N. P. and Danziger, R. Centralization, certification and monitoring: Readmission and complications after surgery. *Medical Care* 1986; **24**: 1044–1066.
6. Roos, L. L., Roos, N. P. and Sharp, S.M. Monitoring adverse outcomes of surgery using administrative data. *Health Care Financing Review* 1987; Annual Supplement: 5–16.
7. Wennberg, J.E., Roos, N. P., Sola, L., Schori, A. and Jaffe, R. Use of claims data systems to evaluate health care outcomes: Mortality and reoperation following prostatectomy. *Journal of the American Medical Association* 1987; **257**: 933–936.
8. Farley, D. E. and Ozminkowski, R. J. Volume–outcome relationships and in hospital mortality: The effect of changes in volume over time. *Medical Care* 1992; **30**: 77–94.
9. Heckman, J. and Singer, B. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 1984; **52**: 271–320.
10. Charlson, M. E., Pompei, P., Ales, K. and MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* 1987; **40**: 373–383.
11. Romano, P. S., Roos, L. L. and Jolles, J. G. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *Journal of Clinical Epidemiology* 1993; **46**: 1075–1079.
12. Lancaster, T. *The Econometric Analysis of Transition Data.* Cambridge: Cambridge University Press, 1990.
13. Pickles, A. and Crouchley, R. A comparison of frailty models for multivariate survival data. *Statistics in Medicine* 1995; **14**: 1447–1461.
14. Hughes, R. G., Garnick, D. W., Luft, H. S., McPhee, S. J. and Hunt, S. S. Hospital volume and patient outcomes: the case of hip fracture patients. *Medical Care* 1988; **26**: 1057–1067.
15. Phillips, K. A., Luft, H. S. and Ritchie, J. L. The association of hospital volumes of percutaneous tranluminal coronary angioplasty with adverse outcomes, length of stay and charges in California. *Medical Care* 1995; **33**: 502–514.
16. Jollis, J. G., Peterson, E. D., DeLong, E. R., *et al.* The relation between the volume of coronary angioplasty procedures at hospitals treating medicare beneficiaries and short-term mortality. *New England Journal of Medicine* 1994; **334**: 1625–1629.

