

NBER WORKING PAPER SERIES

COMPARING HOSPITAL QUALITY AT
FOR-PROFIT AND NOT-FOR-PROFIT
HOSPITALS

Mark McClellan
Douglas Staiger

Working Paper 7324
<http://www.nber.org/papers/w7324>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1999

This paper was prepared for the NBER not-for-profit hospitals conference, Palm Beach FL, Nov. 20-22 1997. Our thanks to David Cutler, Karen Norberg, Catherine Wolfram and seminar participants at the NBER and various universities for their helpful comments. We thank Dhara Shah and Yu-Chu Shen for outstanding research assistance, and the Health Care Financing Administration and the National Institute on Aging for financial support. All errors are our own. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 1999 by Mark McClellan and Douglas Staiger. All rights reserved. Short sections of text, not to exceed

two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals

Mark McClellan and Douglas Staiger

NBER Working Paper No. 7324

August 1999

JEL No. I10, L31, L15

ABSTRACT

Do not-for-profit hospitals provide better care than for-profit hospitals? We compare patient outcomes in for-profit and not-for-profit hospitals between 1984 and 1994 using a new method for estimating differences across hospitals that yields far more accurate estimates of hospital quality than previously available. We find that, on average, for-profit hospitals have higher mortality among elderly patients with heart disease, and that this difference has grown over the last decade. However, much of the difference appears to be associated with the location of for-profit hospitals. Within specific markets, for-profit ownership appears if anything to be associated with better quality care. Moreover, the small average difference in mortality between for-profit and not-for-profit hospitals masks an enormous amount of variation in mortality within each of these ownership types. Overall, these results suggest that factors other than for-profit status *per se* may be the main determinants of quality of care in hospitals.

Mark McClellan
Department of Economics
Stanford University
Palo Alto, CA 94305
and NBER
markmc@texas.stanford.edu

Douglas Staiger
Department of Economics
Dartmouth College
Hanover, NH 03755
and NBER
doug.staiger@dartmouth.edu

1. Introduction

Do not-for-profit hospitals provide better care than for-profit hospitals? While many studies have compared care delivered by for-profit and not-for-profit hospitals, these studies have provided relatively little empirical evidence on the performance of not-for-profits and for-profits.¹ The ultimate measure of hospital performance is the impact of its care on important patient outcomes, such as death or the development of serious complications that compromise quality of life. Assessing this impact is very difficult. First, collecting reliable long-term outcome data can be challenging. Second, without comprehensive controls for differences in patient case-mix, such measures leave open the possibility that differences between hospitals reflect differences in patient disease severity and comorbidity rather than differences in quality of care. Finally, measures of important patient outcomes are notoriously noisy, due to the small numbers of patients on which they are based and the relative rarity of serious adverse outcomes for most patients. Thus, many policymakers and health care managers have expressed reservations about whether measures of serious outcomes are informative enough to identify useful differences in quality of care among hospitals.² The problem is particularly onerous for comparisons of quality of care between individual hospitals (e.g., for choosing among hospitals in a given market area).

¹ For example, see Gaumer (1986), Gray (1986), Hartz *et al.* (1989), Keeler *et al.* (1992), Staiger and Gaumer (1995).

² For example, see Ash (1996), Hofer and Hayward (1996), Luft and Romano (1993), McNeil *et al.* (1992), Park *et al.* (1990), and the cites in footnote 1.

We readdress the question of assessing hospital quality using longitudinal data sources and methods that we have recently developed (McClellan and Staiger, 1997). We discuss the data and methods below. We study important health outcomes – all-cause mortality, and major cardiac complications – for all elderly Medicare beneficiaries hospitalized with heart disease in the past decade. Our measures optimally combine information on patient outcomes from multiple years, multiple diagnoses, and multiple outcomes (e.g. death and readmission with various types of complications). As a result, we are able to develop measures that are far more accurate indicators of hospital quality than previously used in hospital outcome studies. In our previous work, we have shown that these measures far outperform previously-used methods in terms of forecasting hospital mortality rates in future years, and in terms of signal-to-noise ratios. Thus, we can expect these measures to enhance our ability to determine whether quality of care differs across hospitals.

After we introduce our data and methods, we present two sets of results. First, we examine how these new hospital quality measures vary across for-profit and not-for-profit hospitals, controlling for other characteristics of the hospital. In addition, we examine how these relationships have changed over our study period. We then examine the experience of three market areas closely: (1) a city in which a few large for-profit and not-for-profit hospitals have coexisted with stable ownership, (2) a city in which a large not-for-profit hospital was purchased by a for-profit chain, and later by another for-profit chain, and (3) a city in which the only for-profit hospital was converted to not-for-profit status.

Based on these new measures of hospital quality, our analysis uncovers a number of interesting differences between for-profit and not-for-profit hospitals. On average, we find that

for-profit hospitals have higher mortality among elderly patients with heart disease, and that this difference has grown over the last decade. However, much of the difference appears to be associated with the location of for-profit hospitals: When we compare hospital quality within specific markets, for-profit ownership appears if anything to be associated with better quality care. Moreover, the small average difference in mortality between for-profit and not-for-profit hospitals masks an enormous amount of variation in mortality within each of these ownership types. Overall, these results suggest that factors other than for-profit status *per se* may be the main determinants of quality of care in hospitals.

2. Background

Comparisons of hospital quality, and of provider quality more generally in health care and other industries, must address three crucial problems: measurement, noise, and bias.

The first problem involves measurement. Without measures of performance, there is no basis for comparing quality of care. One of the major obstacles to research on provider performance is the development of reliable data on important medical processes and health outcomes. For example, a major obstacle to comparisons of different managed-care plans today, including for-profit and not-for-profit comparisons, is that many plans simply do not have reliable mechanisms in place for collecting data on the care and outcomes their patients, especially for outpatient care. While the problem is somewhat less severe for care during an inpatient admission, many hospitals do not have reliable methods for collecting followup data on their patients, and health plans do not have mechanisms for tracking patients across hospitals. For example, until several years ago, HCFA published diagnosis-specific mortality rates for

Medicare patients. But because these outcome measures were admission-based, they could be favorably affected by hospital decisions about discharging or transferring patients, even though such actions may have no effect or adverse effects on meaningful patient outcomes. We use longitudinal data from the Medicare program linked to complete records of death dates to address the problem of collecting followup data on important outcomes for patients. But data limitations exist here as well: Medicare collects no reliable information on the care or outcomes of their rapidly-growing managed care population.

The second problem involves noise. Important health outcomes are determined by an enormous number of patient and environmental factors; differences in the quality of medical care delivered by hospitals are only one component. Moreover, most of these outcomes are relatively rare. For example, even for a common serious health problem such as heart attacks, most hospitals treat fewer than 100 cases per year, and death within a year occurs in fewer than one-fourth of these patients. Even though a one or two percentage-point difference in mortality may be very important to patients, few hospitals treat enough patients with heart disease in a year to detect such differences in outcomes. While data on other related health outcomes or on multiple years of outcomes might help reduce the noise problem, combining multiple outcome measures raises further complications. Hospital quality may improve or worsen from year to year, and the extent to which different outcomes are related to each other may not be obvious. We develop a general framework for integrating a potentially large number of outcomes over long time periods to address the noise problem. Our methods are designed to distinguish the signal of hospital quality from a potentially large number of noisy outcome measures.

The third problem involves bias. Patient selection may result in differences in outcomes across hospitals for reasons unrelated to quality. In particular, higher-quality hospitals are likely to attract more difficult cases. A range of methods, including multivariate case-mix adjustment, propensity scores, and instrumental variables, have been developed to address the selection problem. In this paper, we address the problem by focusing on an illness – heart attacks, and heart disease more generally – for which urgency limits the opportunities for selection across hospitals. A more comprehensive analysis of the selection problem is beyond the scope of this paper. In the Conclusion, we discuss some of the further evidence we have developed on the magnitude of the selection bias in our outcome measures.

In the next section, we outline our steps for addressing the measurement problems and noise problems that have complicated comparisons between for-profit and not-for-profit hospitals. Our results follow.

3. Data and Methods

Data

We use the same data as in McClellan and Staiger (1997) for this analysis. Our hospital performance measures include serious outcomes – mortality, and cardiac complications requiring rehospitalization -- for all elderly Medicare beneficiaries hospitalized with new occurrences of acute myocardial infarction (AMI, or heart attacks) from 1984 through 1994, as well as for all elderly beneficiaries hospitalized for ischemic heart disease (IHD) from 1984 through 1991. To evaluate quality of care from the standpoint of a person in the community experiencing heart disease, we assign each patient to the hospital to which she was first admitted

with that diagnosis. Our population includes over 200,000 AMI patients and over 350,000 IHD per year. We limit our analysis of hospital performance to U.S. general short term hospitals with at least two admissions in each year, a total of 3991 hospitals that collectively treated over 92 percent of these patients. In this paper, we focus exclusively on outcome differences for AMI patients, but we use information on IHD patient outcomes to help improve our estimates of hospital quality for AMI treatment.

For each AMI and IHD patient, our mortality measure is whether the patient died within 90 days of admission. In principle we could use other patient outcomes as well (e.g. death at other time periods, readmission for a cardiac complication). We focus on these two outcomes, and AMI patients in particular, for a number of reasons. First, death is an easily measured, relatively common adverse outcome for AMI, and many acute medical treatments have been shown to have a significant impact on mortality following AMI. Second, AMI cases that are not immediately fatal generally result in rapid admission to a nearby hospital, so that questions of hospital selection of patients are less of a problem for AMI. Finally, McClellan and Staiger (1997) find that measures of hospital quality based on AMI have a relatively high signal-to-noise ratio, and are strong predictors of hospital quality for other outcomes and diagnoses.³

³ In particular, McClellan and Staiger (1997) also consider performance measures for ischemic heart disease, and for a patient's quality of life following a heart attack (the occurrence of hospital readmission with congestive heart failure, ischemic heart disease symptoms, and recurrent heart attack).

For each hospital, we construct risk-adjusted mortality rates (RAMR) for each year and each diagnosis. These are the estimated hospital-specific intercepts from a patient level regression (run separately by year and by diagnosis) that estimates average all-cause mortality rates with fully-interacted controls for age, gender, black or nonblack race, and rural location. These RAMRs provide the outcome measures on which our hospital comparisons are based.

To describe hospital ownership status and other characteristics, we use data on hospital and area characteristics from the annual American Hospital Association (AHA) survey of hospitals. We use data from the 1985, 1991, and 1994 surveys in this analysis. AHA data are not available for some hospitals, limiting our final sample to 3718 hospitals.

Empirical methods

Past work comparing quality of care in hospitals has generally relied on a single hospital outcome measure in a given year. For example, to compare quality of care at two hospitals, one would simply calculate the estimated RAMR and the precision of the estimate for each hospital, and assess whether the difference in the RAMRs is statistically significant. The limitation of this approach is that the standard errors are often quite large.

Alternatively, one can combine information from all the outcome measures available for a given hospital (e.g. other years, other patients, other outcomes for the same patients) in order to more precisely estimate a hospital's current quality. This is the approach taken by McClellan and Staiger (1997). We briefly outline the method below.

Suppose we observe AMI_DTH90 and IHD_DTH90. These are noisy estimates of the true hospital intercepts that are of interest:

$$\text{AMI_DTH90}_{it} = \mu^1_{it} + \varepsilon^1_{it}$$

$$\text{IHD_DTH90}_{it} = \mu^2_{it} + \varepsilon^2_{it}$$

Where μ is the true parameter of interest (the hospital-specific intercept in the 90-day mortality equations), ε is the estimation error, and we observe each outcome for T years. Note that $\text{Var}(\varepsilon^1_{it} \varepsilon^2_{it})$ can be estimated, since this is simply the variance of regression estimates.

Let $M_i \equiv \{\text{AMI_DTH90}_i, \text{IHD_DTH90}_i\}$ be a $1 \times (2T)$ vector of the T years of data on each outcome, and let $\mu_i \equiv \{\mu^1_i, \mu^2_i\}$ be a $1 \times (2T)$ vector of the true hospital intercepts. Our problem is how to use M_i to predict μ_i . More specifically, we wish to create a linear combination of each hospital's observed outcomes data in such a way that it minimizes the mean square error of our predictions. In other words, we would like to run the following hypothetical regression:

$$(1) \quad \mu_{it} = \{\text{AMI_DTH90}_i, \text{IHD_DTH90}_i\} \beta_{it} + \nu_{it} \equiv M_i \beta_{it} + \nu_{it}$$

but cannot, since μ_i is unobserved and β will vary by hospital and time.

Equation (1) helps to highlight the problem with using a single years RAMR as a prediction of the true hospital level intercept. Since the RAMR is estimated with error, we can improve the mean squared error of the prediction by attenuating the coefficient towards zero, and this attenuation should be greater for hospitals in which the RAMR is not precisely estimated. Moreover, if the true hospital specific intercepts from other outcomes equations (e.g. other years, other patients) are correlated with the intercept we are trying to predict, then using their estimated values can further improve prediction ability.

McClellan and Staiger (1997) develop a simple method for creating estimates of μ_i based on equation (1). The key to the solution is noting that to estimate this hypothetical

regression (e.g. get coefficients, predicted values, R-squared) we only need three moment matrices:

$$(i) \quad E(M_i'M_i) = E(\mu_i'\mu_i) + E(\varepsilon_i'\varepsilon_i)$$

$$(ii) \quad E(M_i'\mu_i) = E(\mu_i'\mu_i)$$

$$(iii) \quad E(\mu_i'\mu_i)$$

We *can* estimate the required moment matrices directly as follows:

1. We can estimate $E(\varepsilon_i'\varepsilon_i)$ with the patient-level OLS estimate of the variance-covariance for the parameter estimates M_{it} . Call this estimate S_i .
2. We can estimate $E(\mu_i'\mu_i)$ by noting that: $E(M_i'M_i - S_i) = E(\mu_i'\mu_i)$. If we assume that $E(\mu_i'\mu_i)$ is the same for all hospitals, then it can be estimated by the sample average of $M_i'M_i - S_i$.

Finally, it helps to impose some structure on $E(\mu_i'\mu_i)$ for two reasons. First, this improves the precision of the estimated moments by limiting the number of parameters that need to be estimated. Second, a time series structure allows for out-of-sample forecasts. Thus, we assume a non-stationary first-order Vector Autoregression structure (VAR) for μ_{it} (1x2). This VAR structure implies that $E(\mu_i'\mu_i)=f(\Gamma)$, where Γ are the parameters of the VAR. These parameters can be estimated by GMM, i.e. by setting the theoretical moment matrix, $f(\Gamma)$, as close as possible to its sample analog, the sample average of $M_i'M_i - S_i$. For details, see McClellan and Staiger (1997).

With estimates of $E(\mu_i'\mu_i)$ and $E(\varepsilon_i'\varepsilon_i)$, we can form estimates of the moments (i-iii) needed to run the hypothetical regression in equation (1). By analogy to simple regression, our predictions of a hospitals true intercept are given by:

$$(2) \hat{\boldsymbol{\mu}} = \mathbf{M}_i \mathbf{E}(\mathbf{M}_i' \mathbf{M}_i)^{-1} \mathbf{E}(\mathbf{M}_i' \boldsymbol{\mu}_i) = \mathbf{M}_i [\mathbf{E}(\boldsymbol{\mu}_i' \boldsymbol{\mu}_i) + \mathbf{E}(\boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i)]^{-1} \mathbf{E}(\boldsymbol{\mu}_i' \boldsymbol{\mu}_i)$$

where we use our estimates of $\mathbf{E}(\boldsymbol{\mu}_i' \boldsymbol{\mu}_i)$ and $\mathbf{E}(\boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i)$ in place of their true values. We refer to estimates based on equation (2) as “filtered RAMR” estimates, since these estimates are attempting to filter out the estimation error in the raw data (and because our method is closely related to the idea of filtering in time series).

4. National Estimates

One common method of comparing quality of care across hospitals is to run cross-section regressions using a quality measure such as RAMR as the dependent variable and using hospital characteristics such as patient volume, ownership and teaching status as independent variables. In this section we investigate the extent to which using a filtered RAMR as the dependent variable affects the inferences that can be drawn from such regressions. *A priori*, we would expect that using the filtered RAMR (as opposed to the actual RAMR in a given year) would improve the precision of such regression estimates because the dependent variable is measured with less noise. The gain in efficiency is likely to be particularly large for smaller hospitals, since the RAMR estimates in any single year for these hospitals have the lowest signal-to-noise ratio.

Figure 1 illustrates this difference between filtered and actual RAMR by plotting each against volume using data from 1991. Throughout the remainder of the paper we focus on RAMRs based on 90-day mortality among Medicare AMI admissions (although the filtered estimates incorporate the information from 90-day mortality among IHD admissions as well). Keep in mind that the unit for the RAMR measures is the probability of death, so that a RAMR

of 0.1 means that the hospital had a mortality rate that was ten percentage points higher than expected (e.g. 30% rather than 20%).

There are two interesting features of Figure 1. First, the filtered RAMR estimates have much less variance than the actual RAMR estimates, particularly for smaller hospitals. This is the result of two distinct effects. Most importantly, the filtered estimates for small hospitals are relying more heavily on data from other years and other diagnoses, and this improves their precision. In addition, the filtered estimates assume the actual RAMR estimates for small hospitals have a very low signal-to-noise ratio, and therefore attenuate them back towards the average (similar to shrinkage estimators).

A second interesting feature of Figure 1 is that the relationship between outcomes and volume is much more apparent in the filtered data. High volume hospitals clearly seem to have lower mortality. Thus, these filtered RAMRs appear to be a useful tool for uncovering quality differences across hospitals.

Table 1 provides regression estimates that further suggest that these filtered RAMR estimates improve our ability to uncover differences in quality across hospitals. This table contains coefficient estimates from regressions of RAMR estimates (either actual or filtered) on dummies for ownership (for-profit and government, with not-for-profit the reference group), a dummy for being a teaching hospital, and the number of Medicare AMI admissions in the given year (in 100s). Since volume is potentially endogenous (and since Medicare volume is a crude proxy for total volume) we also report estimates from regressions that do not control for volume. The table contains estimates for 1985, 1991, and 1994. The regressions using actual RAMR are weighted by the number of Medicare admissions, while the regressions using filtered

RAMR are weighted by the inverse of the estimated variance of each hospital's filtered RAMR estimate.

As one would expect, the regressions based on the filtered RAMR yield much more precise coefficient estimates. The standard errors in regressions using the actual RAMR are 2-3 times larger than the corresponding standard errors from regressions using the filtered RAMR. For example, using actual RAMR in 1985, mortality in for-profit hospitals is estimated to be 0.16 percentage points higher than in not-for-profit hospitals. But the standard error for this estimate is so large (0.43 percentage points) that the difference would have to be near a full percentage point before we could be confident of a real difference in mortality. In contrast, using filtered RAMR in 1985, mortality in for-profit hospitals is estimated to be 0.30 percentage points higher than in not-for-profit hospitals and this difference is borderline significant because of the much smaller standard error.

More generally, the coefficients in the regressions using filtered RAMR are precise enough to uncover a number of interesting facts. For-profit hospitals have higher mortality than do not-for-profits (by 0.30 to 1.15 percentage points depending on the year and specification). Government hospitals have higher mortality and teaching hospitals lower mortality than do not-for-profit hospitals. These differences are larger in specifications that do not control for volume, because (1) government and for-profit hospitals tend to be smaller than average, while teaching hospitals tend to be larger than average and (2) there is a strong negative relationship between volume and mortality. For example, in 1985 we estimate that an additional 100 Medicare AMI admissions was associated with 1.5 percentage points lower mortality.

The most striking finding in Table 1 is the apparent change in the coefficients between 1985 and 1994. In the specifications using the filtered RAMR, the coefficient estimates for for-profit and teaching hospitals rise by roughly half of a percentage point in absolute value between 1985 and 1994. At the same time, the coefficient on volume fell in absolute value by about half a percentage point.

These regression estimates suggest that the filtered RAMR can be a useful tool for uncovering general relationships between mortality and hospital characteristics. Based on the filtered data, three facts are clear in the data: (1) there is a negative relationship between volume and mortality, (2) for-profit hospitals and government hospitals have higher mortality than not-for-profit hospitals, while teaching hospitals have lower mortality, and (3) between 1985 and 1994, mortality differences increased between for-profit and not-for-profit hospitals, and between teaching and non-teaching hospitals.

These findings are generally consistent with the existing literature, although our estimates tend to be more precise. Studies examining a variety of patient populations and outcomes measures have found that higher volume is associated with better patient outcomes.⁴ Comparisons by ownership and teaching status, to the extent they have found any differences, have found not-for-profit and teaching hospitals to have better patient outcomes.⁵ The most novel of our findings is that these differences have widened over the last decade. This decade has been a period of rapid change in hospitals, spurred by dramatic changes in the way that

⁴ See Luft *et al.* (1990) for a fairly comprehensive study of the volume-outcome relationship.

⁵ See the cites in footnote 1 and 2.

both government and private insurers pay for hospital care. The extent to which these market changes might explain the growing differences in hospital mortality is an important area for future research.

5. A Tale of Three Counties

The Sample

If the filtered RAMR helps to compare hospitals at the national level, can it also help at a more micro level? One important use for any measure of hospital quality is to compare individual hospitals within a given market. In this section we look more closely at the mortality performance of particular hospitals in three counties. Our goals are: (1) to learn whether these quality measures are able to identify meaningful differences (and changes over time) in mortality among hospitals in a given city; and (2) to explore whether these patterns in mortality could be attributed to for-profit ownership or other factors affecting the market. At the same time, by going to the county level and focusing on a fixed group of hospitals, we are able to address some of the general results discussed in section 4 from a “case study” perspective.

The three counties were chosen on the following basis. First, since we wanted to compare individual hospitals (but not too many hospitals) we limited our search to counties with 2-10 hospitals in our sample. In order to focus on for-profit hospitals, the county had to have at least one for-profit hospital and one other hospital with an average of at least 50 Medicare AMI admissions per year from 1984-1994. Within this subset we considered three categories of counties:

1. CASE 1: no change in for-profit ownership over the study period.

2. CASE 2: at least one hospital converted in to for-profit over the study period.
3. CASE 3: at least one hospital converted away from for-profit over the study period.

Within each category we eliminated counties that were obviously not distinct markets, e.g. the suburbs of Miami. Finally, we chose the county that had the highest average volume in its primary hospitals.

The resulting counties all contain relatively isolated midsize cities. To preserve the confidentiality of individual hospitals, we refer to each hospital according to its rank in terms of AMI volume between 1984 and 1994.

Case 1 contains a small southern city with four larger than average hospitals. The largest (hospital 1) and smallest (hospital 4) are for-profit hospitals, both affiliated with the same for-profit chain. Hospital 2 is government run, while hospital 3 is a not-for-profit. Relative to the other two cases, this city had experienced rapid growth in population and income during the 1980s and has a high number of hospital beds per capita. The population is somewhat older, less educated and less likely to be white, with 10-20% enrolled in HMOs by 1994.⁶

Case 2 contains a mid-sized midwestern city with three larger than average hospitals and one very small hospital (hospital 4). Hospitals 1, 3 and 4 are not-for-profit. Hospital 2 was a not-for-profit until the mid 1980s, at which time it was purchased by a large for-profit chain. The ownership of hospital 2 was transferred to a different for-profit chain in the early

⁶ Information on each city/county comes from the County and City Data Book for 1988 and 1994. Information on HMO penetration in each county was provided by Laurence Baker, based on his calculations using HMO enrollment data from InterStudy.

1990s. Relative to the other two cases, this city had average growth in population and income during the 1980s and has a low number of hospital beds per capita. The population has higher income, is somewhat younger, more educated, and more likely to be white, with 10-20% enrolled in HMOs by 1994.

Case 3 contains a mid-sized southern city with 5 larger than average hospitals.

Hospitals 1, 2 and 4 are not-for-profit. Hospital 3 was initially government owned and hospital 5 was initially for-profit. Both hospital 3 and 5 converted to non-profit status in the late 1980s.

Relative to the other two cases, this city had low population growth during the 1980s.

Otherwise, this city has fairly average population characteristics with 10-20% enrolled in HMOs by 1994.

Evidence on Quality in Each County

In keeping with the exploratory nature of this analysis, Figure 2 simply plots the RAMR (left panel) and filtered RAMR (right panel) annually from 1984-1994 for each hospital in Case 1. Note that the vertical scale differs between the two plots (in order to preserve the detail of the filtered RAMR plot). Figure 2 plots this data slightly differently. Each panel corresponds to a hospital, and plots the actual RAMR along with the filtered RAMR and its 90% confidence band. Confidence bands for the actual RAMR are too large to fit on the figure. A horizontal line denoting the RAMR at the average hospital in our sample is added to each panel for reference. The data for Case 2 are similarly plotted in figures 4 and 5, and for Case 3 in figures 6 and 7.

For Case 1, it is impossible to detect quality differences across the hospitals or over time based on the actual RAMR (see the left panel of Figure 2). Obviously, the problem is the variability in the actual RAMR: Even the largest hospital (#1) experiences year-to-year changes in its actual RAMR of over five percentage points.

In contrast, the filtered RAMR is much more stable and displays three interesting features. First, the for-profit hospitals (1 and 4) have, if anything, lower mortality than the other hospitals in the market. The fact that the smallest hospital also has the lowest filtered RAMR seems surprising, but may be the result of its affiliation with hospital 1 (recall that they are members of the same chain). A second interesting feature of the filtered data in figure 2 is that every hospital appears to experience an improvement in mortality of about 1-2 percentage points in the mid 1980s relative to other hospitals nationally. Although it is beyond the scope of this paper, an interesting topic for further research is the analysis of the cause of this general improvement in quality of care in this area.⁷ Finally, it is notable that the range of filtered RAMR estimates, while much larger than the differences estimated between the average for-profit and not-for-profit in Table 1, are still relatively compressed. Based on national data, McClellan and Staiger (1997) estimate that the standard deviation across hospitals is around 4 percentage points for the true hospital-specific intercepts for 90-day mortality.

⁷ Recall that the RAMR measures mortality relative to the average hospital, so this improvement does not simply reflect the downward national trend in heart attack mortality rates. Mortality in these hospitals improved relative to the national average over this time.

Figure 3 Plots each hospital's data separately, and adds 90% confidence bands to the filtered RAMR (thick line with vertical bars). The horizontal line at RAMR=0 represents the national average in that year, so when the confidence bands lie entirely below or above this line it is likely that the hospital is, respectively, better or worse than average. Relative to the size of the confidence bands, there are not large differences either across these hospitals or over time. Hospital 1 (the large for-profit) is the only hospital that is consistently better than the national average, and this seems to be consistent with its general status in the community.

Thus, the overall picture for Case 1 seems to be one of fairly homogeneous quality, perhaps slightly above the national average. There are hints of improvement over time, and of better quality in the for-profit hospitals, but no dramatic differences.

As Figure 4 illustrates, Case 2 is quite different. The only similarity is that it is impossible to detect quality differences across hospitals or over time based on the actual RAMR data plotted in the left panel of the figure. Using the filtered RAMR, there is a clear ranking of quality across hospitals that roughly corresponds to size. The largest hospital (a not-for-profit) consistently has the lowest mortality, while hospital 4 (a very small hospital) has the highest mortality. The difference in mortality between the largest and smallest hospital is substantial, from 6 to over 8 percentage points. These differences are large even relative to the 90% confidence bounds for the filtered RAMR (see figure 5). As in Case 1, the hospital which we identify as having the lowest mortality is recognized in the community as the leading hospital.

Hospital 2 is of particular interest, because it was taken over by a for-profit chain in the mid 1980s and then became part of a different for-profit chain in the early 1990s. Around both of these ownership changes, there is a notable decline in the hospital's filtered RAMR of about

2 percentage points. In fact, it is the only hospital in Case 2 that has an apparent trend (downward) in its mortality, going from being worse than average to better than average. While it is not clear that the change in ownership *per se* led to these improvements, it is at least suggestive that this may be the case.

The overall picture for Case 2 seems to be one of more diversity of quality, although fairly average quality overall. The purchase of a hospital first by one and then another for-profit chain seemed, if anything, to improve quality. However, the purchased hospital is still not clearly any better than the national average in terms of mortality.

Case 3 presents yet another situation. Again, there is a wide range of quality across hospitals in this area, with the range in filtered RAMR of 5-8 percentage points (see Figure 6). There is a clear downward trend in mortality occurring in this area, which is even seen in the actual RAMR (although the actual RAMR is still very noisy). Using the filtered RAMR, each of the hospitals in this area experienced a decline in mortality of between 2 and 8 percentage points. Hospital 1, the largest not-for-profit, had the lowest mortality throughout almost the entire period. Hospital 3, which converted from government to not-for-profit in the late 1980s, clearly had the highest mortality initially but also experienced one of the largest declines by 1994. Hospital 5, which converted from for-profit to not-for-profit in the late 1980s, had the largest mortality decline of all five hospitals to the point where it had the lowest filtered RAMR in the area in 1994.

Thus, the overall picture for Case 3 is one of rapidly improving quality in the area as a whole. At the same time, the for-profit and government hospitals converted to not-for-profit and had the most dramatic quality improvements in the area.

There are two common themes across all of these cases. First, filtered RAMRs appear to be a useful tool for analyzing quality of care differences across hospitals and over time. More importantly, our micro level evidence from these specific cases is not consistent with the common belief (supported by our aggregate regressions) that for-profit hospitals provide lower quality of care. In two of our three markets, for-profits appeared to be associated with higher quality of care: hospitals that were for-profit throughout our study period tended to have lower mortality rates, and changes to for-profit status were associated with mortality reductions.

What might explain this apparent conflict between the case-study evidence and the aggregate cross-section evidence, which showed a poorer performance overall for the for-profits? Some of the explanation may come from the way in which we chose our case studies, relying on areas with relatively large for-profit hospitals that were perhaps likely to represent “flagship” hospitals in their communities. These features may not be representative of the market status of a typical for-profit hospital.

One possible explanation for these results could be that for-profit hospitals selectively locate in areas with low quality (see e.g. Norton and Staiger, 1994). Thus, the aggregate evidence would tend to find that for-profit ownership was correlated with lower quality, while within their markets the for-profit hospitals could provide higher quality (as in Case 1), or at least improve quality in the hospitals they acquire (as in Case 2). This explanation would also imply that for-profit hospitals would tend to leave markets in which the quality was rising (as in Case 3). If the cross-section correlation is being generated by location, then we would expect within county differences between for-profit and not-for-profit hospitals to be smaller than across county differences. In fact, when we include county-level fixed-effects in the regressions

from Table 1, the estimated mortality difference between for-profit and not-for-profit hospitals falls by roughly half. Thus, it appears that at least some of the difference in quality is generated by the different location patterns of for-profit hospitals.

Why might for-profit hospitals tend to locate in areas with low hospital quality? One possible reason would be a relationship between poor hospital management and lower quality of care. Poorly-managed hospitals might make attractive takeover targets for for-profit chains, but as a byproduct the for-profits would tend to enter markets with low quality of care.

Alternatively, patients may demand high quality care in some markets, either because of demographic factors such as high income or because of an existing high-quality hospital in the market (e.g. a teaching hospital). If providing such high quality care results in lower patient margins, then for-profits would be less likely to locate in these areas.

These speculative explanations are based on the results of only a few market case studies. We will leave a more systematic exploration of this question to future work. Clearly, however, a final important conclusion of this research is that the “average” differences in mortality between for-profit and not-for-profit hospitals, or among any other general system for classifying hospitals such as bed size, account for only a small share of the variation in outcomes across hospitals. Many not-for-profit hospitals are below average, many for-profit hospitals are above average, and these relationships vary enormously at the market level. More extensive market-level analyses using the methods we have developed to evaluate quality could yield new insights into these complex relationships.

6. Conclusion

In this paper, we have summarized new methods for evaluating the quality of care of for-profit and not-for-profit hospitals. These methods address two of the major problems that have limited the value of previous hospital quality assessments: measurement of important outcomes, and the high level of noise in these measures. In McClellan and Staiger (1997), where we describe these techniques in more detail, we also present evidence on a third major problem: bias in the hospital comparisons because of unmeasured differences in “case mix” across hospitals. We use detailed medical chart review data to show that hospital performance measures for heart attack care which account for patient disease severity and comorbidity in a much more extensive way are highly correlated with the measures we report in this research. In other words, our measures with limited case-mix adjustment provide reasonably good predictions of hospital performance in terms of measures based on detailed case-mix adjustment. Our results to date on the bias problem are by no means conclusive; hard-to-measure patient factors may differ systematically across hospitals, particularly for less acute conditions than heart attacks. At a minimum, however, by providing relatively precise measures of hospital performance for important dimensions of hospital quality of care, our approach allows further research to focus on this final key problem.

The results of our analysis provide a range of new insights for policy issues related to for-profit and not-for-profit hospital ownership. On average, the performance of not-for-profit hospitals in treating elderly patients with heart disease appears to be slightly better than that of for-profit hospitals, even after accounting for systematic differences in hospital size, teaching status, urbanicity, and patient demographic characteristics. This average difference in mortality performance between for-profits and not-for-profits appears to be increasing over time.

However, this small average difference masks an enormous amount of variation in hospital quality within the for-profit and not-for-profit hospital groups. Our case study results also suggest that for-profits may provide the impetus for quality improvements in markets where, for various reasons, relatively poor quality of care is the norm. Understanding the many market- and hospital-specific factors that contribute to these variations in hospital quality is a crucial topic for further research. Using the methods and results developed here, such detailed market analyses can be based on rather precise assessments of differences in hospital performance, rather than on speculation necessitated by imprecise or absent outcome measures.

References

Ash, A., "Identifying Poor-Quality Hospitals with Mortality Rates

Medical Care, 34:735-736, 1996

County and City Data Book, Washington DC: U.S. Department of Commerce, Bureau of the Census, 1988 and 1994 editions.

Gaumer, G., "Medicare Patient Outcomes and Hospital Organizational Mission," in B.H. Gray (ed) *For-Profit Enterprise in Health Care*, Washington DC: National Academy Press, 354-384, 1986

Gray, B.H., *For-Profit Enterprise in Health Care*, Washington DC: National Academy Press, 1986.

Hartz, A.J., H. Krakauer, E.M. Kuhn, *et al.*, "Hospital Characteristics and Mortality Rates," *New England Journal of Medicine*, 321:1720-1725, 1989.

Hofer, T.P. and R.A. Hayward, "Identifying Poor-Quality Hospitals: Can Hospital Mortality Rates Detect Quality Problems for Medical Diagnoses?" *Medical Care*, 34:737-753, 1996

Keeler, E.B., L.V. Rubenstein, K.L. Kahn, *et al.*, “Hospital Characteristics and Quality of
JAMA, 268:1709-1714, 1992.

Luft, H.F., D.W. Garnick, D.H. Mark and S.J. McPhee, *Hospital Volume, Physician
Volume, and Patient Outcomes*, Ann Arbor, Mich.: Health Administration Press Perspectives,
1990.

Luft, H.F. and P.S. Romano, “Chance, Continuity, and Change in Hospital Mortality Rates,”
JAMA, 270:331-337, 1993.

McClellan, M. and D. Staiger, “The Quality of Health Care Providers,” working paper,
December 1997.

McNeil, B.J., S.H. Pedersen and C. Gatsonis, “Current Issues in Profiling Quality of Care,”
Inquiry, 29:298-307, 1992.

Norton, E.C. and D. Staiger, “How Hospital Ownership Affects Access to Care for the
RAND Journal of Economics, 25:171-185, 1994.

Park, R.E., R.H. Brook, J. Kosecoff, *et al.*, “Explaining Variations in Hospital Death Rates,”
JAMA, 264:484-490, 1990.

Staiger, D. and G. Gaumer, "Price Regulation and Patient Mortality in Hospitals," working paper, August 1995.

</ref_section>

Table 1
Regression estimates of the relationship between hospital characteristics
and the Risk-Adjusted Mortality Rate (RAMR)
based on 90-day mortality for AMI admits
(3718 hospitals)

| | 1985 | | | 1991 | | | 1994 | | |
|--------------------------------|---------------------|--------------------------|---------------------|---------------------|--------------------------|---------------------|---------------------|--------------------------|---------------------|
| | Actual RAMR | Filtered Version of RAMR | | Actual RAMR | Filtered Version of RAMR | | Actual RAMR | Filtered Version of RAMR | |
| #Medicare Admits in AMI (100s) | -0.0178 (0.0022) | -0.0153 (0.0010) | | -0.0148 (0.0018) | -0.0143 (0.0009) | | -0.0093 (0.0014) | -0.0110 (0.0007) | |
| Government | 0.0151 (0.0033) | 0.0104 (0.0012) | 0.0148 (0.0012) | 0.0219 (0.0033) | 0.0120 (0.0013) | 0.0178 (0.0013) | 0.0169 (0.0033) | 0.0109 (0.0012) | 0.0156 (0.0013) |
| For-Profit | 0.0016 (0.0043) | 0.0030 (0.0016) | 0.0061 (0.0016) | 0.0115 (0.0038) | 0.0071 (0.0016) | 0.0100 (0.0016) | 0.0102 (0.0038) | 0.0087 (0.0015) | 0.0115 (0.0015) |
| Teaching | -0.0031 (0.0033) | 0.0022 (0.0014) | -0.0066 (0.0014) | -0.0047 (0.0030) | -0.0039 (0.0014) | -0.0111 (0.0014) | -0.0083 (0.0028) | -0.0047 (0.0013) | -0.0102 (0.0013) |

Standard errors given in parentheses.

Regressions using the actual RAMR weight by the number of AMI admits.

Regressions using the filtered RAMR weight by $1/\sigma^2$, where σ is the standard error of the estimated RAMR

Table 1
Regression estimates of the relationship between hospital characteristics
and the Risk-Adjusted Mortality Rate (RAMR)
based on 90-day mortality for AMI admits
(3718 hospitals)

| | 1985 | | | 1991 | | | 1994 | | |
|--------------------------------|---------------------|--------------------------|---------------------|---------------------|--------------------------|---------------------|---------------------|--------------------------|---------------------|
| | Actual RAMR | Filtered Version of RAMR | | Actual RAMR | Filtered Version of RAMR | | Actual RAMR | Filtered Version of RAMR | |
| #Medicare Admits in AMI (100s) | -0.0178 (0.0022) | -0.0153 (0.0010) | | -0.0148 (0.0018) | -0.0143 (0.0009) | | -0.0093 (0.0014) | -0.0110 (0.0007) | |
| Government | 0.0151 (0.0033) | 0.0104 (0.0012) | 0.0148 (0.0012) | 0.0219 (0.0033) | 0.0120 (0.0013) | 0.0178 (0.0013) | 0.0169 (0.0033) | 0.0109 (0.0012) | 0.0156 (0.0013) |
| For-Profit | 0.0016 (0.0043) | 0.0030 (0.0016) | 0.0061 (0.0016) | 0.0115 (0.0038) | 0.0071 (0.0016) | 0.0100 (0.0016) | 0.0102 (0.0038) | 0.0087 (0.0015) | 0.0115 (0.0015) |
| Teaching | -0.0031 (0.0033) | 0.0022 (0.0014) | -0.0066 (0.0014) | -0.0047 (0.0030) | -0.0039 (0.0014) | -0.0111 (0.0014) | -0.0083 (0.0028) | -0.0047 (0.0013) | -0.0102 (0.0013) |

Standard errors given in parentheses.

Regressions using the actual RAMR weight by the number of AMI admits.

Regressions using the filtered RAMR weight by $1/\sigma^2$, where σ is the standard error of the estimated RAMR

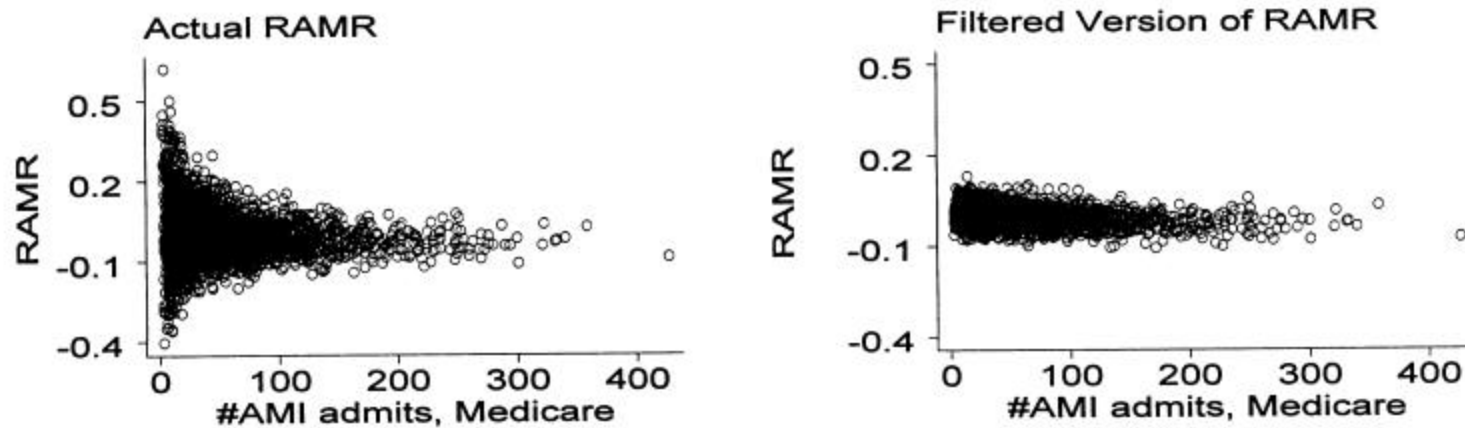


Figure 1. The relationship between Risk Adjusted Mortality Rates (RAMR) and patient volume using actual versus filtered RAMR. Based on 90-day mortality for Medicare AMI admits.

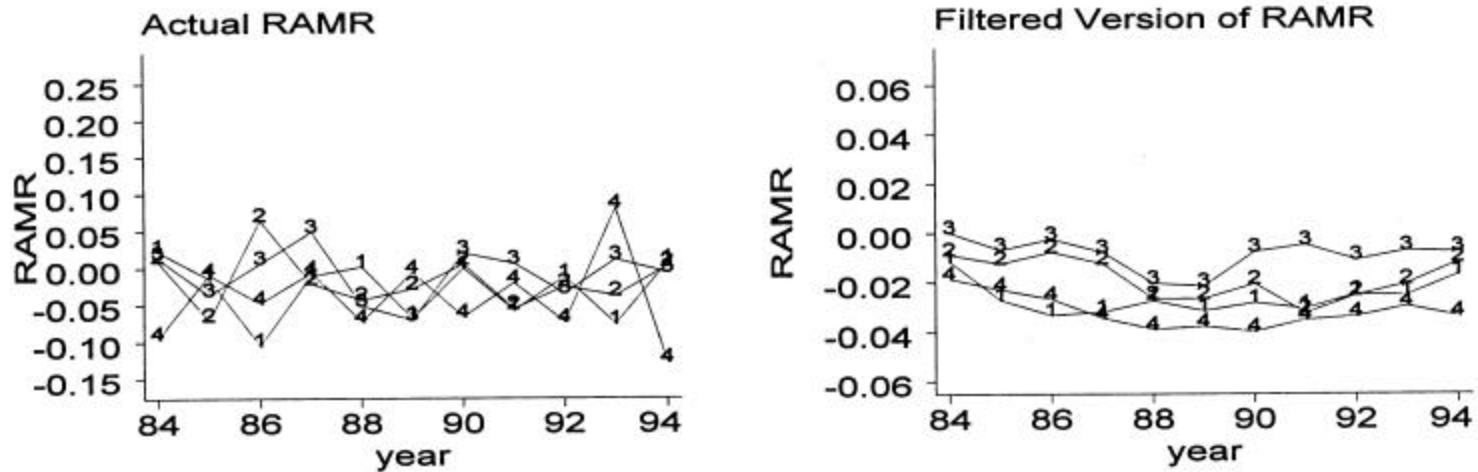


Figure 2. Trends in Risk Adjusted Mortality Rates (RAMR) for Case 1 (a midsize southern city). Left panel based on Actual RAMR and right panel based on filtered RAMR (Note that the horizontal scale of the two panels differs). The hospitals are ranked from largest (1) to smallest (4) according to their number of Medicare AMI admissions from 1984 to 1994. Hospitals 1 and 4 are for-profit hospitals and are affiliated with the same chain. Hospital 2 is government owned, while hospital 3 is not-for-profit.

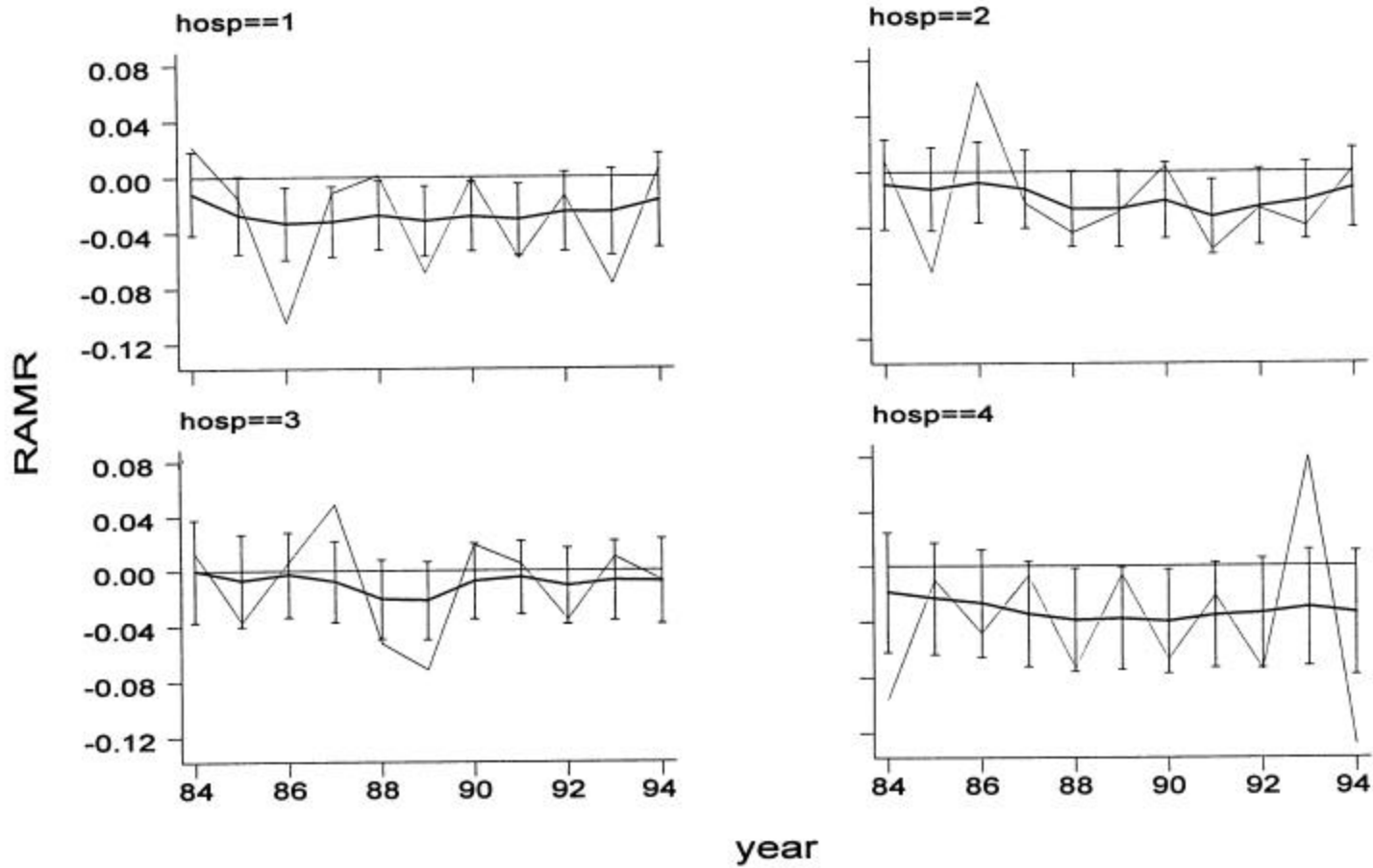


Figure 3. Trends for Case 1 in Actual (thin line) and filtered (thick line with 90% confidence bands) RAMR by hospital. The straight horizontal line denotes the RAMR at the average hospital in our national sample (RAMR=0 by definition). For description of the hospitals see Figure 2.

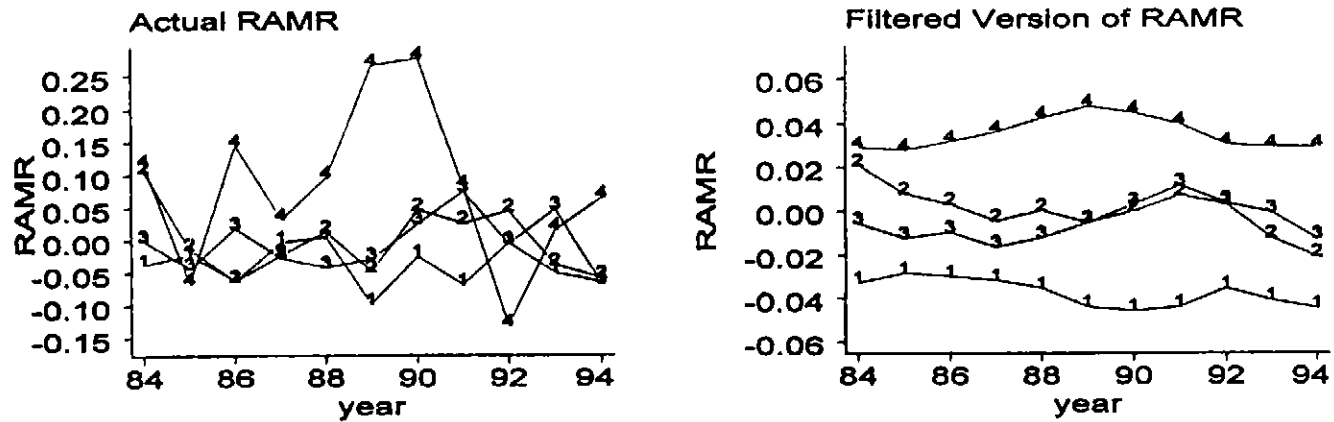


Figure 4. Trends in Risk Adjusted Mortality Rates (RAMR) for Case 2 (a midsize midwestern city). Left panel based on Actual RAMR and right panel based on filtered RAMR (Note that the horizontal scale of the two panels differs). The hospitals are ranked from largest (1) to smallest (4) according to their number of Medicare AMI admissions from 1984 to 1994. Hospital 4 is quite small. Hospitals 1, 3 and 4 are not-for-profit hospitals. Hospital 2 was a not-for-profit that was purchased by a for-profit chain in 1985, and then by a different for-profit chain in 1993.

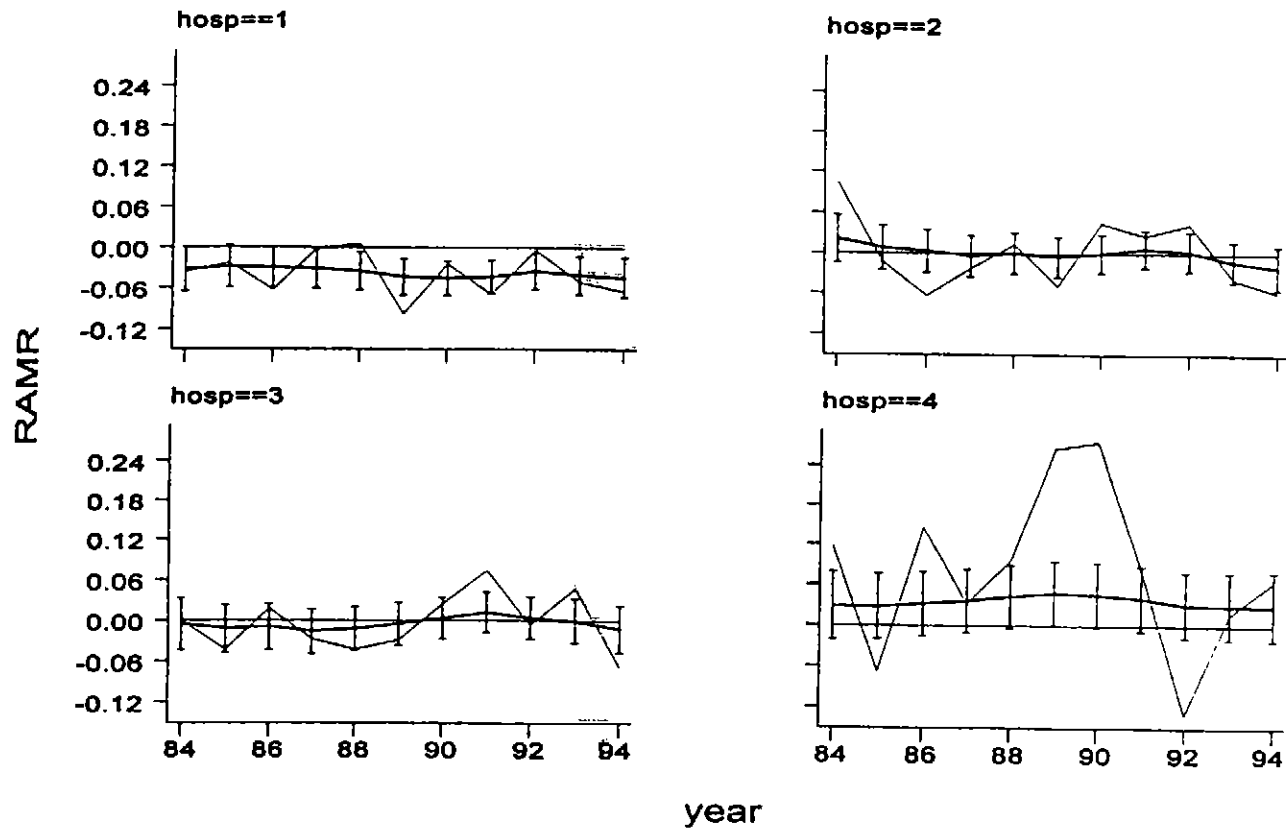


Figure 5. Trends for Case 2 in Actual (thin line) and filtered (thick line with 90% confidence bands) RAMR by hospital. The straight horizontal line denotes the RAMR at the average hospital in our national sample (RAMR=0 by definition). For description of the hospitals see Figure 4.

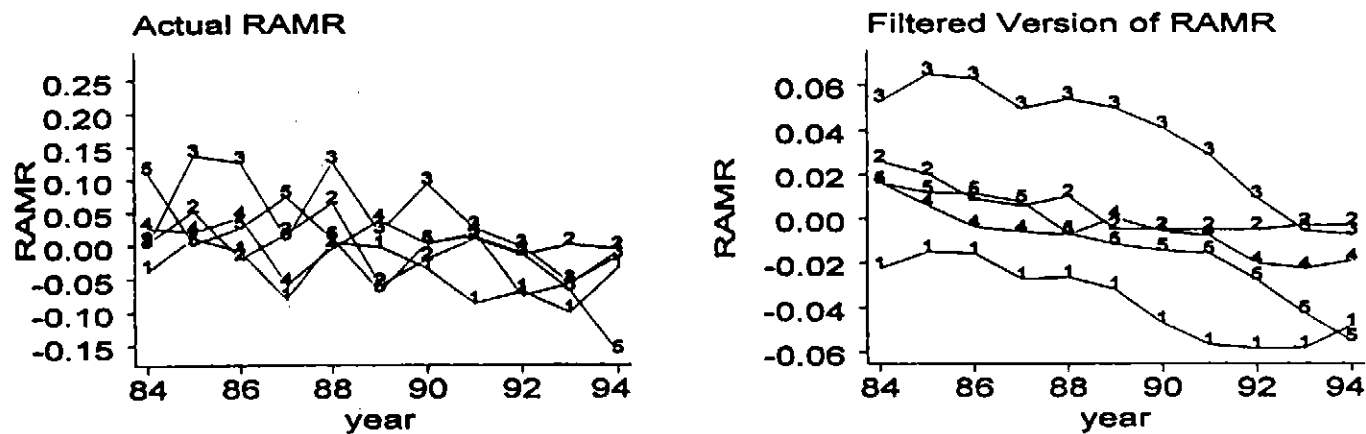


Figure 6. Trends in Risk Adjusted Mortality Rates (RAMR) for Case 3 (a midsize southern city). Left panel based on Actual RAMR and right panel based on filtered RAMR (Note that the horizontal scale of the two panels differs). The hospitals are ranked from largest (1) to smallest (5) according to their number of Medicare AMI admissions from 1984 to 1994. Hospitals 1, 2 and 4 are not-for-profit hospitals. Hospital 3 was initially government owned and then converted to non-profit status in the late 1980s. Similarly, hospital 5 converted from for-profit to not-for-profit status in the late 1980s.

