Hospital Diagnostic Aggregation and Risk-Adjusted Quality

Chun Lok K. Li

Objective. To demonstrate the importance of diagnostic aggregation when assessing hospitals.

Data Sources. Patient data from the Victorian Admitted Episodes Database (VAED), 1999/2000 to 2004/2005. Financial statements from public hospitals, 2002/2003 to 2004/2005.

Study Design. Risk-adjusted quality computed for each hospital using two aggregation levels. Each is then used to estimate the relationship between hospital efficiency and quality using two-stage DEA/Tobit model by Wilson and Simar (2006).

Data Collection. Selected variables from the VAED were obtained from the Department of Health in Victoria, then linked anonymously with financial statements.

Principal Findings. Hospital quality and, in some cases, its relationship with efficiency differs depending on aggregations.

Conclusions. Patient risk adjustment should be conducted using more than one aggregation level whenever possible.

Key Words. Risk adjustment, diagnostic aggregation, hospital quality

Risk-adjusted quality measures are frequently used in empirical settings. To do this, patients are grouped into diagnostic aggregations because there are too many to individually account for. This study aims to demonstrate the importance of the choice of aggregation method in the context of risk adjustment for empirical analysis. It is necessary because there is little discussion in the literature as to how aggregation affects quality estimates.

One of the earliest well-known examples was a discussion on how the unit of time measurement can produce very different incidence measures (Farr 1838). What makes it interesting is perceived relevance of Farr's study even after all the years, along with its supporters (Vandenbroucke and Vandenbroucke-Grauls 1996) and detractors (Iezzoni 1996). More recently, the level of interest in statistical methods for computing them seemed to have decreased (Gross et al. 1996), opting instead to use tools developed in other disciplines.

To risk adjust, patients have to be aggregated for statistical analysis, but this step is seldom discussed. A well-known example where it has been more explicitly mentioned is a study that compares the differences in patient aggregation definitions between records from the Department of Veterans Affairs and the Medicare system, both of which are datasets commonly used in the United States (Burgess et al. 2011). But even there, the significance of aggregation in a risk-adjustment context is not mentioned.

Another motivating example of the significance of aggregation comes from an article investigating a policy change that demanded increased transparency in hospital quality statistics (Dranove et al. 2003). Their aim was to measure the effect of increased transparency on physician and patient choice, with a concern that difficult patients would be turned away by hospitals because accepting them may have a negative impact on their hospital report card. The performance evaluation method is a secondary issue.

However, readers may wonder whether the risk-adjustment mechanism behind the quality measure sufficiently controlled for the differences in patient makeup. This question is important because, if the quality measure was fair to hospitals treating sicker patients, then they may be less averse to accepting high-risk patients even with increased transparency. Therefore, some form of robustness checking to verify the validity of the risk-adjustment process would have strengthened the results of that study. And if such robustness checks were in place, the aggregation method would have definitely been examined, as it is one of the most important factors for patient risk adjustment.

Hence, we believe that this is a suitable empirical setting to demonstrate the importance of aggregation. The next section will explain how this is done.

METHODS

The analysis is designed to demonstrate the importance of aggregation in two stages. First, we will compute risk-adjusted measures of hospital quality using two aggregation levels and compare the differences. Second, we will apply these measures to estimate the relationship between hospital quality and efficiency. The emphasis is on comparing differences between estimated results

Address correspondence to Chun Lok Kris Li, Ph.D. Economics, Level 5, Faculty of Business and Economics Building, 111 Barry Street, The University of Melbourne, Victoria 3010, Australia; e-mails: clli@student.unimelb.edu.au; krisli@sqwchinagroup.com.

generated using quality measures computed under different aggregations. A data description follows.

Risk-Adjusted Quality Measure: The Hospital Standardized Mortality Ratio (HSMR)

Hospitals with sicker patients are expected to have higher incidence rates. We can use the hospital standardized mortality ratio (HSMR), which can be applied to any type of episode, important because there is no agreed-upon method for appropriating hospital overheads to each episode.

Let h and p index, respectively, hospitals and patient admissions, $h = 1, \dots, H$ and $a = 1, \dots, A$. The patient associated with each admission is represented by a vector of demographic characteristics δ_{ha} with the marginal effect of each element in δ_{ha} represented by the corresponding element in the vector λ . Each time a patient is admitted, it is discharged alive ($D_{ha} = 0$) or dead ($D_{ha} = 1$). The probability of death for admission a in hospital h is represented by $E(D_{ha})$. The functional form for the logistic regression is given as follows:

$$E(D_{hp}) = \exp(\lambda' \delta_{ha}) / 1 + \exp(\lambda' \delta_{ha})$$

Assume for simplicity that the expected number of deaths in a hospital is the sum of the probabilities of all its patients. It is denoted as $E(D_h)$, defined as follows:

$$E(D_h) = \sum_a (E(D_{ha})); a = 1 \text{ to } A_h$$

The definition of HSMR for each hospital follows:

$$\text{HSMR}_h = D_h / E(D_h) \in [0, \infty)$$

Each hospital will have an HSMR estimate for each level of aggregation. The higher the HSMR, the worse the hospital quality is. A total of three increasingly refined aggregation levels are used, to be explained later in the data description. This change in aggregation level will be reflected in different estimates of $E(D_h)$ and consequently HSMR values.

The main weakness of HSMR is it only captures immediate patient deaths, without capturing incidents that occur during later episodes that result from suboptimal treatment quality at an earlier episode (Bottle, Jarman, and Aylin 2011). It does not attempt to attribute blame between hospitals if a patient died after being treated in multiple locations either (Kahn, Kramer, and Rubenfeld 2007). And if one medical procedure results in complications that are classified under a different ICD code, it is counted as a separate

episode and the HSMR may be artificially lower than it should because the denominator will be enlarged.

Relationships between Hospital Quality and Efficiency

Besides direct comparisons, hospital quality measures are also used as variables for answering other research questions. Therefore, if the estimated values of quality variables can be affected by aggregation method, then they might also have an effect on the results of any subsequent estimation. This study will investigate this possibility using the typical setting of a study that estimates the relationship between hospital quality and efficiency. The goal is to compare results across broader versus narrower aggregation methods.

We use a two-stage model with a first-stage data-envelopment analysis (DEA) followed by a second-stage Tobit regression bounded by full efficiency. This gives us two sets of results to study with, the technical efficiency of each hospital and factors driving those efficiency scores.

The first stage is a DEA frontier model solved using linear programming, commonly used in health services research. Details of the estimation process can be found in operations research textbooks (e.g., Banker, Charnes, and Cooper 1984; Ramanathan and Ramanathan 2003). We use input-oriented, variable-returns-to-scale efficiency scores with BCC specification (Charnes et al. 1994). In practice, differences in DEA specification used seldom produce significant variations in the results (Golany and Yu 1997).

Quality measures are introduced as input variables, representing the extra effort required for hospitals to improve service. Specifically, the negative of the HSMR rating is included because the less negative the HSMR rating is, the better the hospital quality. Details of the analytical properties associated with similar approaches in quality-adjusted frontier analysis are available in Eckermann and Coelli (2013).

Frontier estimates also contain statistical error, but this not naturally using DEA because it is deterministic. This study will address the above by simulating the DEA distribution using naive bootstrap with subsamples (Wilson and Simar 2006). Recall that there are H hospitals in the entire sample. For each hospital, 2,000 bootstrap resamples are generated, indexed $n = 1, \dots, N$. Each resample contains R hospitals drawn with replacement, where R is less than H . We set R at one-third of the sample size. Notably, Wilson and Simar (2006) stated that R should be less than H , but no explanation or guidelines for choosing R were given.

The second stage is concerned with the factors that are related to the technical efficiency of hospitals. This is estimated using a bootstrap-adjusted Tobit regression as specified by Simar and Wilson (2007). Note that the method requires efficiency scores to be inverted by transforming $[0, 1]$ into $[1, \infty)$, to reduce the problem from double-bounds to single-bound. Point estimates and BCa percentile intervals (Efron, 1987) are presented in the findings section.

Data Description: The Victorian Admitted Episodes Database

Two sources of data are used in this study, a patient admissions database and a list of assembled hospital financial statements. Patient data are extracted from the Victorian Admitted Episodes Database (VAED) for the years 1999–2004. It contains the typical information such as patient demographics and illness types.

We use the variables age, gender (Chambers and Clarke 1990), and the ICD-10-based Charlson index for risk adjustment, using logistic regression. Each aggregation is separately estimated. Patient names and identities of private hospitals are encoded due to confidentiality reasons, eliminating the possibility of panel estimation techniques.

A useful feature of this data is the two-tiered hierarchy of aggregations. The smaller one is the diagnostic-related group (AR-DRG V5.0), of which a total of 648 DRGs were included, leaving out a few administrative entries. These DRGs are then aggregated into 23 broader major diagnostic categories (MDCs) based on VAED definitions during the study period.

The other data source required for estimating hospital efficiency is operating expenses. These are extracted from the financial statements for public hospitals in Victoria during the years 2002–2004. The data are separated into metropolitan and rural hospitals in Table 1.

In the first stage of the analysis for estimating hospital efficiency, financial statement entries are aggregated into the three input prices of capital (K), labor (L), and materials (M). This leaves a total of five variables in the frontier estimate: three input types in dollars, output in number of patient-days, and quality in estimated HSMR values.

Hospitals are then further separated into two hospital types, metropolitan and rural hospitals, with classifications taken directly from the Australian Department of Health. This separation is relevant because Victoria has a clear contrast between urban and rural regions, leading to very different operational needs and funding arrangements (Commission 2010). Results

generated from both the entire hospital set and from the separated metropolitan/rural sets will be presented.

Quality is one of the five variables used to estimate the frontier, measured by the HSMR statistic of each hospital. The risk-adjustment portion of the HSMR can be estimated using one of three tiers of aggregations. That is, for each aggregation tier, patients are divided into their corresponding aggregations. The expected mortality of patients from each aggregation is estimated using logistic regression. This produces three versions of expected mortality for each patient, each culminating into different HSMR estimates for the hospitals. Other variables remain unchanged each time the frontier analysis is conducted.

In the second stage, explanatory variables used include the quality measure, a computed Herschmann–Herfinhdal index (HHI), and a dummy variable for teaching hospital status. Quality is the primary variable and is included

Table 1: Summary Statistics of Hospital Finances and Types of Patients

	<i>Pooled</i>		<i>Metro</i>		<i>Rural</i>	
	<i>Mean</i>	<i>Standard Deviation</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Mean</i>	<i>Standard Deviation</i>
Capital ('000)	2,868	4,428	5,362	6,273	543	372
Labor ('000)	40,582	61,890	87,366	80,688	5,800	3,870
Materials ('000)	11,296	19,149	26,090	25,685	1,116	648
Acute	26,369	42,755	57,017	56,983	2,256	1,940
Maternity	3,753	9,341	8,075	14,618	245	271
Alcohol/drugs	2,820	5,407	6,484	7,199	129	111
Rehabilitation	5,700	9,836	12,728	13,679	654	677
Mort. rate (%)	3.52	7.57	4.60	12.10	2.22	1.58
Quality (MDC)	1.3879	0.8820	1.3769	1.2776	1.3934	0.5937
Quality (DRG)	1.2675	0.6108	1.0876	0.7041	1.3584	0.5372
Data points (2002)	114		33		81	
Data points (2003)	115		34		81	
Data points (2004)	69		33		36	

Note. Profile of basic statistics for hospitals in the VAED sample for the years 2002–2004. The sample is treated as a cross-section because hospitals cannot be identified. This may result in underestimation of statistical dispersion during empirical analysis and make the sampled hospitals appear more similar than they are in reality. The reduction in number of rural hospitals in 2004 is an administrative-level change; no major rural service centers were closed down that year. Dollar values are not adjusted for inflation. “Quality” refers to HSMR values that are separately computed using two different levels of aggregation during the risk-adjustment stage. See the Methods section for details.

by default. HHI is a common measure of hospital competition and is usually considered to be an important factor for efficiency. This study computes an adjusted HHI according to Zwanziger and Melnick (1988). Finally, teaching status is, according to Huerta et al. (2008), a basic measure of “organizational and market characteristics” that is widely considered to affect efficiency.

RESULTS

Estimates of Hospital Quality by Level of Aggregation

Mean and variance of the collective HSMR values of each hospital are listed in Table 1. In general, the more refined the aggregation method, the less varied the estimated quality scores become. A consistent explanation is that a part of the perceived differences in quality can be attributed to unobserved patient differences. This effect is stronger for metropolitan hospitals, perhaps because they deal with more severe cases.

Effect of Aggregation on Hospital Quality

The empirical setting is a two-stage DEA/Tobit analysis of efficiency. Summary statistics of first-stage results are included in Table 2. Three specifications are estimated: MDC, DRG, and a control set without quality.

There are two main observations. The first one is a confirmation that quality is important. This can be seen by the large increase in efficiency score after quality is included. This is consistent with the notion that hospitals with sicker patients may appear to be less efficient. The second is that a change in aggregation did not make a large difference to hospital efficiency levels. The

Table 2: Mean and Variance of Hospital HSMR Values

<i>Hospital Type</i>	<i>Pooled</i>		<i>Metro</i>		<i>Rural</i>	
	<i>Mean</i>	<i>Standard Deviation</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Mean</i>	<i>Standard Deviation</i>
No quality variable	0.3936	0.2109	0.5609	0.2478	0.5568	0.1811
MDC (23)	0.5634	0.1878	0.7725	0.1674	0.6237	0.1582
DRG (570~)	0.5647	0.1856	0.7685	0.1565	0.6335	0.1529
Corr (DRG, MDC)	0.9615		0.9602		0.9688	

Note. Estimates of hospital efficiency by aggregation level using pooled, metropolitan, and rural hospitals. Mean and standard deviations of the scores are presented, as well as the correlation coefficients between the aggregation levels.

mean efficiency scores were similar across the aggregation methods and the correlation coefficients between the different sets of efficiency scores are close to 1. A larger dataset of hospitals could take this analysis further.

Effect of Aggregation on the Relationships between Quality and Efficiency

The second stage examines the relationship between quality and efficiency, with hospital-level factors added as control variables. The result of interest from the Tobit regression is whether the 95 percent CIs of the quality parameters overlap or not. It is a measure of whether aggregation has a statistically significant impact on the quality–efficiency relationship. Parameter results are listed in Table 3.

For the pooled data, estimates obtained at the DRG level are significantly different from those under the MDC and service-type levels. This is consistent with our claim that the choice of aggregation mechanism can have a significant effect on subsequent empirical analysis. As for the separated regression, the 95 percent CIs for the MDC and DRG categories are no

Table 3: Parameter Estimates of Independent Variables Driving Hospital Efficiency

	<i>Constant</i>	<i>Quality</i>	<i>HHI</i>	<i>Teaching</i>
MDC				
Pooled	1.3502 [1.4779, 1.7373]	*0.3726 [0.3120, 0.4515]	0.7885 [−0.1868, 0.2603]	−0.3067 [−0.1428, −0.2029]
Metro	1.1619 [1.3461, 1.6294]	*0.2194 [0.1533, 0.2809]	−3.2978 [−0.1111, 0.2746]	−0.0809 [−0.0040, −0.0959]
Rural	1.2745 [1.3896, 1.7576]	*0.2045 [0.1393, 0.3527]	*0.5668 [0.3863, 0.9686]	
DRG				
Pooled	1.2181 [1.3374, 1.6277]	*0.5806 [0.4829, 0.6933]	0.2955 [−0.2325, 0.1889]	−0.3472 [−0.1859, −0.2941]
Metro	1.1951 [1.3007, 1.5602]	*0.2345 [0.1572, 0.3505]	−3.4066 [−0.1352, 0.1525]	−0.1277 [−0.1144, −0.0866]
Rural	1.1876 [1.2714, 1.5911]	*0.2938 [0.2218, 0.4347]	*0.3346 [0.2140, 0.8070]	

Note. Parameter estimates and their approximate 95% CIs of the second-stage analysis. Results with * indicate statistical significance at the 95% level. The dependent variable is the inefficiency score of hospitals, defined as the reciprocal of efficiency, to simplify the Tobit regression. Note that lower and upper bounds of bootstrapped CIs are, in general, not symmetrical around the point estimate because the normal distribution assumption is relaxed. Similarly, point estimates with a skewed underlying distribution may fall outside a 95% CI, indicating that lower and upper bounds estimated using the normal distribution assumption would have been grossly inaccurate.

longer disjoint. This may be caused by a reduction in data points, or that the differences in risk characteristics of the various DRGs within most MDCs coincide with a metropolitan-rural divide. Nevertheless, there may be other empirical settings where the sample cannot be divided in a similar way, and as such, it is still important to conduct risk adjustment using different possible aggregations, if possible, to ensure that the results are not dependent on this empirical detail.

DISCUSSION

Researchers should design their studies to account for the possibility that results may be sensitive to the choice of aggregation for computing risk-adjusted quality. Adjustments to the aggregation method within the study, if possible, would improve the robustness of the results. Similarly, policy makers should insist that some form of cross-comparison be conducted with a study or against similar studies using significantly different aggregation mechanisms. More adequate funding for creating and maintaining more detailed patient databases would also help.

To show this, we first estimated hospital risk-adjusted quality measures using two separate levels of aggregation: 23 MDCs and 648 DRGs. The marginal factors for probability of death are estimated separately in each aggregation, using logistic regression. Then, these quality measures were used in a two-stage DEA/Tobit analysis, first to estimate quality-adjusted efficiency, then to investigate factors that explain these differences in efficiency. Every stage of the exercise is repeated for all levels of aggregation so that their results can be compared.

The variance in reported quality between hospitals decreased when using the aggregation with smaller, more numerous groups. If a researcher has reason to be concerned with bias in terms of hospital evaluation, a narrower aggregation should be used; indeed, this is consistent with the common belief that unobserved patient heterogeneity can exaggerate differences in hospital performance. However, it is also important to be aware of the bias-variance trade-off in that the standard error of the individual (estimated) quality measures themselves may be higher as more explanatory factors are added.

Finally, we note that the design of the empirical demonstration was constrained by the hospital costing data, which provide total dollar amounts for their inputs. This meant that prices had to be imputed, as

unit quantities are unavailable, which may have reduced the accuracy of the results.

ACKNOWLEDGMENTS

Joint Acknowledgments/Disclosures Statement: The author acknowledges the research funding support provided by the Melbourne Institute of Applied Economic and Social Research.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Banker, R., A. Charnes, and W. Cooper. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 30 (9): 1078–92.
- Bottle, A., B. Jarman, and P. Aylin. 2011. "Strengths and Weaknesses of Hospital Standardised Mortality Ratios." *British Medical Journal* 342: 749–52.
- Burgess, J. F., M. L. Maciejewski, C. L. Bryson, M. Chapko, J. C. Fortney, M. Perkins, N. D. Sharp, and C. F. Liu. 2011. "Importance of Health System Context for Evaluating Utilization Patterns across Systems." *Health Economics* 20 (2): 239–51.
- Chambers, M., and A. Clarke. 1990. "Measuring Readmission Rates." *British Medical Journal* 301 (6761): 1134–6.
- Charnes, A., W. W. Cooper, A. Y. Lewin, and L. M. Seiford. 1994. *Data Envelopment Analysis: Theory, Methodology, and Application*. Boston, MA: Kluwer Academic Publishers.
- Commission, P. 2010. *Public and Private Hospitals: Multivariate Analysis, Supplement to Research Report*. Technical report. Canberra, Australia: Commission.
- Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite. 2003. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers." *Journal of Political Economy* 111: 555–88.
- Eckermann, S., and T. Coelli. 2013. "Including Quality Attributes in Efficiency Measures Consistent with Net Benefit: Creating Incentives for Evidence Based Medicine in Practice." *Social Science and Medicine* 76: 159–68.
- Efron, B. 1987. "Better Bootstrap Confidence Intervals (with Discussion)." *Journal of the American Statistical Association* 82: 171–85.
- Farr, W. 1838. On Prognosis. *British Medical Almanack* 1: 199–218.
- Golany, B., and G. Yu. 1997. "Estimating Returns to Scale in DEA." *European Journal of Operational Research* 103: 28–37.

- Gross, P. A., B. I. Braun, S. B. Kritchevsky, and B. P. Simmons. 1996. "Comparison of Clinical Indicators for Performance Measurement of Health Care Quality: A Cautionary Note." *British Journal of Clinical Governance* 5: 202–11.
- Huerta, T. R., E. W. Ford, L. T. Peterson, and K. H. Brigham. 2008. "Testing the Hospital Value Proposition: An Empirical Analysis of Efficiency and Quality." *Health Care Management Review* 4: 341–9.
- Iezzoni, L. I. 1996. "100 Apples Divided by 15 Red Herrings: A Cautionary Tale from the Mid-19th Century on Comparing Hospital Mortality Rates." *Annals of Internal Medicine* 124 (12): 1079–85.
- Kahn, J. M., A. A. Kramer, and G. D. Rubenfeld. 2007. "Transferring Critically Ill Patients out of Hospital Improves the Standardized Mortality Ratio." *Chest* 131: 68–75.
- Ramanathan, R., and R. Ramanathan. 2003. *An Introduction to Data Envelopment Analysis: A Tool for Performance Measurement*. New Delhi, India: SAGE Publications.
- Simar, L., and P. W. Wilson. 2007. "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes." *Journal of Econometrics* 136 (1): 31–64.
- Vandenbroucke, J. P., and C. M. Vandenbroucke-Grauls. 1996. "In Defense of Farr and Nightingale." *Annals of Internal Medicine* 125 (12): 1014.
- Wilson, P. W., and L. Simar. 2006. "Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives." In *The Measurement of Productive Efficiency, Chapter 4*, 2d Edition, edited by H. O. Fried, K. C. Lovell, and S. S. Schmidt, pp 421–521. Oxford, England: Oxford University Press.
- Zwanziger, J., and G. A. Melnick. 1988. "The Effects of Hospital Competition and the Medicare PPS Program on Hospital Cost Behavior in California." *Journal of Health Economics* 7 (4): 301–20.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table S1. Overview of MDCs in the VAED, 1999/2000 to 2004/2005.

Table S2. VAED Mapping of AR-DRG Aggregations into MDCs, 1999/2000 to 2004/2005.