Volume Title: Health Care Issues in the United States and Japan

Volume Author/Editor: David A. Wise and Naohiro Yashiro, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-262-90292-7

Volume URL: http://www.nber.org/books/wise06-1

Conference Date: May 1-3, 2003

Publication Date: September 2006

Title: The Volume-Outcome Relationship in Japan

Author: Koichi Kawabuchi, Shigeru Sugihara

URL: http://www.nber.org/chapters/c7362

5

The Volume-Outcome Relationship in Japan

The Case of Percutaneous Transluminal Coronary Angioplasty (PTCA) Volume on Mortality of Acute Myocardial Infarction (AMI) Patients

Koichi Kawabuchi and Shigeru Sugihara

5.1 Introduction

Everyone agrees that high quality health care is a very important policy objective, but there are disagreements on how to measure quality. A popular quality measure is the volume of procedures performed by a hospital or physician. The presumption is that as the number of procedures increases, the quality will improve due to, for example, the learning-by-doing (or "practice makes perfect") effect. We refer to an inverse relationship between volume and adverse medical outcomes such as mortality as the *volume effect.* Many studies have examined whether volume affects outcomes, and a consensus seems to have emerged that a volume effect does exist.

Against this background, the American College of Cardiology/American Heart Association (ACC/AHA 2001) recommends that a physician should perform more than 75 percutaneous transluminal coronary angioplasty (PTCA) procedures per year and a hospital should perform at least 200 PTCA procedures per year, and ideally more than 400, to ensure the quality of PTCA procedures.[1] In Japan, the Ministry of Health, Labor and Welfare adopts differential reimbursement policy for PTCA procedures based on hospital PTCA volume, with no adjustment for risk factors.

Koichi Kawabuchi is a professor of health economics at the Tokyo Medical and Dental University. Shigeru Sugihara is a professor at the Osaka School of International Public Policy, Osaka University.

The views expressed in this paper are those of authors and not those of institutions to which the authors belong. We are grateful to David Cutler, Jonathan Skinner, and Chapin White for helpful comments and to Isao Igarashi for skillful editorial work. Remaining errors are ours.

1. In addition, the ACC/AHA recommends that a low-volume physician with fewer than 75 procedures should work only in a high-volume hospital with more than 600 procedures because of direct correlation between both hospital and physician volume and outcomes.

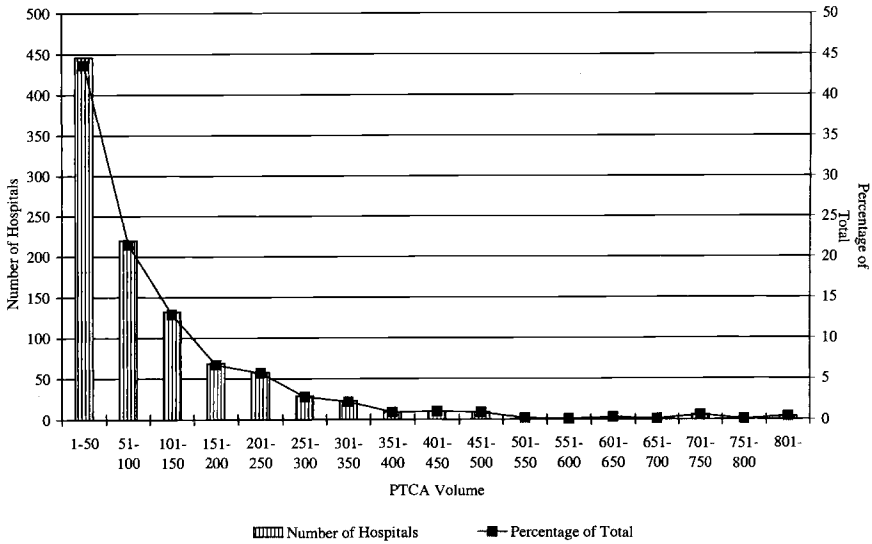

Fig. 5.1 Number of PTCA procedures per hospital

Specifically, if a hospital performs fewer than 100 PTCA procedures per year, its reimbursement rate is reduced by 30 percent.

In Japan, hospitals perform limited numbers of PTCA procedures each year. Figure 5.1 shows the distribution of hospitals according to their PTCA volume in 1997.[2] Nearly half of the hospitals performed fewer than 50 PTCA procedures, and only 15 percent of the hospitals performed more than 200 PTCA procedures per year. Hospitals with more than 400 PTCA procedures are quite rare.[3]

It is sometimes suggested that by increasing PTCA volume per hospital, the quality of health care can be improved. However, empirical studies of the volume effect in Japan are scarce. The reimbursement policy of the Ministry of Health, Labor and Welfare is, therefore, not strongly evidence based. Even in the U.S. context, evidence for the volume effect is, arguably, still mixed. At the least, there is great uncertainty about the nature of the volume effect (how does volume relate to outcomes?), and little is known about the channels through which a volume effect might operate.

This paper examines the empirical relevance of a hypothesized volume effect in Japan in the case of PTCA performed on acute myocardial infarc-

2. This figure is taken from Takeshita (2000).

3. Even in the United States, PTCA volume per hospital or physician is low. Jollis et al. (1997) document that the median annual PTCA procedures on Medicare patients are 98 for hospitals and 13 for physicians. They note that median annual PTCA volumes for all patients including non-Medicare patients are 196 to 294 for hospitals and 26 to 39 for physicians considering that Medicare patients consist of one-third to one-half of total patients.

tion (AMI) patients. We also investigate the nature and channels of the volume effect. The results have implications for reimbursement policy as well as competition policy. If there is a strong volume effect, policies should favor the concentration of PTCA procedures in a small number of hospitals or physicians. If this is not the case, policies favoring concentration of PTCA procedures may be inappropriate.

5.2 Literature Review

A comprehensive review of the volume effect was conducted by Halm, Lee, and Chassin (2000), covering a wide range of diseases and operations. The overall conclusion supports the existence of a volume effect in most diseases and operations including PTCA, coronary artery bypass graft (CABG), and treatment of AMI patients.

However, we suspect that the evidence on the volume effect is still mixed for PTCA. In fact, Halm, Lee, and Chassin (2000) report that, of the seven articles that examined hospital PTCA volume, only three found an inverse relationship between volume and mortality and that, of the four studies that analyzed physician volume, only one found a significant association between volume and mortality and one found a trend toward such a relationship.[4] Empirical studies on the existence of a PTCA volume effect in Japan are not easy to find.[5] Tsuchibashi et al. (2003) find no significant relationship between hospital volume and in-hospital death or CABG.

Even when a volume effect is reported, its statistical significance and empirical relevance need to be scrutinized carefully. For example, one of the most reliable studies identified by Halm, Lee, and Chassin (2000) is Hannan et al. (1997). This study classified hospitals into five categories according to their PTCA volume. They compared risk-adjusted–in-hospital mortality rates across these categories and concluded that patients undergoing PTCA in hospitals with annual PTCA volume less than 600 experienced significantly higher risk-adjusted mortality rates and risk-adjusted–same-stay CABG rates.[6] However, the statistical significance of these differences is not certain because the confidence intervals for the estimated mortality rates are very wide and often overlap. In the same spirit, Ellis et

4. The inverse relationship is found more often for emergency CABG. All seven articles that examined hospital volume and three out of four articles that examined physician volume found that low volume are associated with higher rates of emergency CABG. Very recently, an increasing number of studies are being published that demonstrate a volume effect at the hospital level as well as the physician level.

5. A series of papers by Fujita et al. (2000); Fujita and Hasegawa (1999, 2000); Fujita, Hasegawa, and Hasegawa (2001); and Hasegawa, Hasegawa, and Fujita (2000) find support for a volume effect in the context of operations on cancer patients, cardiovascular operations, and treatment of AMI patients.

6. Hannan et al. (1997) also examined physician volume effects on mortality and emergency CABG and found analogous results.

al. (1996, 1997) purport to find an inverse relationship between PTCA volume and adverse outcomes, but they caution that the magnitude of the relationship is not estimated exactly.

Further, in Hannan et al. (1997), the differences in mortality rates across categories are very small. For example, risk-adjusted mortality rates are 1.12 percent for hospitals with fewer than 400 PTCA procedures, around 0.80 percent for hospitals with 400 to 999 PTCA procedures, and 0.95 percent for hospitals with more than 1,000 PTCA procedures.[7] Arguably, these differences may be statistically significant, but clinically insignificant.

Many of the studies focus on short-term outcomes such as in-hospital mortality and death within thirty days after PTCA. However, Doucet et al. (2002) and Kimmel et al. (2002) show that a relationship between PTCA volume and outcomes may exist for short-term outcomes but not for longer-term outcomes.

And there is great uncertainty about the nature of volume effect. First, the volume effect does not look like a simple linear relationship. In Hannan et al. (1997), mortality and emergency CABG rates are highest among hospitals with low volumes, and they are lower among hospitals with intermediate volumes. However, hospitals with the highest volumes have higher mortality rates than hospitals with intermediate volumes (although they still have lower mortality rates than hospitals with the lowest volume). Also, Ellis et al. (1997), Ho (2000) and Thiemann et al. (1999) found an inverse exponential relationship between PTCA volume and adverse outcomes using the logarithm of PTCA volume as an independent variable in either logistic regression or a Cox proportional hazard model. However, Vakili, Kaplan, and Brown (2001) concludes that there is no significant departure from linearity for the hospital and physician volume-outcome relationship. The question remains whether the volume effect is nonlinear and, if so, why. Policies based on a simple "the more, the better" principle may not be appropriate.

Second, as mentioned in the preceding, the systematic review by Halm, Lee, and Chassin (2000) indicates that hospital volume and physician volume may have different effects on outcomes. In the same vein, Vakili, Kaplan, and Brown (2001) find that physician volume has an effect on in-hospital mortality rate but that hospital volume does not. Complicating the picture further, McGrath et al. (2000) show that hospital volume has a significant effect on mortality but an insignificant effect on emergency CABG, while physician volume does not have a significant effect on mortality but does have a significant effect on emergency CABG.[8] Further, Mc-

7. These mortality rates are for all patients who underwent PTCA, including AMI and non-AMI.

8. See also Jollis et al. (1997).

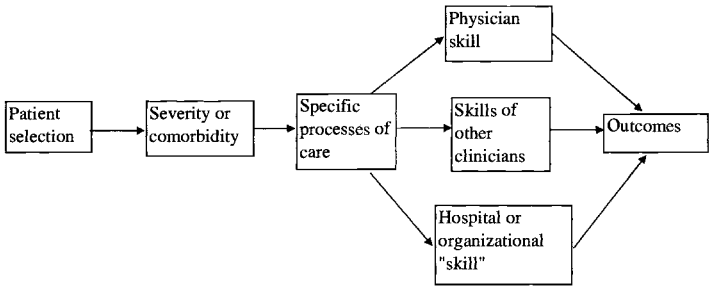

Fig. 5.2 Conceptual framework proposed by Halm, Lee, and Chassin (2000)—How could volume affect quality?

Grath et al. (2000) and Vakili, Kaplan, and Brown (2001) find significant interaction or spillover effects between hospital volume and physician volume. Recently, Birkmeyer et al. (2003) examined the relative importance of hospital and physician volumes in eight cardiovascular procedures or cancer operations and concluded that for many procedures, the observed association between hospital volume and operative mortality is largely mediated by physician volume. The exact relationship between hospital and physician volume deserves further scrutiny.

Fundamentally, it is not obvious why volume should affect outcomes. To date, the literature has presented the "practice makes perfect" hypothesis and the selective referral hypothesis. The former asserts that as physicians perform more PTCA procedures, they become more skillful.[9] This hypothesis implies causality from PTCA volume to outcomes at the level of the individual physician. The latter hypothesis states that physicians or hospitals with better outcomes tend to receive more referrals. According to this hypothesis, the volume effect is not causal and occurs at both the hospital level and the physician level. Admittedly, these hypotheses are rather patchy and do not cover all the aspects of volume effect.

An overall conceptual framework for understanding the volume effect is proposed by Halm, Lee, and Chassin (2000). This framework identifies various channels through which a volume effect could operate, including improved physician skills per se, spillover effects from the skills of other physicians, and organizational skill. Figure 5.2 is a schematic representation of these channels.

We begin with patient selection. The relationship between volume and patient selection may result from selective referral patterns mentioned previously, or patients of high-volume hospitals or physicians may be more appropriate candidates for PTCA than patients of low-volume providers. In

9. Ellis et al. (1997) report that, even though a PTCA volume effect is present, it is not attributable to the physician's years of experience.

relation to this analysis, note that patients suffering AMI usually have little time to select hospitals or physicians.

The severity of patients' illness and the presence of comorbid conditions also affect outcomes. If high-volume hospitals or physicians treat patients who are systematically healthier than their low-volume counterparts, they would tend to have better outcomes. Hence, risk adjustment is essential.

Halm, Lee, and Chassin (2000) note that volume cannot directly produce better outcomes. If volume is related to outcome, that association must be due to differences in the components of care or in the skills with which patients are treated. In the case of AMI patients, components of care might include PTCA, CABG, thrombolytic therapy, aspirin, beta-blockers, angiotensin converting enzyme (ACE) inhibitors, and so on. Provision of different components of care by high-volume hospitals will result in outcomes that differ from low-volume hospitals.

Then, Halm, Lee, and Chassin (2000) distinguish three kinds of skills, namely, physician skill, skills of other clinicians, and hospital or organizational skill. Physician experience (volume) may be a proxy measure of a certain skill level that results in superior performance. Further, the volume or experience of other physicians, the staff team performing PTCA, and the hospital as a whole may be important. We will examine these skills in more detail in the following.

A summary of the current state of research on the volume effect shows that the evidence is still mixed as to whether the volume effect exists; which volume effect is important, hospital or physician; what kind of effects they are; and the direct channel through individual physician's skills, spillover effects from other physicians, or organizational skills of the hospital as a whole.

Investigation into these questions is especially urgent in Japan because no reliable measures of the quality of care are published beyond the number of procedures or operations, a measure that is not only very popular but also adopted by the Japanese government as a criterion for determination of reimbursement. This paper will contribute to our knowledge about the existence, the nature, and the channels of the volume effect and will pave the way to evidence-based policy making in health care.

5.3 Research Strategy

Our research strategy in this paper is as follows. We restrict our analysis to AMI patients who underwent PTCA.[10] This, we believe, minimizes the potential for bias due to patient selection and different combinations of treatments. As explained later, we also adjust for risk of mortality based on

10. Canto et al. (2000) and Thieman et al. (1999) also examine volume effect on AMI patients.

individuals' severity of illness and comorbidities, as measured by International Classification of Diseases (ICD) codes.

Our main focus is the nature of the volume effect and the channels through which it operates. The first question about the nature of the volume effect is whether it operates at the hospital level or physician level. A direct channel for the volume effect would be through individual physicians' skills. However, the volume effect may also operate at the hospital level if there are interactions among physicians or spillover effects from other physicians or overall hospital skills. Therefore, we will analyze the volume effect at the hospital level as well as the physician level.[11]

The second question about the nature of the volume effect is whether the volume effect is linear or nonlinear. A "practice makes perfect" effect does not necessarily imply a linear relationship between volume and outcomes. The marginal effects of volume may be decreasing, that is, may exhibit decreasing returns to scale. And the volume effect could be negative after a certain point due to, for example, congestion effects. Physicians and hospitals have limited capacity due to constraints on time, physical strength, mental acuity (especially concentration), operating space, equipment, staff, and so on. Hence, very high volumes could result in worse outcomes.

As for the channels through which PTCA volume affects outcomes, we will examine externalities or spillover effects among physicians and from team staff or the hospital as a whole. We can imagine the existence of organizational skills or teamwork effects because physicians do not treat patients alone. Physicians may benefit from good team work, suggestions from experienced mentors, and peer pressures from other physicians. Furthermore, high-volume hospitals may have superior equipment and systems that support complex treatments, or they may have the advantage of ample staff and physicians to provide high quality care. In particular, hospitals may adopt continuous quality improvement (CQI), which consists of the repetitive cycle of process and outcomes measurement, design and implementation of interventions to improve the process of care, and remeasurement to determine the effect on quality of care (Ferguson et al. 2003). Halm, Lee, and Chassin (2000) note that the more complex the treatment process, the more likely it is that physician or surgeon skill will be only one of many important components of the full complement of effective care. One advantage of our data set is that it permits the identification of individual physicians, although the identification is not perfect. By using hospital volume in conjunction with physician volume, we can distinguish between the effects of physician volume per se and overall hospital effects and the spillover effect from other physicians.

As we will see later, a physician in a high-volume hospital does not nec-

11. The unit of analysis is patients. Hospital or physician volume is common to patients who underwent PTCA at the same hospitals or by the same physicians.

essarily perform a large number of PTCA procedures. If we were to judge hospital quality by hospital volume, either organizational skills or spillover effects would have to exist so that even a low-volume physician would have better outcomes if he or she works at a high-volume hospital.[12]

An important question is whether volume might simply be a proxy for other effects specific to a hospital or physician. In the case of AMI, Thiemann et al. (1999) report that adjustment for differences in process of care such as use of aspirin, thrombolytic agents, beta-blockers, and ACE inhibitors account for a good part of the survival benefits attributed to high-volume hospitals. Canto et al. (2000) also point out that the lower mortality rates at high-volume hospitals may be due in part to the earlier administration of primary angioplasty after hospitalization.[13] These results suggest that the reported volume effects only represent unobserved effects specific to hospitals or physicians.[14] We are currently investigating this type of factor specific to hospitals or physicians (see Kawabuchi and Sugihara 2003a,b,c).

5.4 Data and Models

The data used in this paper were collected by Kawabuchi in collaboration with the Japan Medical Association. He conducted three waves of surveys of hospitals gathering data on patients' disease diagnoses (ICD-9 or ICD-10), the main operations or procedures patients underwent, and hospital characteristics such as teaching status and the number of beds.

In this paper, we use data from the second and third waves, conducted in 2000 and 2001, respectively.[15] In each wave, thirty-six hospitals participated, of which thirty hospitals took part in both waves. Of the forty-two hospitals in total, sixteen are public (established by the central govern-

12. Another interesting question about the channels through which the volume effect operates is whether experience on AMI patients per se may be important, rather than experience on PTCA. If the volume effect represents the direct effect of physicians' PTCA techniques, then experience on PTCA is essential. However, if it represents overall management of the disease, experience in handling AMI patients may be valuable. Canto et al. (2000) report that even after controlling for the number of patients with myocardial infarction, hospital PTCA volume is inversely related to mortality. Another interesting question is whether the volume of related operations such as CABG is relevant. Percutaneous transluminal coronary angioplasty and CABG are performed by different categories of physicians, but those physicians face the same problems in treating AMI patients and share common knowledge, skills and equipments.

13. Canto et al. (2000) note, however, that there were no important differences in the use of antiplatelet agents, beta-blockers, or heparin among quartiles of hospitals classified by volume.

14. In the case of hip fracture patients in Quebec, Hamilton and Ho (1998) find a volume effect when hospital-specific effects are not included, but when he controls for hospital-specific effects, the volume effect vanishes.

15. The data in the first wave are less reliable due to coding errors, so we exclude it from the analysis.

ment, prefectures, or municipals), and the rest are private. The distribution of hospitals according to the number of beds is as follows: six hospitals have fewer than 200 beds, twelve have 200 to 299, sixteen have 400 to 599, three have 600 to 799, three have 800 to 999, and two have more than 1,000 beds. The total number of patients of all diagnoses is 482,000, of which 3,220 are AMI patients. The number of AMI patients who underwent PTCA is 906.

Diseases and operations or procedures are identified by ICD-9 or ICD-10 codes. Acute myocardial infarction patients are defined by having an ICD-10 code of I21. Percutaneous transluminal coronary angioplasty is identified by ICD-9-CM codes of 3601, 3602, 3605, and 3606, and CABG is identified by ICD-9-CM codes of 3610, 3611, 3612, 3613, 3614, and 3615. In the following we discuss other ICD codes for comorbidities, which we use for risk adjustment.

The main limitations of our data set include the small number of PTCAs per hospital or physician and limited risk adjustment because we obtain information on severity of illness only from ICD-9-CM codes. Hence, as a sensitivity analysis, we report in the appendix supplemental results using the data collected by the Japanese Society of Interventional Cardiology (JSIC). The JSIC data set contains detailed clinical indicators and a large number of PTCA procedures per hospital. Details are described in the appendix.[16]

We focus on the volume effect among AMI patients who underwent PTCA. Focusing on AMI patients has the benefit of reducing referral bias because AMI patients usually have little time or opportunity to select hospitals or physicians, in contrast with elective PTCA.

We use the annual number of PTCA procedures as the volume variable, which is a flow concept.[17] It may be more appropriate to measure volume based on the stock of experience, such as the cumulative number of PTCA procedures performed by a hospital or physician. Due to the limitation of our data set, however, we measure volumes for only up to two years. We leave the question of stock measures of volume for future research. As an outcome we adopt the hazard rate, $h(t)$, which is the instantaneous probability of death at a point in time, conditional on the patient having survived up to that point. This is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{p(t \leq T < t + \Delta t \mid t \leq T)}{\Delta t},$$

where $p(\cdot)$ denotes a conditional probability and T is a random variable that represents the time of the occurrence of the event (death). By inte-

16. See also Chino, Nakanishi, and Isshiki (2000) and Chino et al. (2001).

17. We use the number of PTCA procedures performed on all patients, not just AMI patients because PTCA is essentially the same skill when it is performed on AMI patients as when it is performed on patients with other diseases.

grating this hazard function with respect to time, T, one can infer a patient's probability of death and, hence, the probability of survival.[18]

Conventional practice is to compare in-hospital mortality rates among hospitals with different volumes after adjusting for severity of illness of individual patients. This approach has at least two drawbacks, however. One is that the in-hospital mortality rate contains information only on whether a patient died in the hospital or was discharged alive, neglecting information on whether the patient died soon after PTCA or survived for some time. Our assumption is that a patient who survived for thirty days but then died, for example, was closer to recovery than a patient who died on the first day. This distinction is in line with Doucet et al. (2002) and Kimmel, Sauer, and Brensinger (2002) who show that longer-term outcomes differ from short-term outcomes, implying that the time dimension is important in the evaluation of the quality of health care. The other drawback to the conventional approach is that the first-stage–risk-adjustment regression omits PTCA volume.[19] If PTCA volume truly affects outcomes, the regression without volume variables will result in biased estimates. Therefore, we estimate hazard functions using volume variables as well as other risk factors as independent variables and directly test the hypothesis that the coefficients on the volume variables are significantly different than zero or the associated hazard ratios are significantly different than one.[20]

On the other hand, our estimates may be biased because flow volume is endogenously determined. If, for example, a hospital or physician that is very good at performing PTCA due to reasons other than the volume effect tends to perform a large number of PTCA, as is implied by the selective referral hypothesis, this will result in correlation between the volume variable and the error term in the regression of mortality rates on PTCA volume. To resolve this difficulty, one may explicitly specify the simultaneous determination of volume and quality. Or one may estimate hospital- or physician-specific effects on mortality first and relate such specific effects to volume. We are now investigating this line of research in Kawabuchi and Sugihara (2003a,b,c).

Because the dates of PTCA procedures are unknown in our data set, we cannot specify how long patients survived after they underwent PTCA procedures. Hence, we analyzed survival time after the beginning of hospitalization, not after PTCA. This treatment can be justified on the ground that the timing of performing PTCA is chosen as part of the process of care

18. A survival function, $S(t) \equiv \text{prob}(T > t)$, is related to the hazard function by the following formula: $\int_0^t h(u)du = -\log S(t)$.

19. If a volume variable is included in the regression, the risk-adjusted mortality rate will be independent of volume.

20. This kind of regression approach to the volume effect is adopted by Ellis et al. (1997), Ho (2000), and Thiemann et al. (1999).

so that it is appropriate to measure survival time from the time physicians accept patients and become responsible for them, not from the time they decide to perform PTCA.

Because analysis of the volume effect using the hazard rate is rare,[21] we repeated the analysis using more conventional logistic regression as a sensitivity analysis. We obtained almost identical results, some of which are reported in the context of emergency CABG as an alternative indicator of quality.

We employ three functional forms: linear, log-linear, and quadratic. These cover a wide variety of nonlinearity and have very different policy implications. In the log-linear model, the hazard ratio declines indefinitely as volume increases if the coefficient on the log-linear term is negative. On the other hand, in the quadratic model, the hazard ratio declines up to a certain volume, but increases after that. How much concentration of PTCA procedures is desirable differs between these two models.

In the literature, emergency CABG is often used as an alternative indicator of the quality of health care, as the fact that a patient needs CABG after a failed PTCA clearly represents a bad outcome. It is often found that the volume effect exists for emergency CABG even when no evidence is found for a volume effect on mortality.[22]

Therefore, we conducted multinominal logit analysis using CABG as the dependent variable. However, our data set has two limitations. One is that we cannot distinguish CABG after failed PTCA from other CABG. Hence, we treat CABG in the same hospitalization as emergency CABG. This choice is common in the literature but may be problematic. The second limitation is that only ten patients underwent CABG in the same hospitalization as PTCA. Due to these limitations, we may be unable to obtain reliable estimates. Therefore, in the appendix we conduct a sensitivity analysis using the JSIC data set, which contains more-accurate information on emergency CABG.

We adjust for mortality risk using age, age squared, sex, comorbidities, and the number of occlusions (single- or multivessel disease). We include the following comorbidities (ICD-10 codes in parentheses): diabetes mellitus (E10 to E14), hypometabolism of lipoprotein (E78), hypertension (I10), angina pectoris (I20), chronic ischemic heart disease (I25), heart failure (I47), paroxysmal tachycardia (I48), ventricular fibrillation and flutter

21. As far as we know, Thiemann et al. (1999) is the only study which explicitly presents results of estimating Cox proportional hazard models with hospital volume as an independent variable.

22. For example, Halm, Lee, and Chassin (2000) report that, of seven articles that examined hospital volume, only three found an inverse relationship between volume and mortality, but all seven found that lower volume was associated with higher rates of emergency CABG as well as the combined endpoint of inpatient death or emergency CABG.

(I49), other arrhythmia (I50), shock (R57), and transplant or graft (Z95). The ICD-10 code for multivessel angioplasty (multiple occlusions) is 3605.

We use a parametric hazard model, in which the survival time is distributed as the Weibull distribution.[23] That is, let t_{ij} be failure time of the jth patient treated by the ith hospital. In the proportional hazard model, $\lambda_{ij}(t_{ij}) = \lambda_0(t_{ij})\exp(\mathbf{X}_{ij}\beta)$, where $\mathbf{X}_{ij}$ is a matrix of explanatory variables. We specify the baseline hazard as $\lambda_0(t_{ij}) = rt_{ij}^{r-1}$. Then, the survival time follows the Weibull distribution with two parameters, r and μ_{ij}: $f(t_{ij}, \mathbf{X}_{ij}) = r\mu_{ij}t_{ij}^{r-1}\exp(-\mu_{ij}t_{ij}^r)$, where $\mu_{ij} \equiv \exp(\mathbf{X}_{ij}\beta)$.

It is often pointed out that patients are heterogeneous. Even after controlling for severity of illness, some patients are likely to recover, while others are not. This may be due to unobservable patient characteristics or due to the patient's situation, such as distance to the nearest suitable hospital.

To allow for this kind of patient heterogeneity, we incorporate frailty into our model as an unobservable multiplicative effect, α_{ij}.[24] Let $\lambda_{ij}(t_{ij})$ be the usual hazard function. Incorporating frailty, the hazard function becomes $\lambda_{ij}(t_{ij} \mid \alpha_{ij}) = \alpha_{ij}\lambda_{ij}(t_{ij})$. If the realized value of α_{ij} is greater than (less than) 1, then the jth patient treated at the ith hospital tends to fail at a faster (slower) rate. To achieve identification, it is assumed that α_{ij} is a random variable with mean zero and variance θ and that the frailty density function, $g(\alpha)$, is distributed as Gamma$(1/\theta, \theta)$.[25]

Most patients survive their hospital stay and are discharged alive. For these patients we only know that they survived up to the discharge date, which means that our data are right censored. If we ignore censoring by, for example, including only the patients who died or by regarding time to discharge as time to failure, we are certain to obtain incorrect estimates of the survival probabilities. Let T_j be a possibly censored failure time for the jth patient and C_j be the censoring time. Then, the observed time is $Y_j = \min(T_j, C_j)$. If Y_j is not censored, the contribution of the jth observation to the likelihood function is the density function, $f(Y_j)$, for T evaluated at Y_j. If Y_j is censored, we only know that T_j is greater than Y_j, so that the contribution to the likelihood function is the probability that $T_j > C_j$, that is, prob$(T_j > Y_j)$, which is a survival function, $S(Y_j)$. Hence, the joint likelihood function over all observations, $j = 1, 2, \ldots, n$, is $L = \Pi^n_{j:Y_j\text{uncensored}}$

23. The following exposition of the model and estimation methods is fairly standard. See, for example, Harrell (2001) and Klein and Moeschberger (1997).

24. The frailty here is specified at individual-patient level. This treatment is different than the shared frailty models usually encountered in survival analysis. The modeling here is similar to that of stochastic frontier analysis (see Kumbhakar and Lovell 2000).

25. That is, $g(\alpha) = (\alpha^{1/\theta-1})(e^{\alpha/\theta})/\Gamma(1/\theta)\theta^{1/\theta}$. Then, the survival function, $S_\theta(t)$, of a frailty model is related to the survival function, $S(t)$, of a nonfrailty model as $S_\theta(t) = \{1 - \theta \ln[S(t)]\}^{-1/\theta}$. If $\theta = 1$, the frailty model is just a usual model without frailty (take the log of both sides). Therefore, we can test the relevance of frailty by checking whether $\theta = 1$. Empirically, we cannot reject the hypothesis that $\theta = 1$ in our sample. We use the frailty model, however, in light of the often expressed concern with patient heterogeneity.

$f(Y_j) \times \Pi^{n}_{j:Y_j \text{censored}} S(Y_j)$. Taking the logarithm and maximizing L gives estimates of the parameters of $S(t)$.[26]

If hospitals or physicians have different skills, then patients who are treated by different hospitals or physicians will tend to have different outcomes. On the other hand, patients who are treated by the same hospitals or physicians will tend to have similar outcomes. For example, if a hospital or physician is very skillful, patients who are treated by that hospital or physician will have higher probability of recovery relative to the average. Patients who are treated by an unskilled hospital or physician will have higher probability of death relative to the average.

Therefore, in regression analysis, the residuals among patients who are treated by the same hospital or physician are likely to be correlated. Hence, in calculating our standard errors, patients are assumed to be clustered by hospital or physician.[27] Conventional standard errors assume there is no correlation of random errors among patients.

5.5 Overview of the Data

Of 3,220 AMI patients, 906 patients underwent PTCA, and 51 patients died, resulting in a mortality rate of 5.6 percent. There were twenty-three hospitals that performed PTCA, with an average of 96.6 PTCAs per year.[28] This ranges from a minimum of 1 to a maximum of 312. The number of PTCA procedures per hospital in our sample is generally small, but this is not atypical for Japanese hospitals. As we saw in figure 5.1, a vast majority of Japanese hospitals perform very few PTCA procedures. We should be careful in interpreting the results of the analysis because the small number of procedures could potentially mask volume effects. However, in the appendix, we performed a sensitivity analysis with a data set that contains hospitals with much higher volumes. The results of this sensitivity analysis are consistent with the results in the text.

We identified forty-nine physicians who performed more than 5 PTCA procedures on AMI patients in our sample. In our analysis of physician volume, we exclude patients of physicians who performed fewer than 5

26. Here, we are assuming that censoring is uninformative in the sense that censoring occurs independently of the risk of death. This assumption implies that the contribution of the censoring to the likelihood function simply multiplies L and that the censoring distribution contains little information on the survival distribution. Although this is not necessarily true in our sample, we adopt this assumption for the sake of estimation. Incorporating dependence between the censoring time and the risk of death is left for future research.

27. This relates to the concept of shared frailty in the survival literature (see Hougaard [2000], Kalbfleisch and Prentice [2002] and Klein and Moeschberger [1997]). In this analysis we do not incorporate shared frailty into the hazard functions because of technical limitations. We leave this for future research.

28. Nine hospitals performed PTCA in both the second and third waves. In this calculation, we calculated the number of PTCA procedures per hospital on an annual basis for hospitals with survey periods less than twelve months.

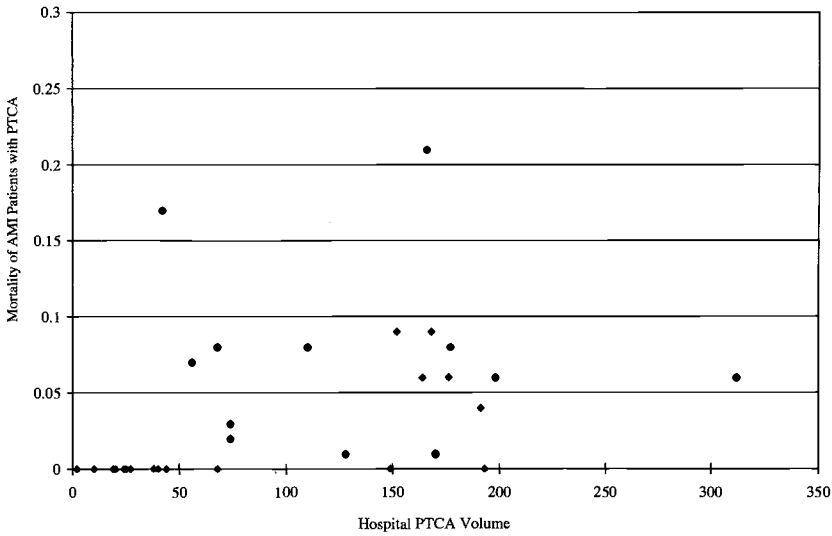

Fig. 5.3 Hospital volume and mortality

PTCA procedures on AMI patients in the data set because these physicians are likely to have extreme outcomes (0 percent or 100 percent mortality, for example) by sheer chance. The average number of PTCAs per physician is 29.5 per year.[29] The maximum number of PTCAs is 144. The average mortality rate is 6.0 percent. The number of observations is 571 when we exclude patients treated by physicians with fewer than 5 PTCA procedures on AMI patients.

Figure 5.3 is a scatter diagram of hospital-level PTCA volume and mortality rates for AMI patients. No clear relationship is apparent. Although this figure shows raw data without risk adjustment, this casts some doubts on the existence of the volume effect at the hospital level.

Figure 5.4 is an analogous scatter diagram for physicians. For physicians with very low volume, mortality rates tend to be high, while for the physicians with higher volume, mortality rates tend to be low. And the relationship seems to be nonlinear, convex to the origin.

Figure 5.5 plots physician PTCA volume against hospital PTCA volume. This figure clearly shows that even within the same hospital, physician volume differs substantially. Physicians in a high-volume hospital do not necessarily have high volume. If there are spillover effects or hospital organizational skills, physicians with low PTCA volume in high-volume

29. Three physicians performed more than five PTCA procedures on AMI patients in both the second and third waves. As with hospitals, we calculated the number of PTCA procedures per physician on an annual basis for physicians with a survey period of less than twelve months.

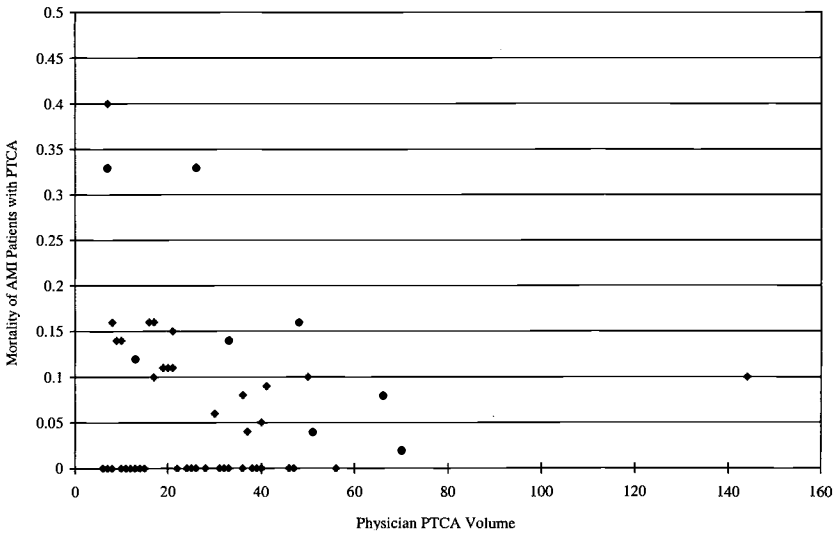

Fig. 5.4 Physician volume and mortality

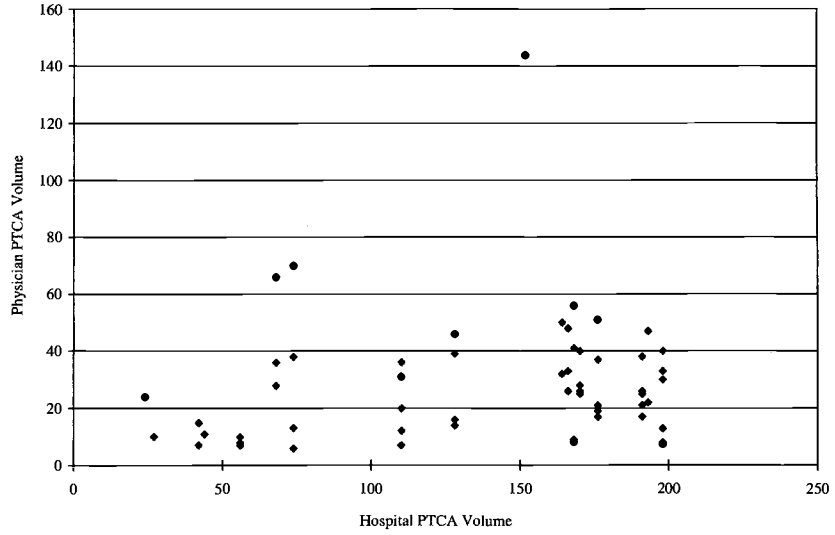

Fig. 5.5 Hospital volume and physician volume

hospitals would tend to have better outcomes. Without such external effects, however, low-volume physicians should have inferior outcomes even if they are at high-volume hospitals. If this is the case, then we should not expect aggregate outcomes at high-volume hospitals to be better than those at low-volume hospitals.

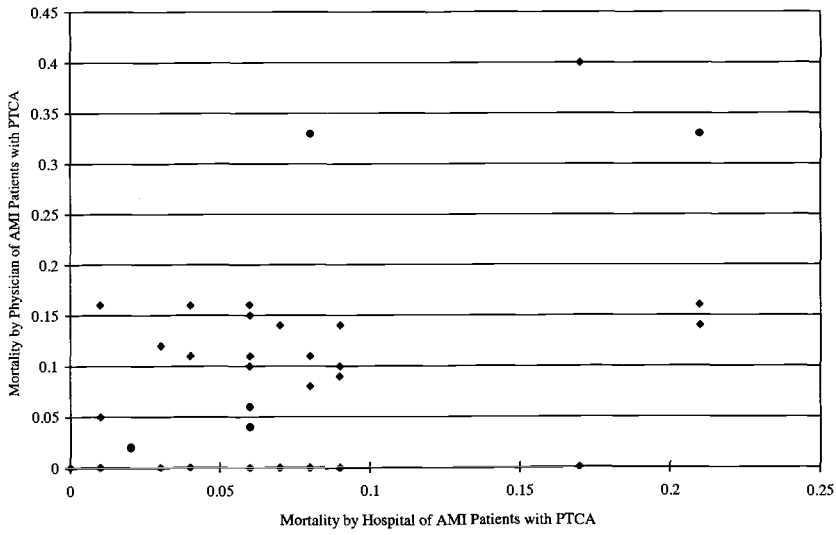

Fig. 5.6 Hospital mortality and physician mortality

Figure 5.6 plots physician-level mortality rates against hospital-level mortality rates. We can see a modest but positive relationship between the two. This might be surprising given the fact mentioned previously that physicians in a high-volume hospital do not necessarily have high PTCA volume. One possibility is that hospital volume has a significant influence on physicians' outcomes. The other possibility is that unobserved common factors affect hospital- and physician-level mortality rates while physician-level mortality is independent of hospital volume.

5.6 Survival Analysis I: Hospital Volume Effect

We estimate hazard functions using three functional forms. One is a linear model in which simply the number of PTCA procedures of each hospital enters as an independent variable. The second is a log-linear model in which the logarithm of PTCA volume enters as an independent variable. The third is a quadratic model in which PTCA volume and squared PTCA volume enter as independent variables.

The results are shown in table 5.1, where we report the hazard ratio for each variable, its standard error, and the p-value for the hypothesis that the hazard ratio is 1. If the hazard ratio of a variable is larger than 1, it means that the variable significantly raises the mortality rate.

We first examine the results of the linear model. Percutaneous transluminal coronary angioplasty volume is not statistically significant at

Table 5.1 **Hospital volume effect**

Independent variables	Linear model			Log-linear model			Quadratic model		
	Hazard ratio	Standard error	*p*-value	Hazard ratio	Standard error	*p*-value	Hazard ratio	Standard error	*p*-value
Volume variables									
PTCA volume	1.001	0.003	0.752	1.050	0.324	0.874	1.004	0.009	0.653
PTCA volume squared							1.000	0.000	0.652
Demographic characteristics									
Age	0.947	0.104	0.620	0.946	0.102	0.610	0.949	0.103	0.628
Age squared	1.001	0.001	0.385	1.001	0.001	0.372	1.001	0.001	0.387
Sex	1.612	0.634	0.225	1.604	0.620	0.222	1.621	0.639	0.220
Risk adjustment									
Diabetes mellitus	0.904	0.264	0.729	0.906	0.263	0.734	0.897	0.260	0.709
Hypometabolism of lipoprotein	0.337	0.259	0.156	0.336	0.257	0.154	0.335	0.258	0.155
Hypertension	0.365	0.111	0.001	0.364	0.112	0.001	0.364	0.111	0.001
Angina pectoris	0.485	0.266	0.188	0.486	0.269	0.192	0.491	0.274	0.202
Chronic ischemic heart disease	1.113	0.448	0.791	1.113	0.441	0.787	1.095	0.426	0.816
Heart failure	1.139	0.441	0.736	1.129	0.437	0.755	1.125	0.444	0.766
Paroxysmal tachycardia	0.448	0.485	0.458	0.451	0.486	0.459	0.443	0.476	0.449
Ventricular fibrillation/flutter	1.330	1.303	0.771	1.333	1.308	0.770	1.337	1.307	0.766
Other arrhythmia	1.296	0.417	0.420	1.301	0.416	0.410	1.298	0.423	0.423
Shock	2.082	1.411	0.279	2.105	1.427	0.272	2.083	1.415	0.280
Transplant/graft	1.116	0.680	0.858	1.108	0.703	0.871	1.128	0.682	0.842
Multiple occlusions	3.955	0.896	0.000	3.966	0.942	0.000	3.813	0.954	0.000
Number of observations		906			906			906	
Log likelihood		–232.999			–233.031			–232.930	

conventional significance levels. Among risk factors, multiple occlusions significantly raises the hazard ratio.

The overall results are similar in the log-linear and quadratic models. In the log-linear model, the logarithm of PTCA volume is not statistically significant, and neither is the squared PTCA volume in the quadratic model.

We find no evidence that hospital volume has a significant influence on the hazard ratio in either the linear or nonlinear case.[30] In other words, we do not find either a learning-by-doing effect or a congestion effect at the hospital level. This result is in sharp contrast with the common belief that the more PTCAs a hospital performs, the better the hospital becomes in the provision of PTCA. This could be due to the fact that physicians in a high-volume hospital are not necessarily high-volume physicians. Even when the volume effect is operative at the physician level, it may not be translated into a volume effect at the level of the hospital as a whole if there are no spillover effects from other physicians or organizational skills. If this is the case, it is not appropriate to judge the quality of care of a hospital based on the hospital's PTCA volume.

5.7 Survival Analysis II: Physician Volume Effect

Next, we examine the effects of physician volume on mortality. From the outset, caution is in order. In our data set, the identification of physicians is uncertain because a physician code does not necessarily identify the physician who performed PTCA, but may instead identify the physician who was responsible for the overall management of the patient.

As with hospital volume, we estimate hazard functions using three functional forms: linear, log-linear, and quadratic. For each functional form, we also examine additional effects of organizational skill and spillover effects from other physicians by including hospital PTCA volume and the volumes of other physicians at the same hospital.

Table 5.2 reports the results when only physician volume is used as an independent variable. We show the coefficients and standard errors for the volume variables in parentheses in addition to the hazard ratio, its standard error, and p-values for the hypothesis that the hazard ratio is 1. In the linear model, physician volume is not statistically significant. Among risk factors, shock and multiple occlusions significantly raise the hazard ratio.

In the log-linear model, the logarithm of physician volume significantly affects the mortality rate. The parameter estimates for the other variables are similar to those from the linear model. We can infer from the estimates how much the hazard ratio decreases as physician volume increases. The

30. To check the sensitivity of the results to small numbers of PTCA per hospital, we repeated the analysis excluding patients who were treated at hospitals with fewer than five PTCA procedures on AMI patients recorded in the data set. The results (not shown) were quite similar.

Table 5.2 **Physician volume effect**

Independent variables	Linear model			Log-linear model			Quadratic model		
	Hazard ratio[a]	Standard error[b]	*p*-value	Hazard ratio[a]	Standard error[b]	*p*-value	Hazard ratio[a]	Standard error[b]	*p*-value
Volume variables									
PTCA volume	0.99369	0.00521	0.227	0.65616	0.12410	0.026	0.96807	0.01529	0.040
	(–0.00633)	(0.00524)		(–0.42135)	(0.18913)		(–0.03245)	(0.01579)	
PTCA volume squared							1.00019	0.00009	0.047
							(0.00019)	(0.00009)	
Demographic characteristics									
Age	1.052	0.186	0.772	1.047	0.172	0.780	1.035	0.168	0.830
Age squared	1.000	0.001	0.963	1.000	0.001	0.983	1.000	0.001	0.967
Sex	3.259	1.277	0.003	3.191	1.187	0.002	3.035	1.191	0.005
Risk adjustment									
Diabetes mellitus	0.846	0.500	0.778	0.865	0.496	0.800	0.827	0.472	0.739
Hypometabolism of lipoprotein	0.220	0.239	0.164	0.229	0.253	0.183	0.227	0.248	0.175
Hypertension	0.467	0.281	0.206	0.470	0.283	0.210	0.455	0.277	0.196
Angina pectoris	0.393	0.278	0.187	0.392	0.265	0.166	0.394	0.262	0.161
Chronic ischemic heart diseases	1.908	0.803	0.125	2.053	0.882	0.094	2.071	0.898	0.093
Heart failure	0.811	0.701	0.809	0.944	0.799	0.945	0.99	0.839	0.989
Paroxysmal tachycardia	0.363	0.440	0.404	0.334	0.406	0.367	0.325	0.395	0.355
Ventricular fibrillation/flutter	2.092	1.961	0.431	2.223	1.990	0.372	2.134	1.850	0.382
Other arrhythmia	1.072	0.419	0.859	1.078	0.398	0.840	1.108	0.410	0.782
Shock	7.508	6.742	0.025	6.351	5.147	0.023	5.985	4.928	0.030
Transplant/graft	2.522	1.976	0.238	2.529	1.902	0.218	2.422	1.832	0.242
Multiple occlusions	6.464	1.838	0.000	7.058	2.200	0.000	5.952	1.602	0.000
Number of observations		571			571			571	
Log likelihood		–155.317			–154.343			–154.175	

[a]Coefficients in parentheses.

[b]Standard errors of coefficients in parentheses.

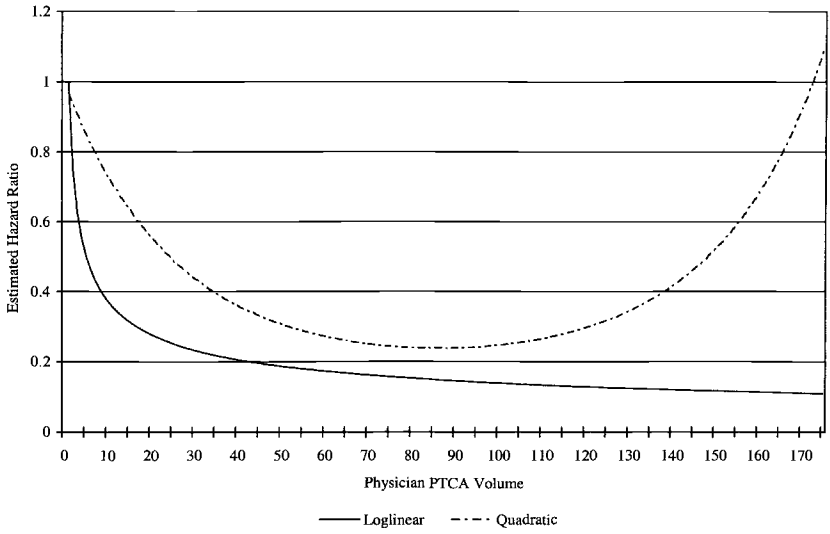

Fig. 5.7 Volume-mortality curve

solid line in figure 5.7 shows this relationship. By assumption, the log-linear model implies that the hazard ratio decreases indefinitely, asymptotically converging to zero. In the very low-volume region, the decrease in the hazard ratio is dramatic. However, the rate of decrease decelerates quickly. Around sixty or seventy PTCA procedures, the decrease becomes quite moderate. Note, however, that the standard error of the estimate on the logarithm of physician volume is so large that the 95 percent confidence interval for the hazard ratio is very wide, ranging from 0.45 to 0.95, with a point estimate of 0.66.

In the quadratic model of physician volume, both the quadratic term and the linear term are statistically significant. The coefficients in parentheses indicate that the lowest hazard ratio is attained at around 85 PTCA procedures. Up to that point, the hazard ratio decreases as physician volume increases, but after that point the hazard ratio increases as physician volume increases. The dotted line in figure 5.7 depicts the relationship between the estimated hazard ratio and PTCA volume in the quadratic model. Again, the initial reduction in hazard ratio is impressive. At the lowest point, the hazard ratio is slightly over 0.2. However, after around 85 PTCA procedures, the hazard ratio increases rapidly. Of course, we should be cautious in extrapolating hazard ratios because we have only one physician with more than 100 PTCA procedures. However, the nonlinearity found in our regression results makes us skeptical of the simple "the more, the better" principle.

We find that performing more PTCA procedures produces learning-by-

doing effects at first but congestion effects later.[31] This result seems reasonable because physicians have limited capacity so that with too many PTCA procedures, they may be too busy to perform well. This also implies that an incentive to increase volume over certain levels may have adverse effects.

Although we find significant physician volume effects, it is important to note that risk factors such as shock and multiple occlusions significantly raise the hazard ratio. In the log-linear and quadratic models, the existence of shock or multiple occlusions raises the hazard ratio by a factor of about 6 to 7. These estimates may appear to be extreme, but it is certainly true that risk factors have large effects on mortality. When we analyze the JSIC data with different and more detailed risk adjustment in the appendix, we also find strong effects from risk factors such as the existence of occlusion in the left main trunk, the number of occlusions, and an AMI severity index. The AMI severity index takes the value 1 if any of the following are observed and zero otherwise: sustained ventricular tachycardia/ventricular fibrillation (VT/VF), shock, heart failure, insertion of a pacemaker, or cardiopulmonary resuscitation. This result reinforces the fact that risk adjustment is essential for the evaluation of the quality of health care. Even with significant volume effects, judging the quality of health care only by volume is inadequate.

We then examine the question of whether organizational skills related to overall hospital volume are important. To do this, we include the number of PTCA procedures performed at each hospital in the regression. In this case, hospital volume represents the effects of hospital volume after controlling for physician-specific effects. If hospital volumes are significant in addition to physician volumes, it provides evidence for effects of organizational skills and spillover effects from other physicians. Three functional forms are again employed, linear, log-linear, and quadratic.

Table 5.3 shows the results. In the linear model, hospital volume is not statistically significant, and physician volume is also insignificant. In the log-linear model, hospital volume is not statistically significant, while physician volume is significant. In the quadratic model, both physician and hospital volumes are strongly significant. However, the effect of hospital volume is difficult to interpret. The estimated coefficients imply that the hazard ratio increases as hospital volume increases up to a little more than ninety PTCA procedures and decreases after that.

Overall, hospital volume does not seem to be an additional contributing factor to the higher quality of health care. The fact that we find a

31. As a sensitivity analysis, we analyzed in-hospital mortality as the dependent variable in a logistic regression treating the observations as a panel data with random effects. We obtained similar results, including the volume at which the lowest mortality is attained. See also the first equation of the multinominal model in table 5.4 below in the context of emergency CABG as an alternative indicator of quality of health care.

Table 5.3 **Physician volume effect with hospital volume**

Independent variables	Linear model			Log-linear model			Quadratic model		
	Hazard ratio[a]	Standard error[b]	*p*-value	Hazard ratio[a]	Standard error[b]	*p*-value	Hazard ratio[a]	Standard error[b]	*p*-value
Volume variables									
Physician PTCA volume	0.99388	0.00491	0.215	0.65216	0.14086	0.048	0.96289	0.01477	0.014
	(–0.00613)	(0.00494)		(–0.42747)	(0.21599)		(–0.03781)	(0.01534)	
Physician PTCA volume squared							1.00020	0.00009	0.033
							(0.00020)	(0.00009)	
Hospital PTCA volume	0.99734	0.00336	0.429	1.02584	0.38686	0.946	1.05265	0.02688	0.044
	(–0.00266)	(0.00337)		(0.02551)	(0.37712)		(0.05131)	(0.02553)	
Hospital PTCA volume squared							0.99979	0.00010	0.035
							(–0.00021)	(0.00010)	
Demographic characteristics									
Age	1.047	0.182	0.793	1.048	0.170	0.773	1.103	0.196	0.580
Age squared	1.000	0.001	0.993	1.000	0.001	0.977	1.000	0.001	0.751
Sex	3.094	1.314	0.008	3.209	1.331	0.005	3.203	1.337	0.005
Risk adjustment									
Diabetes mellitus	0.895	0.531	0.851	0.861	0.498	0.796	0.941	0.562	0.919
Hypometabolism of lipoprotein	0.221	0.238	0.162	0.229	0.254	0.184	0.221	0.249	0.180
Hypertension	0.469	0.278	0.198	0.470	0.286	0.215	0.434	0.264	0.170
Angina pectoris	0.375	0.266	0.167	0.394	0.267	0.169	0.412	0.281	0.194
Chronic ischemic heart disease	2.090	0.859	0.073	2.037	0.858	0.091	2.192	0.846	0.042
Heart failure	0.805	0.694	0.801	0.944	0.799	0.946	0.716	0.623	0.701
Paroxysmal tachycardia	0.374	0.451	0.414	0.333	0.404	0.365	0.362	0.462	0.426
Ventricular fibrillation/flutter	2.421	2.328	0.358	2.199	2.023	0.392	3.256	3.029	0.204
Other arrhythmia	1.085	0.410	0.829	1.077	0.400	0.841	1.248	0.459	0.546
Shock	9.016	8.177	0.015	6.242	5.314	0.031	11.276	10.355	0.008
Transplant/graft	2.323	1.857	0.291	2.545	1.923	0.216	2.154	1.709	0.333
Multiple occlusions	7.018	2.323	0.000	7.021	2.400	0.000	5.296	1.795	0.000
Number of observations		571			571			571	
Log likelihood		–155.070			–154.341			–152.120	

[a]Coefficients in parentheses.

[b]Standard errors of coefficients in parentheses.

significant volume effect only at the physician level, not at the hospital level, can be interpreted in two ways. One interpretation is that the volume effect operates only through a direct channel, namely physicians' PTCA techniques and that spillover effects and effects of organizational skill are weak. Another interpretation is that organizational skills resulting from teamwork, good management, and so on are not related to hospital volume. Even a hospital with a small number of PTCA procedures could provide an excellent medical team and good management and, in particular, could adopt CQI, which stresses that outcomes of health care are direct results of the properties of the system of care (Berwick 1996), although the evidence for the effectiveness of CQI is rather mixed (Shortell, Bennett, and Byck 1998). To disentangle these possibilities will be an important agenda for future research.

5.8 Alternative Outcome Indicator: Emergency CABG

In this section, we analyze another indicator of the quality of health care. There are several candidates, including readmission rates and emergency CABG. Our data set includes information on single hospitalizations, so readmission could not feasibly be measured.

Therefore, we focus on emergency CABG after failed PTCA. However, due to the limitations of our data set, we define *emergency CABG* as all CABGs occurring in the same hospitalization as a PTCA. Though this definition is not ideal, many other studies define emergency CABG in the same way. This analytic strategy is justified on the assumption that if a patient's medical condition requires CABG at the beginning of the hospitalization, the patient would not be subjected to a PTCA. We must be careful in interpreting the results, however, because what we define as *emergency CABG* likely includes some CABGs performed prior to PTCA and some nonemergency CABGs after PTCA.

We estimate multinominal logit models for patients who underwent PTCA, in which the dependent variable can take the following three values: discharged alive without CABG, died in hospital without CABG, and received emergency CABG (whether discharged alive or died in hospital). Conventional studies compare emergency CABG rates among hospitals or physicians independently of whether patients are discharged alive or died in hospital. In that case, there are only two outcomes, received emergency CABG and did not. In reality, however, there are three outcomes after PTCA: discharged alive without CABG, died in hospital without CABG, and received emergency CABG. Therefore, we simultaneously model outcomes of PTCA using multinominal logit models. Of course, our approach is still incomplete in that multinominal logit model assumes independence of irrelevant alternatives (IIA). This assumption is almost certainly violated in this case. But we do not have enough information to conduct multi-

nominal probit analysis, which does not require the IIA assumption. We leave for future research the question of robustness of the IIA assumption.

We analyze the effects of hospital volume as well as physician volume using, as usual, three functional forms: linear, log-linear, and quadratic. Table 5.4 reports the results, in which two equations are being estimated simultaneously. One is the equation on the probability of death, denoted "Died." The other is the equation on the probability of CABG in the same hospitalization, denoted "CABG." Only the parameter estimates for the PTCA volume variable are reported.

The left half of the table shows the hospital volume effects. In both the death equation and the CABG equation, the effect of hospital volume is

Table 5.4 Emergency CABG as an alternative indicator of quality of care

	Hospital volume			Physician volume		
Independent variables	Hazard ratio[a]	Standard error[b]	*p*-value	Hazard ratio[a]	Standard error[b]	*p*-value
Died						
PTCA volume	1.00242	0.00324	0.456	0.99405	0.00619	0.338
	(0.00241)	(0.00324)		(–0.00597)	(0.00623)	
CABG						
PTCA volume	1.00213	0.00600	0.723	1.00434	0.02294	0.850
	(0.00212)	(0.00598)		(0.00433)	(0.02284)	
Number of observations		906			571	
Log likelihood		–205.842			–114.183	
Died						
PTCA volume	1.35706	0.52742	0.432	0.63929	0.14454	0.048
	(0.30532)	(0.38865)		(–0.44739)	(0.22609)	
CABG						
PTCA volume	0.86768	0.69850	0.860	1.48060	1.77685	0.744
	(–0.14193)	(0.80502)		(0.39245)	(1.20009)	
Number of observations		906			571	
Log likelihood		–205.842			–113.135	
Died						
PTCA volume	1.00423	0.01030	0.681	0.96348	0.01682	0.033
	(0.00422)	(0.01026)		(–0.03720)	(0.01746)	
PTCA volume squared	0.99999	0.00003	0.809	1.00023	0.00011	0.026
	(–0.00001)	(0.00003)		(0.00023)	(0.00011)	
CABG						
PTCA volume	1.00728	0.02481	0.769	1.07816	0.11898	0.495
	(0.00725)	(0.02463)		(0.07525)	(0.11035)	
PTCA volume squared	0.99998	0.00008	0.804	0.99946	0.00061	0.376
	(–0.00002)	(0.00008)		(–0.00054)	(0.00061)	
Number of observations		906			571	
Log likelihood		–2205.778			–112.358	

[a]Coefficients in parentheses.

[b]Standard errors of coefficients in parentheses.

not statistically significant in either the linear model (upper panel), the log-linear model (middle panel), or the quadratic model (lower panel).

The right half of the table examines the physician volume effect. In the linear model, physician volume is not statistically significant in either the first equation on the probability of death or in the second equation on the probability of CABG. In the log-linear and quadratic models, physician volume is statistically significant in the death equation. These results are consistent with the earlier survival analysis using a parametric hazard function. In the second equation on the probability of CABG, physician volume is not statistically significant in either the log-linear or quadratic models.

In sum, the probability of death results presented here confirm the earlier results from the survival analysis using parametric hazard function. We do not, however, find evidence for an effect of hospital or physician volume on emergency CABG. This could be due to imprecise identification of emergency CABG, however. When we do the same exercise in appendix using the JSIC data set, which has more accurate identification of emergency CABG, we find a significant effect of hospital volume on emergency CABG in the quadratic model, although the results are not consistent across specifications.

5.9 Conclusion

This paper has examined the empirical relevance of the volume effect in Japan and investigated the nature and channels of volume effect. The main conclusions are as follows:

1. *The volume effect operates not at the hospital level but at the physician level.* This seems plausible because in a hospital different physicians have different volumes, so we should not expect an aggregate volume effect to exist for a hospital as a whole. This finding is robust even when we utilize a data set with detailed risk adjustment and larger PTCA volume per hospital (shown in the appendix). This result implies that policies focused on the hospital as a whole are not appropriate.
2. *The volume effect is nonlinear.* We find significant volume effects for physicians, but the relationship is not linear. The principle of "the more, the better" applies only up to a certain volume. If the quadratic model holds, volumes above a certain level result in worse outcomes, or, as an old sage in China said, "too much is as bad as too little." We suspect that this is due to congestion effects resulting from the fact that physicians themselves have a limited capacity to perform PTCA. The result implies that incentives to increase volume over certain levels may have adverse effects.

However, the log-linear model seems to be equally plausible. We cannot determine which model is more appropriate because very few physicians

performed sufficiently large numbers of PTCA procedures. Note, however, that even if the log-linear model holds, in which mortality rate decreases indefinitely, the benefit of higher volume is exhausted rather quickly. In terms of competition policy, nonlinearity of the volume effect implies that, from a medical point of view, a highly concentrated market is not required.

3. *Risk adjustment is essential for the evaluation of the quality of health care.* Although we find significant physician volume effects, it is important to notice that risk factors such as shock and multiple occlusions greatly increase the hazard ratio. Even with significant volume effects, judging the quality of health care by volume alone is inadequate.

4. *We observe virtually no spillover effects nor organizational skill.* We find no evidence for spillover effects or organizational skills as represented by hospital volumes, which may imply that physicians learn by themselves. This does not necessarily mean, however, that there is no role for peer groups, teamwork, mentors, and so on. Presumably, it simply means that these effects are independent of volume.

5. *More intensive as well as extensive data collection is needed.* Although we believe that the data set we used is one of the best currently available, it has limitations. Possible improvements include a much larger sample size, more-detailed clinical indicators, information on the timing of treatments, more-accurate outcome measures, precise physician identification, proper measures of volume and emergency CABG, and so on.

6. *Future research should focus on the question of whether volume effects simply reflect other factors specific to hospitals or physicians.* These factors include the style of care, adequate staffing and good teamwork, sufficient equipment, the role of the hospital in the health care system (designated emergency hospitals, teaching status, etc.) and internal as well as external governance mechanisms (not-for-profit or incorporated status, the extent to which appointments of physicians are controlled by the professors in the university departments from which physicians are graduated, whether the head of the hospital is dictatorial or not, etc.). We are measuring hospital and physician-specific effects and making a first step in this direction in Kawabuchi and Sugihara (2003a,b,c).

Appendix

Sensitivity Analysis Using the JSIC Data Set

Our sensitivity analysis focuses on three aspects of our methodology. The first is concerned with our strategy of limiting the analysis to include only AMI patients receiving PTCA. The second is the sensitivity of our results to the limited risk adjustment and the small number of PTCA procedures

per hospital. The third is the incomplete identification of emergency CABG in the analysis of emergency CABG as an alternative indicator of the quality of health care services.

As for the first aspect, we could have included in our analysis non-AMI patients who underwent PTCA. In the main text we showed that the number of PTCAs per hospital has no effects on the outcomes of AMI patients. One may wonder, therefore, whether we obtain different results if we analyze the volume effect on all patients, including non-AMI patients.

For the second aspect, our data set contains limited information on risk factors, obtained from ICD codes so that insufficient risk adjustment may be responsible for our results. Furthermore, hospitals in our data set perform very few PTCA procedures per year. In the literature, a volume effect is often found among the highest-volume hospitals, although even in the United States typical hospitals perform only a limited number of PTCA procedures.

As for the third aspect, we regard CABG in the same hospitalization as emergency CABG and find no evidence for an effect of volume on CABG. However, this result may be due to the limitation of our data set, namely that we cannot distinguish between emergency CABG post-PTCA, CABG before PTCA, or nonemergency CABG.

For these sensitivity analyses, we use a data set collected by the JSIC, which includes both AMI and non-AMI patients who underwent PTCA. However, the volume variable only captures the number of PTCA procedures performed at the hospital level. We cannot determine the number of PTCA procedures by each physician.

The main advantage of using this data set is that it contains detailed clinical indicators such as which vessels were occluded, the number of occlusions, and a severity index (which reflects sustained VT/VF, shock, heart failure, etc.). Furthermore, this data set has the merit of identifying emergency CABG explicitly. We have information on whether a patient underwent CABG operation after PTCA during the same hospitalization. In addition, the hospitals in the sample performed large numbers of PTCA procedures per year.

Of thirty-eight hospitals that participated in the survey, thirty-four hospitals performed PTCA on patients with AMI, unstable angina, or other diseases. The average number of PTCAs per hospital is 261, with the maximum of 750 and the minimum of 40. There were 2,011 patients who underwent PTCA procedures in the data set, of whom 640 were AMI patients, and 1,370 were non-AMI patients. Mortality rates in the sample are 6.2 percent for AMI patients and 0.5 percent for non-AMI patients. See Chino, Nakanishi, and Isshiki (2000) and Chino et al. (2001) for more detail.

Risk factors included in the regression are age, age squared, sex, the vessel where occlusions occurred (right coronary artery, left coronary artery, circumflex artery, left main trunk), the number of vessels occluded,

and the AMI severity index. The severity index takes the value 1 if any of the following occurred and zero otherwise: sustained VT/VF, shock, heart failure, insertion of a pacemaker, or cardiopulmonary resuscitation. We experimented with including other risk factors such as diabetes mellitus, comorbidities, and Type C lesion but found these generally to be insignificant so that we drop them in view of the limited number of observations.

We exclude variables specifying the type of procedure, such as intra-aortic balloon pumping (IABP) and primary PTCA. This is because the procedures are endogenously determined in the process of care so that the inclusion of procedural variables could result in biased estimates.[32] If we were to include this type of variable, we would have to take into consideration explicitly the simultaneity of procedures and outcomes.

Volume Effect of PTCA on AMI versus Non-AMI Patients

We first examine the effects of each hospital's number of PTCA procedures on all patients including both AMI and non-AMI patients. We include dummy variables representing indications of AMI and unstable angina. This is because AMI patients are, arguably, more likely to die than patients with other diseases, including unstable angina.

We estimate three functional forms as in the main text: linear, log-linear, and quadratic. The results (not shown) are that in each of the three models, hospital PTCA volume does not significantly affect the hazard ratio, although an indication of AMI, the number of occlusions, and an occlusion in the left main trunk significantly raise the hazard ratio. These results justify our strategy of analyzing only AMI patients, leaving out non-AMI patients.

Detailed Risk Adjustment and Higher Hospital Volume

Henceforth we concentrate on AMI patients. We estimate a parametric hazard function using three functional forms: linear, log-linear, and quadratic. Table 5A.1 shows the results. None of the volume variables in the linear, log-linear, or quadratic models are statistically significant. Among the risk factors, the number of occlusions and AMI severity index are highly statistically significant.

These results are obtained with detailed risk adjustment and higher hospital volumes. Therefore, our results in the main text are robust to risk adjustment methods and hospital volume.

Emergency CABG as an Alternative Indicator of Quality

As in the main text, we estimate a multinominal logit model that describes the probabilities of discharge alive, death after PTCA, and emergency

32. When we include IABP or primary PTCA, we find that IABP is strongly statistically significant, but primary PTCA is not.

Table 5A.1 Hospital volume effect on mortality of AMI patients

Independent variables	Linear model			Log-linear model			Quadratic model		
	Hazard ratio[a]	Standard error[b]	*p*-value	Hazard ratio[a]	Standard error[b]	*p*-value	Hazard ratio[a]	Standard error[b]	*p*-value
Volume variables									
Hospital PTCA volume	0.998	0.001	0.155	0.753	0.207	0.302	1.000	0.005	0.924
	(–0.002)	(0.001)		(–0.284)	(0.275)		(–0.000)	(0.005)	
Hospital PTCA volume squared							1.000	0.000	0.733
							(–0.000)	(0.000)	
Risk adjustment									
Age	0.735	0.137	0.097	0.733	0.142	0.108	0.729	0.144	0.109
Age squared	1.003	0.002	0.083	1.003	0.002	0.097	1.003	0.002	0.095
Sex	0.740	0.510	0.662	0.732	0.504	0.651	0.735	0.505	0.654
Right coronary artery	3.768	5.146	0.331	3.316	4.967	0.423	4.029	5.609	0.317
Left anterior descending coronary artery	6.238	9.526	0.231	5.472	8.961	0.299	6.607	10.132	0.218
Left circumflex coronary artery	1.778	1.554	0.510	1.445	1.269	0.675	1.866	1.619	0.472
Left main trunk	45.841	136.407	0.199	41.678	130.094	0.232	48.878	140.191	0.175
Number of occlusions	2.766	1.056	0.008	2.760	1.104	0.011	2.755	0.978	0.004
AMI severity index	32.745	31.446	0.000	30.652	31.876	0.001	34.765	35.018	0.000
Number of observations		544			544			544	
Log likelihood		–122.731			–122.956			–122.695	

[a]Coefficients in parentheses.

[b]Standard errors of coefficients in parentheses.

CABG. The dependent variable takes the value zero if the patient is discharged alive without emergency CABG, 1 if he or she died without emergency CABG, and 2 if he or she underwent emergency CABG with or without dying. The sample for this analysis consists of AMI patients who underwent PTCA first.

Table 5A.2 reports the results, where only the results for hospital PTCA volume are shown. In the linear model, PTCA volume is not significant in

Table 5A.2 **Emergency CABG as a quality indicator: Multinominal logit models Chino data**

	Hospital volume		
Independent variables	Hazard ratio[a]	Standard error[b]	*p*-value
	A. Linear model		
Died			
PTCA volume	0.998	0.001	0.111
	(–0.002)	(0.001)	
CABG			
PTCA volume	0.996	0.005	0.430
	(–0.004)	(0.005)	
Number of observations		547	
Log likelihood		–94.324	
	B. Log-linear model		
Died			
PTCA volume	0.630	0.134	0.030
	(–0.463)	(0.213)	
CABG			
PTCA volume	0.441	0.257	0.160
	(–0.819)	(0.584)	
Number of observations		547	
Log likelihood		–94.096	
	C. Quadratic model		
Died			
PTCA volume	0.995	0.003	0.121
	(–0.005)	(0.003)	
PTCA volume squared	1.000	0.000	0.378
	(0.000)	(0.000)	
CABG			
PTCA volume	0.987	0.006	0.033
	(–0.013)	(0.006)	
PTCA volume squared	1.000	0.000	0.013
	(0.000)	(0.000)	
Number of observations		547	
Log likelihood		–93.581	

[a]Coefficients in parentheses.

[b]Standard errors of coefficients in parentheses.

either the first equation concerning the probability of death or in the second equation on the probability of emergency CABG.

In the log-linear model, hospital volume is a significant predictor in the death equation, but not in the emergency CABG equation. In the quadratic model, hospital volume is significant in the emergency CABG equation, but not in the death equation.

These results are somewhat inconsistent with the results of the survival analysis estimating parametric hazard function with the same JSIC data set and the results in the main text. In view of the fact that hospital volume is not consistently significant across various specifications, we cannot say definitely that hospital volume has significant effects on mortality or emergency CABG. However, in the case of emergency CABG, considering the problems identifying emergency CABG in our data set, we may give more faith to the results obtained in the JSIC data set than in our data set.

References

American College of Cardiology/Society for Cardiac Angioplasty and Interventions. 2001. Guidelines for Percutaneous Coronary Intervention. *Journal of the American College of Cardiology* 37 (8): 2239i–lxvi.

Berwick, Donald. 1996. A primer on leading the improvement of systems. *British Medical Journal* 312:619–22.

Birkmeyer, John, Therese Stukel, Andrea Siewers, Philip Goodney, David Wennberg, and Lee Lucas. 2003. Surgeon volume and operative mortality in the United States. *New England Journal of Medicine* 349:2117–27.

Canto, John, Nathan Every, David Magid, William Rogers, Judith Malmgren, Paul Frederick, William French, Alan Tiefenbrunn, Vijay Misra, Catarina Kiefe et al. 2000. The volume of primary angioplasty procedures and survival after acute myocardial infarction. *New England Journal of Medicine* 342:1573–80.

Chino, Masao, Narumoto Nakanishi, and Takaaki Isshiki. 2000. The first nationwide database for cost analysis of percutaneous transluminal coronary angioplasty (PTCA) in Japan part 1: Registry and basic characteristics [in Japanese]. *Japanese Journal of Interventional Cardiology* 15:407–12.

Chino, Masao, Narumoto Nakanishi, Takaaki Isshiki, and Hideki Hashimoto. 2001. The JSIC nationwide database for cost analysis of percutaneous transluminal coronary angioplasty (PTCA) in Japan part 2: Multivariate analysis of initial in-hospital charge and the inter-hospital difference [in Japanese]. *Japanese Journal of Interventional Cardiology* 16:401–07.

Doucet, Michael, Mark Eisenberg, Lawrence Joseph, and Louise Pilote. 2002. Effects of hospital volume on long-term outcomes after percutaneous transluminal coronary angioplasty after acute myocardial infarction. *American Heart Journal* 144:144–50.

Ellis, Stephen, Nowamagbe Omoigui, John Bittl, Michael Lincoff, Mark Wolfe, Georgiana Howell, and Eric Topol. 1996. Analysis and comparison of operator-specific outcomes in interventional cardiology. *Circulation* 93:431–39.

Ellis, Stephen, William Weintraub, David Holmes, Richard Shaw, Peter Block, and

Spencer King. 1997. Relation of operator volume and experience to procedural outcome of percutaneous coronary revascularization at hospitals with high intervention volumes. *Circulation* 95:2479–84.

Ferguson, T. Bruce, Eric D. Peterson, Laura P. Coombs, Mary Eiken, Meghan Carey, Frederick Grover, and Elizabeth DeLong. 2003. Use of continuous quality improvement to increase use of process measures in patients undergoing coronary artery bypass graft surgery. *Journal of the American Medical Association* 290:49–56.

Fujita, Takashi, and Toshihiko Hasegawa. 1999. Technology agglomeration of cardiovascular operations and its evaluation [in Japanese]. *Byoin Kanri* 36:76.

———. 2000. Technology agglomeration of treatments of AMI patients and its evaluation [in Japanese]. *Byoin Kanri* 37:106.

Fujita, Takashi, Yuki Hasegawa, and Toshihiko Hasegawa. 2001. An analysis of treatments of AMI patients in the second medical area [in Japanese]. *Nihon Eiseigaku Zasshi* 56:195.

Fujita, Takashi, Hisayoshi Kondo, Yuki Hasegawa, and Toshihiko Hasegawa. 2000. An analysis of operation on cancer patients II [in Japanese]. *Nihon Eiseigaku Zasshi* 55:222.

Halm, Ethan, Clare Lee, and Mark Chassin. 2000. How is volume related to quality in health care? A systematic review of the research literature. In *Interpreting the volume-outcome relationship in the context of health care quality,* Maria Hewitt for the Committee on the Quality of Health Care in America and the National Cancer Policy Board. Washington, DC: Institute of Medicine, National Academy Press. http://books.nap.edu/catalog/10005.html.

Hamilton, Barton, and Vivian Ho. 1998. Does practice makes perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. *Medical Care* 36 (6): 892–903.

Hannan, Edward, Michael Racz, Thomas Ryan, Ben McCallister, Lewis Johnson, Djavad Arani, Alan Guerci, Julio Sosa, and Eric Topol. 1997. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. *Journal of the American Medical Association* 279:892–98.

Harrell, Frank. 2001. *Regression modeling strategies.* New York: Springer-Verlag.

Hasegawa, Yuki, Toshihiko Hasegawa, and Takashi Fujita. 2000. An analysis of operation on cancer patients I [in Japanese]. *Nihon Eiseigaku Zasshi* 55:223.

Ho, Vivian. 2000. Evolution of the volume-outcome relation for hospitals performing coronary angioplasty. *Circulation* 101:1806–11.

Hougaard, Philip. 2000. *Analysis of multivariate survival data.* New York: Springer-Verlag.

Jollis, James, Eric Peterson, Charlotte Nelson, Judith Stafford, Elizabeth DeLong, Lawrence Muhlbaier, and Daniel Mark. 1997. Relationship between physician and hospital coronary angioplasty volume and outcome in elderly patients. *Circulation* 95:2485–91.

Kalbfleisch, John, and Ross Prentice. 2002. *The statistical analysis of failure time data.* 2nd ed. Hoboken, NJ: Wiley.

Kawabuchi, Koichi, and Shigeru Sugihara. 2003a. Hospital- and physician-specific effects on outcomes of PTCA for AMI patients. Mimeograph.

———. 2003b. Hospital-specific effects on after-discharge outcomes of PTCA for AMI patients. Mimeograph.

———. 2003c. The quality and costs of PTCA procedures on AMI patients in Japan. Mimeograph.

Kimmel, Stephen, William Sauer, Colleen Brensinger, John Hirshfeld, Howard Haber, and Russell Localio. 2002. Relationship between coronary angioplasty

laboratory volume and outcomes after hospital discharge. *American Heart Journal* 143:833–40.

Klein, John, and Melvin Moeschberger. 1997. *Survival analysis.* New York: Springer-Verlag.

Kumbhakar, Subal, and Knox Lovell. 2000. *Stochastic frontier analysis.* Cambridge: Cambridge University Press.

McGrath, Paul, David Wennberg, John Dickens, Adrea Siewers, Lee Lucas, David Malenka, Mirle Kellett, and Thomas Ryan. 2000. Relation between operator and hospital volume and outcomes following percutaneous coronary intervention in the era of the coronary stent. *Journal of the American Medical Association* 284:3139–44.

Shortell, Stephen M., Charles L. Bennett, and Gayle R. Byck. 1998. Assessing the impact of continuous quality improvement on clinical practice: What it will take to accelerate progress? *Milbank Quarterly* 76:593–624.

Takeshita, Akira. 2000. *Survey of hospitals on cardiovascular intervention in Japan* [in Japanese]. Tokyo: Ministry of Health, Labor and Welfare.

Thiemann, David, Josef Coresh, William Oetgen, and Neil Powe. 1999. The association between hospital volume and survival after acute myocardial infarction in elderly patients. *New England Journal of Medicine* 340:1640–48.

Tsuchibashi, M., Hiroyuki Tsutsui, Hideo Tada, Miwako Shihara, Suminori Kono, and Akira Takeshita. 2003. The volume-outcome relation for hospitals performing angioplasty for acute myocardial infarction: Results from nationwide Japanese registry [in Japanese]. Mimeograph.

Vakili, Babak, Robert Kaplan, and David Brown. 2001. Volume-outcome relation for physicians and hospitals performing angioplasty for acute myocardial infarction in New York State. *Circulation* 104:2171–76.

This Page Intentionally Left Blank