

# The Effect of Clustering of Outcomes on the Association of Procedure Volume and Surgical Outcomes

Katherine S. Panageas, DrPH; Deborah Schrag, MD, MPH; Elyn Riedel, MA; Peter B. Bach, MD, MAPP; and Colin B. Begg, PhD

**Background:** A large body of literature documents associations between the volume of cases a hospital or surgeon treats and clinical outcomes. Most of these studies have used conventional statistical methods that do not recognize the fact that hospitals or surgeons with similar volumes may have very different outcomes because of systematic differences in processes of care, a phenomenon that exaggerates the true statistical significance of the effect of volume on outcome.

**Objective:** To describe methods to assess the degree of this "clustering" of outcomes and to explore the impact of available statistical techniques that correct for clustering.

**Design:** Reanalysis of 3 previously published volume–outcome studies.

**Setting:** Medicare beneficiaries 65 years of age or older undergoing surgery for colon, prostate, or rectal cancer in the population defined by the Surveillance, Epidemiology, and End Results cancer registries during 1992 to 1996.

**Patients:** 3 data sets were analyzed to assess the impact of surgeon volume on outcomes: 1) 24 166 colectomies performed by 2682 surgeons, 2) 10 737 prostatectomies performed by 999

surgeons, and 3) 2603 rectal resections performed by 1141 surgeons.

**Measurements:** Volume–outcome trends were analyzed by a conventional method (logistic regression) and corrected for clustering. Two widely used statistical methods for analyzing clustered data, a random-effects model and generalized estimating equations, were used and compared, and the degree of clustering was presented graphically.

**Results:** Substantial clustering was observed in the analyses involving morbidity end points. The 2 statistical techniques produced noticeably different results in some analyses.

**Conclusions:** The presence of clustering represents variations in outcomes among providers with similar volumes. Thus, in volume–outcome studies, the degree of clustering of outcomes should be characterized because it may provide insight into variations in quality of care.

*Ann Intern Med.* 2003;139:658-665.

www.annals.org

For author affiliations, see end of text.

See related article on pp 649-657 and editorial comment on pp 696-698.

In an attempt to evaluate the degree to which the choice of provider affects outcomes of major medical procedures, numerous investigators have used procedure volume as a proxy for expertise and have conducted studies correlating volume with outcomes. However, volume is a crude, easily calculated measure, and its use may overlook large variations in quality among providers that are independent of the number of procedures performed. If there are such large variations in outcomes among providers, the outcomes of patients treated by the same provider are necessarily correlated, or "clustered." Clustering of this nature invalidates conventional statistical analyses. Patients treated at the same hospital or by the same surgeon may be more likely to experience similar outcomes if surgical technique or supportive care practices vary among providers and if these factors affect outcomes. It is well established that statistical methods must correct for the effect of clustering of this nature if it exists (1–3). In general, correction for clustering attenuates the statistical significance of observed trends.

In this article, we describe measures that characterize the degree of clustering of outcomes and methods that permit the calculation of tests of statistical significance that are adjusted for the impact of clustering. We reanalyzed 3 large studies of major cancer surgery (4–7) and found examples that illustrate varying degrees of clustering. This reanalysis focused on the relationship of surgeon volume to outcome. We explored the effect of correction for cluster-

ing on the statistical significance of tests for association between surgeon volume and outcome. Graphical methods that visually characterize the degree of between-surgeon variation in quality after adjustment for volume and case mix are described. We also explored analyses that attempt to distinguish between the effects of surgeon volume and those of hospital volume.

## METHODS

### Data Sources

All 3 data sets were obtained by using the linkage of the Surveillance, Epidemiology, and End Results registries with Medicare claims from the Centers for Medicare & Medicaid Services (formerly Health Care Financing Administration) (8). These linkages permit evaluation of the outcomes of surgery for all patients who received a diagnosis of cancer in the geographical regions defined by the registries. Patients were enrolled in Part A and Part B Medicare but were not enrolled in a health maintenance organization, and each had a diagnosis of colon, prostate, or rectal cancer made at age 65 years or older in a region covered by one of the Surveillance, Epidemiology, and End Results registries. All tumors were primary, invasive, and malignant and were diagnosed during 1992 to 1996. The sample sizes for the 3 cohorts ranged from very large (colon cancer,  $n = 24\ 166$ ) to large (prostate cancer,  $n = 10\ 737$ ) to moderate (rectal cancer,  $n = 2603$ ). These analyses en-

compassed 2682, 999, and 1141 surgeons, respectively. The median numbers of patients per surgeon were 5 (range, 1 to 85), 7 (range, 1 to 121), and 1 (range, 1 to 26), respectively. The particular billing codes used to identify the surgeon, the specific surgical procedures, and the outcomes have been described in detail in the original publications (4–7), but our essential approach is summarized as follows.

### Procedure Volume

Surgeon volume was defined as the total number of operations performed by the given surgeon between 1992 and 1996 on members of each cohort, as ascertained from the Medicare files on the basis of International Classification of Diseases, Ninth Revision, procedure codes (9). The Common Procedural Terminology codes submitted in the National Claims History files were used to identify the procedures (10). Surgeon identity was established on the basis of a unique identifier that has been mandatory on all claims since 1991.

### Outcomes

Patient outcomes were binary indicators for 2-year mortality or a particular postoperative complication, depending on the cancer cohort. For the colon and rectal cancer data sets, mortality at 2 years and occurrence of a procedure impairing fecal continence (ileostomy or colostomy in colon cancer and an abdominoperineal resection in rectal cancer) were the outcomes evaluated. These measures of morbidity were selected because they can have an important impact on the patient's subsequent quality of life. For the prostate cancer cohort, postoperative and late urinary complications were examined. Postoperative complications were defined as potentially life-threatening cardiac, respiratory, or vascular events; the need for reoperation; and hemorrhage, renal failure, or shock, all occurring within 30 days of surgery. Late urinary complications were identified by procedures or symptoms recorded more than 30 days after but within 1 year of surgery and included diagnoses of bladder-neck obstruction, urethral or ureteral strictures, intestinal or vesical fistulas, pelvic abscess, and other urinary tract complications that required surgical repair.

### Case Mix

In each primary analysis, clinical and demographic variables were used to adjust for differences in patient case mix among individual surgeons. These included age; sex; ethnicity; disease stage at diagnosis; disease grade; income level; and the presence of particular features such as emergent presentation, obstruction, or perforation. All results shown in the tables and figures are adjusted for case mix.

### Alternative Statistical Methods for Cluster-Related Data

We used and compared 2 general statistical approaches that permit adjustments to account for the effect of clustering of outcomes on statistical significance levels: random-effects models, in which the impact of each surgeon is

### Context

While studies suggest that surgical patients fare best with providers (hospitals and surgeons) that perform a high volume of procedures, most have not accounted for the tendency of patients of 1 provider to have similar outcomes ("clustering"). When outcomes of individual patients are clustered, more patients are required to prove that providers' outcomes differ from one another.

### Contribution

This reanalysis of data from 3 previously published volume–outcome studies of surgery accounted for clustering within provider. Statistical significance of volume–outcome relationships of morbidity end points was attenuated substantially after adjustment for the effects of clustering.

### Implications

Planners considering regionalizing surgery should remember that volume–outcome studies that have not accounted for clustering exaggerate differences in outcomes by provider.

–The Editors

modeled explicitly (11), and generalized estimating equations (GEEs), in which the underlying patient-specific analysis is adjusted to accommodate the effect of clustering on the statistical significance of the analyses (12). The GEE method does not alter the estimate of the magnitude of the impact of surgeon volume on outcome; only the statistical significance (and the width of the confidence interval) is altered. In the random-effects model, the estimate may change after correction for clustering. Both of these methods are widely used in health services research and clinical research. We evaluated the effect of surgeon volume on outcomes in 3 settings: adjustment solely for case mix; adjustment for case mix and hospital volume; and adjustment for case mix, hospital volume, and clustering.

Logistic regression was used throughout to estimate the impact of volume on outcome. The random-effects models were fitted by using the `gllamm6` command in Stata, version 7.0 (Stata Corp., College Station, Texas), and the GEE models were fitted by using the `GENMOD` procedure in SAS, version 8.0 (SAS Institute, Inc., Cary, North Carolina). Volume was modeled as a continuous variable in all analyses. Odds ratios are reported per 100-unit decline in volume increments for the colon and prostate cancer analyses and per 10-unit decline for the rectal cancer analysis.

### A Graphical Assessment and Statistical Test for Clustering of Outcomes

We used a graphical technique to characterize visually the extent of variation among providers. Only surgeons with a minimum number of patients treated were used to ensure that each surgeon had a sufficient number of patients in his

Table 1. Relationship of Surgeon Volume to Outcomes after Surgery for Colon Cancer\*

Outcome	Odds Ratio of Outcome per Decline in Individual Surgeon Volume in 100-Unit Volume Intervals (95% CI)			
	Adjusted for Case Mix	Adjusted for Case Mix and Hospital Volume	Corrected for Clustering of Outcomes within Individual Surgeons†	
			GEE Method‡	Random-Effects Methods§
2-Year mortality	1.80 (1.43–2.28)	1.48 (1.17–1.87)	1.48 (1.14–1.91)	1.47 (1.14–1.90)
P value	<0.001	0.002	0.004	0.003
Ostomy	1.80 (1.40–2.33)	1.48 (1.12–1.94)	1.48 (0.96–2.27)	1.70 (1.12–2.58)
P value	<0.001	0.005	0.083	0.013

\* GEE = generalized estimating equation.

† Corrected for clustering after adjustment for case mix and hospital volume.

‡ Calculated by using the GENMOD procedure in SAS (SAS Institute, Inc., Cary, North Carolina).

§ Calculated by using the gllamm6 command in Stata (Stata Corp., College Station, Texas).

or her profile to allow meaningful estimates of the mean departure from the overall event rate. As a result, we did not construct graphs for the rectal cancer data set, since the vast majority of the surgeons treated very few patients.

The observed outcome relative frequencies, that is, the proportion of patients who experienced the outcome, were evaluated for each surgeon. A conventional multivariable logistic regression analysis that included surgeon volume, hospital volume, and case mix as predictor variables was used to estimate the patient-specific outcome probabilities. The sum of these probabilities for each patient in the surgeon's profile was calculated as the "expected" event rate for the surgeon.

To obtain the graph, the observed frequencies were plotted in a histogram. An "expected" histogram was created to display the outcome histogram that would be expected if there was no additional surgeon-to-surgeon variation after adjustment for known predictors of outcome in the model. For each surgeon, the binomial distribution based on the preceding expected event rate and the number of patients in the surgeon's profile was used to determine probabilities for each possible observed outcome relative frequency (in a histogram grouped by cells of 5 percentage points). For each cell of the histogram, these probabilities were summed across all surgeons to obtain the "expected" histogram. This "expected" histogram reflects the degree of

spread we would expect to observe in the absence of clustering.

The spread of the observed distribution will exceed the spread of the expected distribution in the presence of clustering. To test whether differences between the 2 histograms were statistically significant, the observed frequencies and expected event rates for each surgeon were compared by using a chi-square test. This test evaluated whether the between-surgeon variation in observed outcomes was greater than would be expected on the basis of chance (13). A statistically significant result would indicate the presence of additional surgeon-to-surgeon variation that could not be explained by the factors in the multivariable model, that is, surgeon volume, hospital volume, and case mix. The test is valid only if the expected values are greater than 5. If this assumption is violated, then an exact test with a resampling technique should be used.

### Role of the Funding Source

The funding source had no role in the collection, analysis, and interpretation of the data or in the decision to submit the manuscript for publication.

## RESULTS

Results are shown in Tables 1, 2, and 3 for colon, prostate, and rectal cancer, respectively. For each clinical

Table 2. Relationship of Surgeon Volume to Outcomes after Surgery for Prostate Cancer\*

Outcome	Odds Ratio of Outcome per Decline in Individual Surgeon Volume in 100-Unit Volume Intervals (95% CI)			
	Adjusted for Case Mix	Adjusted for Case Mix and Hospital Volume	Corrected for Clustering of Outcomes within Individual Surgeons†	
			GEE Method‡	Random-Effects Methods§
Postoperative complications	1.73 (1.40–2.15)	1.58 (1.25–2.00)	1.58 (1.16–2.17)	1.46 (1.00–2.14)
P value	<0.001	<0.001	0.004	0.049
Late urinary complications	2.77 (2.19–3.51)	2.32 (1.80–2.99)	2.32 (1.34–4.01)	1.88 (1.19–2.97)
P value	<0.001	<0.001	0.002	0.007

\* GEE = generalized estimating equation.

† Corrected for clustering after adjustment for case mix and hospital volume.

‡ Calculated by using the GENMOD procedure in SAS (SAS Institute, Inc., Cary, North Carolina).

§ Calculated by using the gllamm6 command in Stata (Stata Corp., College Station, Texas).

**Table 3. Relationship of Surgeon Volume to Outcomes after Surgery for Rectal Cancer\***

Outcome	Odds Ratio of Outcome per Decline in Individual Surgeon Volume in 10-Unit Volume Intervals (95% CI)			
	Adjusted for Case Mix	Adjusted for Case Mix and Hospital Volume	Corrected for Clustering of Outcomes within Individual Surgeons†	
			GEE Method‡	Random-Effects Methods§
2-Year mortality	1.51 (1.21–1.89)	1.44 (1.13–1.83)	1.44 (1.17–1.77)	1.44 (1.13–1.83)
<i>P</i> value	<0.001	0.004	<0.001	0.004
Abdominoperineal resection	1.29 (1.09–1.53)	1.21 (1.00–1.46)	1.21 (0.91–1.60)	1.09 (0.81–1.48)
<i>P</i> value	0.003	0.047	0.19	>0.2

\* GEE = generalized estimating equation.

† Corrected for clustering after adjustment for case mix and hospital volume.

‡ Calculated by using the GENMOD procedure in SAS (SAS Institute, Inc., Cary, North Carolina).

§ Calculated by using the gllamm6 command in Stata (Stata Corp., College Station, Texas).

outcome, the results are characterized by the odds ratio of the outcome per 100-unit change in surgeon volume (per 10-unit change for rectal cancer).

### Colon Cancer

The first 2 columns of **Table 1** involve conventional logistic regression analyses adjusting for case mix and for both case mix and hospital volume, respectively. For both end points, adjustment for hospital volume substantially attenuated the strength of the relationship between surgeon volume and outcome, reducing the odds ratio from 1.80 to 1.48. The *P* values were also attenuated but remained highly statistically significant ( $P = 0.002$  and  $P = 0.005$  for 2-year mortality and ostomy, respectively). It should be noted that it is simply a coincidence that both end points lead to identical estimates (1.80 and 1.48) in both cases. The third and fourth columns show the effect of further correction for the effects of clustering, using the GEE and random-effects methods, respectively. The GEE method did not alter the odds ratio, but the analysis led to a widening of the confidence interval and an attenuation of the *P* value. The confidence interval was only slightly widened for 2-year mortality, indicating very little clustering. This is reflected by the similarity of the observed and expected histograms in **Figure 1**, which show that the variation in observed outcomes across surgeons was similar to what would be expected on the basis of random variation (using only surgeons with a minimum caseload of 17 patients). The results of the statistical test to compare these 2 histograms were not significant ( $P > 0.2$ ). Conversely, in **Figure 2**, ostomy rates are more broadly dispersed than the expected rates. For example, 43 surgeons had ostomy rates below 5%, whereas only 23 were expected to have rates this low. Also, several surgeons had rates exceeding 45%, while no rates this high were expected. The differences between these observed and expected frequencies are statistically significant ( $P < 0.001$ ). The substantial clustering is also reflected in the much wider confidence intervals after adjustment for clustering and the attenuation of the *P* value to a nonsignificant value of 0.083 (**Table 1**). Comparison of the third and fourth columns of **Table 1** demonstrates that

both the odds ratio estimate and the *P* value can vary substantially depending on the statistical method used.

### Prostate Cancer

For prostatectomy-related outcomes, the effects of clustering were much more apparent. For both postoperative complications and late urinary complications, there was a substantial difference in the observed and expected distributions (both tests are statistically significant at  $P < 0.001$  in analyses based on a minimum caseload of 20 patients). The most striking difference was observed for late urinary complications (**Figure 3**): Nineteen surgeons had late urinary complication rates below 5% versus only 1% expected, and 11 surgeons had rates greater than 55% versus none expected.

The results in **Table 2** show that the volume–outcome trends were highly statistically significant and remained statistically significant after correction for clustering. However, as for colon cancer, the odds ratios were attenuated considerably by adjustment for hospital volume. There are noticeable differences in the results depending on whether the random-effects method or the GEE method was used.

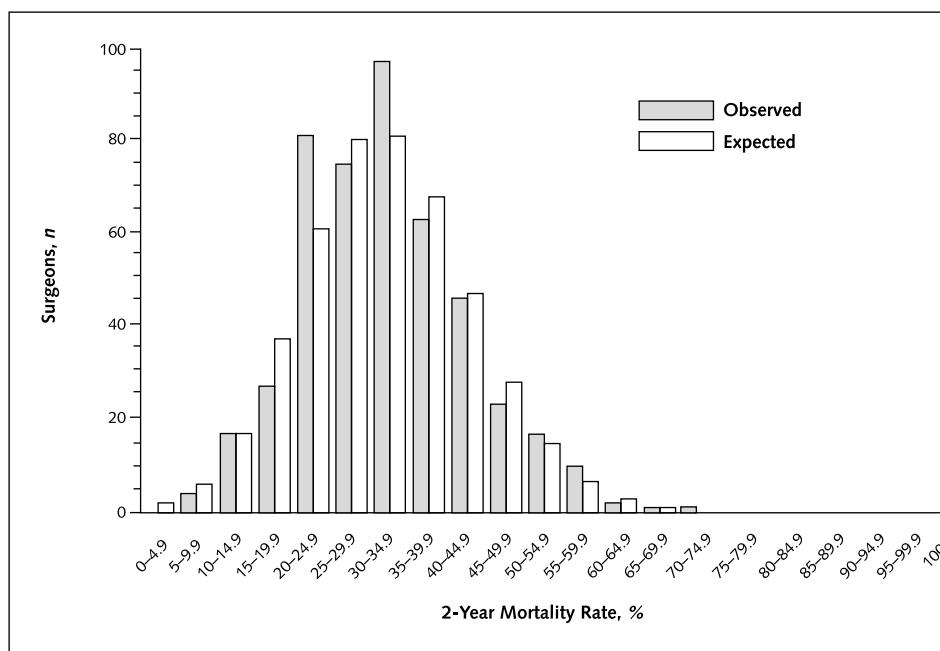
### Rectal Cancer

Adjustment for hospital volume had minimal impact and correction for clustering had essentially no impact on 2-year mortality results, indicating that there was no appreciable between-surgeon variation in outcomes after adjustment for the volume indicators and case mix (**Table 3**). These results contrast with those for the morbidity outcome (abdominoperineal resection). Statistical significance was eliminated after adjustment for clustering, indicating substantial clustering by surgeon. The average number of patients treated per surgeon was too small to permit a meaningful graphical display of the degree of variation in outcomes among surgeons.

### Hospital Volume

We performed an analogous series of analyses to assess the relationship of hospital volume to these outcomes, adjusting for surgeon volume and correcting for clustering of outcomes within hospitals. The trends were similar to

Figure 1. Two superimposed histograms of the numbers of surgeons with observed and expected 2-year mortality rates after colon cancer surgery.



those seen for surgeon volume, with somewhat more pronounced attenuation of statistical significance and odds ratios after adjustments for surgeon volume.

## DISCUSSION

Our analysis highlights important messages that have relevance for the volume–outcome literature in particular and, more broadly, for all researchers studying variation in quality and patterns of care. First, correction for clustering of outcomes within providers can substantially attenuate the statistical significance of an observed volume–outcome trend. We saw this in our analyses of the effect of surgeon volume on the abdominoperineal resection rate after rectal cancer surgery and in corresponding analyses of hospital volume (data not shown). Second, clustering should not be regarded as simply a statistical “nuisance” requiring a standard statistical “adjustment.” The extent of clustering after adjustment for volume and case mix, demonstrated in **Figures 2** and **3**, is likely to reflect variations in surgical skill or technique. Clustering of this nature in studies of hospital volume may reflect important differences in processes of care.

Statistical methods that adjust for the effects of clustering are common in medical research. For example, in clinical trials in which patients are randomly assigned in groups, as in community intervention studies such as the Community Intervention Trial for Smoking Cessation (COMMIT), the focus lies largely with estimating the size of the treatment effect (14, 15). The degree of variation among the communities contributing patients to the trial is of secondary importance and is used to adjust tests of sig-

nificance and confidence intervals for the estimate of the treatment effect. By contrast, in profiling hospitals or surgeons by volume or other characteristics, an important focus of the analysis lies with the degree of variation in outcomes among providers. A key question is whether observed variation in rates is consistent with differences across providers in patient characteristics. We can try to estimate this excess variation and look for the reasons in terms of the quality or organization of health care. These are markedly different perspectives, even if the underlying statistical methods and concepts are similar.

The variations in outcomes among surgeons with similar volumes that we observed in this study can be plausibly interpreted as follows. Some colon cancer surgeons are more likely to perform colostomies (thereby potentially avoiding anastomotic leaks and postoperative infections), while others are substantially more likely to attempt primary reanastomosis. This variation in surgical practice leads directly to observed variation in these outcomes when analyzed on a surgeon-by-surgeon basis. It is likely that analogous variations in surgical “style” or technique explain variations in morbidity after surgery for prostate and rectal cancer. This supports the conclusion that morbidity outcomes vary substantially from surgeon to surgeon even after adjustment for volume. This is an important observation, since it suggests that volume-focused measures to improve outcomes may have limited impact in improving morbidity outcomes.

The large body of literature describing the association between procedure volume and clinically important outcomes spans several disciplines, notably heart disease and

cancer (16–31). The results have led to “regionalization” initiatives that may have major effects on health care delivery (32, 33). These have recently received increasing consideration from private organizations such as the Leapfrog Group and from the Centers for Medicare & Medicaid Services (34, 35). Most of this volume–outcome literature has focused on the effect of hospital volume on surgical mortality rates. The 3 data sets that we reanalyzed reflect the more recent interest of studying surgeon volume alongside hospital volume. Examination of the joint effect of hospital and surgeon volume is challenging because the independent effects are relatively modest and these 2 factors are quite strongly correlated (Pearson correlation coefficients, 0.34, 0.43, and 0.44 for colon, prostate, and rectal cancer, respectively). That is, high-volume surgeons typically practice at high-volume hospitals. Thus, even with large sample sizes, it can be difficult to reliably distinguish the separate effects of surgeons and hospitals. Studies that have analyzed both hospital and surgeon volume, including our own previous work, have not consistently adjusted for both volume measures simultaneously, and this can be problematic given the strong correlation between them. In each of our current analyses, the magnitude of the surgeon volume trend was attenuated after adjustment for hospital volume. Similar results were obtained in analyses of hospital volume adjusted for surgeon volume (data not shown), consistent with another recent study of prostate cancer morbidity (36, 37). In light of these findings, can we be confident that increased hospital volumes will lead to better outcomes? Are we confident enough to take steps to concentrate the care of patients in a few hospitals, or are the data too uncertain to justify this? Our analyses suggest

that decisions regarding triage of patients to high-volume centers should be made with caution.

Although we recommend that attention to clustering should be an important aspect of analyses of volume–outcome studies, the statistical methods are evolving, and their properties are not fully understood in this context. Birkmeyer and colleagues (38), in their recent analysis of the effect of hospital volume on outcomes of several major diseases, corrected for clustering using the GEE method, observing variance inflations not dissimilar to those we observed in this study. Other investigators have used the random-effects method to adjust for clustering (39, 40). Our side-by-side comparison of these 2 approaches yielded some disconcertingly large differences in results, notably for the analyses of ostomy rates (colon cancer), rates of late urinary complications (prostate cancer), and abdominoperineal resection rates (rectal cancer). Both statistical methods endeavor to estimate the same effect, the odds ratio of volume on outcome, and the discrepancies in estimates must reflect their different technical formulations. Although both methods are used widely and have been studied intensively from a theoretical perspective, their applicability in the context of volume–outcome studies has not been studied specifically. In this context, “volume” reflects both the factor under study and the cluster size. As a result, we cannot be fully certain of the statistical validity of these techniques in this context without additional research. Our study does not address the issue of which of the 2 methods is more appropriate.

We reviewed the methods and results of reports that were published between 1995 and 2003 and described the association between hospital or surgeon volume to deter-

Figure 2. Two superimposed histograms of the numbers of surgeons with observed and expected ostomy rates after colon cancer surgery.

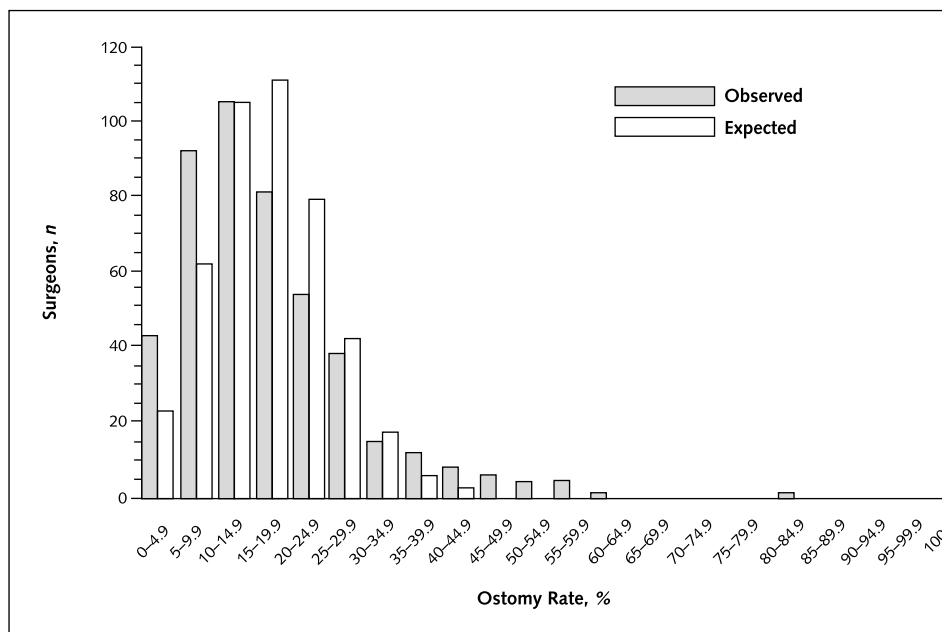
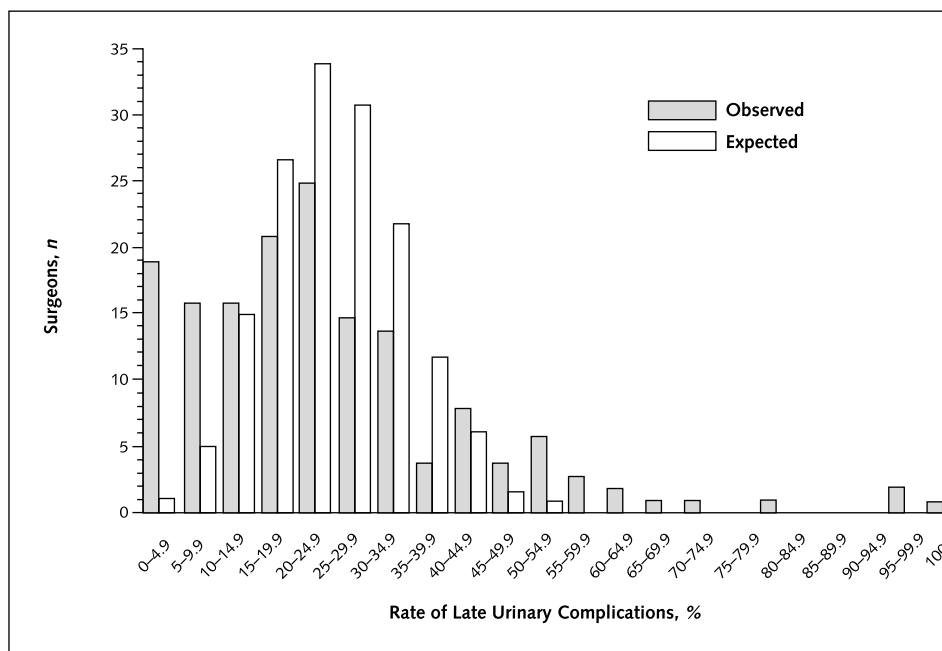


Figure 3. Two superimposed histograms of the numbers of surgeons with observed and expected rates of late urinary complication after prostate cancer surgery.



mine whether correction for clustering was performed and to identify the method used. Using the systematic review of volume–outcome studies by Halm and associates (41), we identified 62 studies from 1995 to 2000; this list was augmented by an additional 10 relevant studies. We found that these studies consistently treated the patient as the unit of analysis and typically assumed that outcomes for patients within a particular hospital or surgeon were independent. Investigators corrected for clustering in only 27% of these studies. Of the studies that adjusted for clustering, the GEE method was more commonly used than the random-effects approach.

In our analyses, we adjusted for hospital volume by including it as a predictor in the model. However, we did not simultaneously correct for clustering of outcomes within individual surgeons and within individual hospitals. Multilevel clustering can be evaluated by using hierarchical random-effects models, and these methods are also used widely, notably in the social science literature. However, surgeon and hospital clustering is not “hierarchical” because many surgeons operate at several hospitals (in our study, this was true of 30%, 37%, and 12% of the surgeons in the colon, prostate, and rectal cancer cohorts, respectively). Although a method for analyzing “cross-classified” binary outcome data of this nature is available in an SAS macro (the GLIMMIX macro) (42), it is extremely computer intensive and its validity has been challenged (43).

Our methodologic advice is directed at individual studies of the relationship between volume and outcome. Since the impact of clustering typically attenuates the strength of the statistical significance, and since volume-

outcome trends are frequently modest, large sample sizes are needed to establish many of these trends in a convincing way. However, volume–outcome relationships have been evaluated in numerous studies, and these studies are overwhelmingly consistent in demonstrating a positive association between hospital volume and selected widely studied outcomes (38). Our findings do not repudiate this substantial literature. However, they indicate that individual studies of specific volume–outcome relationships need to be interpreted cautiously, with due attention to the effects of clustering. Estimation of the degree of clustering is of critical importance in and of itself because it indicates unexplained variation in outcomes. Where there are such variations, there may be opportunities for improving the quality of care.

From Memorial Sloan-Kettering Cancer Center, New York, New York.

**Acknowledgments:** The authors thank 2 anonymous reviewers for help with the article. They also thank the groups responsible for the creation and dissemination of the linked database, including the Applied Research Branch, Division of Cancer Control and Population Sciences, National Cancer Institute; the Office of Strategic Planning and the Office of Informational Services, Centers for Medicare & Medicaid Services; Information Management Services; and the Surveillance, Epidemiology, and End Results tumor registries.

**Grant Support:** In part by grants from the National Cancer Institute (CA83950 [Dr. Schrag], CA90226 [Dr. Bach], and CA08748 [Dr. Begg]).

**Potential Financial Conflicts of Interest:** None disclosed.

**Requests for Single Reprints:** Colin B. Begg, PhD, Memorial Sloan-Kettering Cancer Center, 307 East 63rd Street (3rd Floor), New York, NY 10021; e-mail, [beggc@mskcc.org](mailto:beggc@mskcc.org).

Current author addresses and author contributions are available at [www.annals.org](http://www.annals.org).

## References

- Ashby M, Neuhaus JM, Hauck WW, Bacchetti P, Heilbron DC, Jewell NP, et al. An annotated bibliography of methods for analysing correlated categorical data. *Stat Med*. 1992;11:67-99. [PMID: 1557577]
- Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med*. 2001;135:112-23. [PMID: 11453711]
- Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. *Ann Intern Med*. 2002;136:111-21. [PMID: 11790062]
- Schrag D, Cramer LD, Bach PB, Cohen AM, Warren JL, Begg CB. Influence of hospital procedure volume on outcomes following surgery for colon cancer. *JAMA*. 2000;284:3028-35. [PMID: 11122590]
- Schrag D, Panageas KS, Riedel E, Cramer LD, Guillem JG, Bach PB, et al. Hospital and surgeon procedure volume as predictors of outcome following rectal cancer resection. *Ann Surg*. 2002;236:583-92. [PMID: 12409664]
- Begg CB, Riedel ER, Bach PB, Kattan MW, Schrag D, Warren JL, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med*. 2002;346:1138-44. [PMID: 11948274]
- Schrag D, Panageas KS, Riedel E, Hsieh L, Bach PB, Guillem JG, et al. Surgeon volume compared to hospital volume as a predictor of outcome following primary colon cancer resection. *J Surg Oncol*. 2003;83:68-78; discussion 78-9. [PMID: 12772198]
- Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare-tumor registry database. *Med Care*. 1993;31:732-48. [PMID: 8336512]
- International Classification of Diseases for Oncology. 2nd ed. Geneva: World Health Organization; 1990.
- Current Procedural Terminology. Chicago, IL: American Med Assoc; 1999.
- Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993;88:9-25.
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13-22.
- Hosmer DW Jr, Lemeshow S. *Applied Logistic Regression*. New York: Wiley; 1989.
- Donner A, Donald A. Analysis of data arising from a stratified design with the cluster as unit of randomization. *Stat Med*. 1987;6:43-52. [PMID: 3576016]
- Green SB, Corle DK, Gail MH, Mark SD, Pee D, Freedman LS, et al. Interplay between design and analysis for behavioral intervention trials with community as the unit of randomization. *Am J Epidemiol*. 1995;142:587-93. [PMID: 7653466]
- Bach PB, Cramer LD, Schrag D, Downey RJ, Gelfand SE, Begg CB. The influence of hospital volume on survival after resection for lung cancer. *N Engl J Med*. 2001;345:181-8. [PMID: 11463014]
- Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. *JAMA*. 1998;280:1747-51. [PMID: 9842949]
- Birkmeyer JD, Warshaw AL, Finlayson SR, Grove MR, Tosteson AN. Relationship between hospital volume and late survival after pancreaticoduodenectomy. *Surgery*. 1999;126:178-83. [PMID: 10455881]
- Choti MA, Bowman HM, Pitt HA, Sosa JA, Sitzmann JV, Cameron JL, et al. Should hepatic resections be performed at high-volume referral centers? *J Gastrointest Surg*. 1998;2:11-20. [PMID: 9841963]
- Dimick JB, Cattaneo SM, Lipsett PA, Pronovost PJ, Heitmiller RF. Hospital volume is related to clinical and economic outcomes of esophageal resection in Maryland. *Ann Thorac Surg*. 2001;72:334-9; discussion 339-41. [PMID: 11515862]
- Edwards EB, Roberts JP, McBride MA, Schulak JA, Hunsicker LG. The effect of the volume of procedures at transplantation centers on mortality after liver transplantation. *N Engl J Med*. 1999;341:2049-53. [PMID: 10615076]
- Glasgow RE, Showstack JA, Katz PP, Corvera CU, Warren RS, Mulvihill SJ. The relationship between hospital volume and outcomes of hepatic resection for hepatocellular carcinoma. *Arch Surg*. 1999;134:30-5. [PMID: 9927127]
- Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. *West J Med*. 1996;165:294-300. [PMID: 8993200]
- Hannan EL, O'Donnell JF, Kilburn H Jr, Bernard HR, Yazici A. Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. *JAMA*. 1989;262:503-10. [PMID: 2491412]
- Hannan EL, Siu AL, Kumar D, Kilburn H Jr, Chassin MR. The decline in coronary artery bypass graft surgery mortality in New York State. The role of surgeon volume. *JAMA*. 1995;273:209-13. [PMID: 7807659]
- Hannan EL, Racz M, Ryan TJ, McCallister BD, Johnson LW, Arani DT, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. *JAMA*. 1997;277:892-8. [PMID: 9062327]
- Hannan EL, Popp AJ, Tranmer B, Fuestel P, Waldman J, Shah D. Relationship between provider volume and mortality for carotid endarterectomies in New York state. *Stroke*. 1998;29:2292-7. [PMID: 9804636]
- Yao SL, Lu-Yao G. Population-based study of relationships between hospital volume of prostatectomies, patient outcomes, and length of hospital stay. *J Natl Cancer Inst*. 1999;91:1950-6. [PMID: 10564679]
- Roohan PJ, Bickell NA, Baptiste MS, Therriault GD, Ferrara EP, Siu AL. Hospital volume differences and five-year survival from breast cancer. *Am J Public Health*. 1998;88:454-7. [PMID: 9518982]
- Polednak AP. Hospital volume and survival of breast cancer patients in Connecticut [Letter]. *Am J Public Health*. 1999;89:946-7. [PMID: 10358694]
- Thiemann DR, Coresh J, Oetgen WJ, Powe NR. The association between hospital volume and survival after acute myocardial infarction in elderly patients. *N Engl J Med*. 1999;340:1640-8. [PMID: 10341277]
- Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA*. 1994;271:761-6. [PMID: 8114213]
- Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State's Cardiac Surgery Reporting System: four years later. *Ann Thorac Surg*. 1994;58:1852-7. [PMID: 7979781]
- Birkmeyer JD, Finlayson EV, Birkmeyer CM. Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. *Surgery*. 2001;130:415-22. [PMID: 11562662]
- [www.leapfroggroup.org/](http://www.leapfroggroup.org/). Accessed on 7 May 2003.
- Hu JC, Gold KF, Pashos CL, Mehta SS, Litwin MS. Role of surgeon volume in radical prostatectomy outcomes. *J Clin Oncol*. 2003;21:401-5. [PMID: 12560426]
- Begg CB, Scardino PT. Taking stock of volume-outcome studies [Editorial]. *J Clin Oncol*. 2003;21:393-4. [PMID: 12560422]
- Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128-37. [PMID: 11948273]
- McGrath PD, Wennberg DE, Dickens JD Jr, Siewers AE, Lucas FL, Malenka DJ, et al. Relation between operator and hospital volume and outcomes following percutaneous coronary interventions in the era of the coronary stent. *JAMA*. 2000;284:3139-44. [PMID: 11135777]
- Kee F, Wilson RH, Harper C, Patterson CC, McCallion K, Houston RF, et al. Influence of hospital and clinician workload on survival from colorectal cancer: cohort study. *BMJ*. 1999;318:1381-5. [PMID: 10334746]
- Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*. 2002;137:511-20. [PMID: 12230353]
- Littell R, Millikan GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*. Cary, NC: SAS Institute; 1996.
- Rodriguez G, Goldman N. An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*. 1995;158:73-89.

---

**Current Author Addresses:** Drs. Panageas, Schrag, Bach, and Begg and Ms. Riedel: Memorial Sloan-Kettering Cancer Center, 307 East 63rd Street (3rd Floor), New York, NY 10021.

**Author Contributions:** Conception and design: K.S. Panageas, D. Schrag, E. Riedel, C.B. Begg.  
Analysis and interpretation of the data: K.S. Panageas, D. Schrag, E. Riedel, C.B. Begg.  
Drafting of the article: K.S. Panageas, D. Schrag, C.B. Begg.

Critical revision of the article for important intellectual content: K.S. Panageas, D. Schrag, P.B. Bach, C.B. Begg.

Final approval of the article: K.S. Panageas, P.B. Bach, C.B. Begg.

Provision of study materials or patients: D. Schrag, C.B. Begg.

Statistical expertise: K.S. Panageas, E. Riedel, C.B. Begg.

Obtaining of funding: C.B. Begg.

Administrative, technical, or logistic support: C.B. Begg.

Collection and assembly of data: D. Schrag, C.B. Begg.