Author's response to reviews

Title: Payer leverage and hospital compliance with a benchmark: A population-based observational study

Authors:

John M. Hollingsworth (kinks@med.umich.edu)
Sarah L. Krein (skrein@umich.edu)
David C. Miller (dcmiller@med.umich.edu)
Sonya DeMonner (smdemo@med.umich.edu)
Brent K. Hollenbeck (bhollen@med.umich.edu)

Version: 2 Date: 2 May 2007

Author's response to reviews: see over
May 2, 2007

MS: 1130131151210986

Dear Dr. Saltman:

Thank you for allowing us the opportunity address the reviewers’ comments on our manuscript, entitled “Payer leverage and hospital compliance with a benchmark: A population-based observational study” coauthored by Drs. John Hollingsworth et al. We appreciate the comments of the reviewers and have made the following changes itemized below. Changes to our manuscript are italicized and in bold-face font.

Reviewer 1:

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. I have one major concern regarding the analysis. To measure the association between hospital volume compliance and Medicare market share, the authors estimate a logistic regression (p. 8), which includes Medicare market share and donor organ supply as explanatory variables. They interpret the large and precisely estimated coefficient on the odds ratios for Medicare market share for most procedures as indicating that caseloads with a high proportion of Medicare patients will lead hospitals to work harder to meet the minimum caseload requirements. However, one could also conclude that hospitals with a (high) proportion of Medicare patients merely reflects a hospital’s greater access to a patient population that has (a) high need for transplantation. The authors address this issue by comparing kidney transplantation (which they argue has greater financial incentives) to other transplant procedures. This approach is somewhat convincing, although these procedures also have minimum caseload requirements specified by Medicare.

A more convincing analysis would be to also include (an) analysis of one or two procedures where Medicare does not specify minimum caseload requirements. I don’t know if this is the case, but one could try using angioplasty or open heart surgery, where the American Heart Association has specified minimum requirements, or the Whipple procedure or esophagectomy where the Leapfrog Group has specified minimum volume standards. If the authors’ hypothesis is true, then the odds ratios in similar regression for these procedures would be much smaller in magnitude.

The reviewer recommends a secondary analysis of a procedure for which Medicare has no minimum caseloads specified. However, examining procedures such as percutaneous coronary intervention (the American Heart Association, a professional society, has minimum caseload requirements) and esophagectomy/pancreatectomy (the Leapfrog Group, a consortium of employers, has case volume guidelines) is problematic since neither group is a payer and may have no lever with providers. Indeed, empirical data
suggest that the latter’s guidelines have gained little traction (Galvin RS et al: Has the leapfrog group had an impact on the health care market? Health Aff 2005, 24:228-233.).

However, within our study population, we were able to address the reviewer’s concerns. Medicare’s coverage announcement for lung transplants was made in 1995 enabling us to examine the relationship between compliance and Medicare market share in the absence of a minimum threshold (years 1993 and 1994). For the years 1993 to 1994, during which no lung transplant minimum caseloads were specified, the OR for volume compliance was 1.38 (95% CI, 0.15 – 12.66). Following Medicare’s coverage announcement, the OR rose to 3.73 (95% CI, 1.70 – 8.19). These findings support our interpretation of the association between hospital volume compliance and Medicare market share. The description of this secondary analysis and its results are now described in our manuscript.

2. The manuscript would also benefit from a clearer discussion upfront of exactly what leverage that minimum compliance standards imposed by CMS actually have on hospitals. There is a brief description at the very end of the manuscript recognizing that the volume standards are not strictly enforced. This should be described clearly up front. I was confused while reading most of the manuscript because I could not understand why a non-compliant hospital had any Medicare transplant patients at all.

As recommended by the reviewer, this issue is now discussed in the second paragraph of our background section. Here we state, “Even prior to the recent media attention, concerns have been raised regarding the appropriateness of Medicare’s standards and a possible lack of rigorous transplant program supervision (Department of Health and Human Services: Medicare approved heart transplant centers. OIG Report No. OEI-01-02-00520, 2005.). These issues may explain, in part, the observed low rates of hospital compliance. Another important contextual factor may be Medicare’s leverage with hospitals as a payer of transplantation services to promote compliance (i.e., payer market leverage).”

3. I would also have appreciated a discussion of how CMS arrived at its minimum volume standards.

As requested by the reviewer, we clarify the origin of Medicare’s volume requirements in the second paragraph of our discussion. We state, “Medicare’s policy for transplant programs is predicated on the assumption that hospital compliance with volume standards will translate into better patient outcomes. Along this line of thinking, Medicare has linked both transplant program accreditation and procedural reimbursement to the attainment of its benchmarks. While Medicare’s volume thresholds were initially set without any evidence base, substantial literature has since been published documenting higher mortality after selected high risk procedures, including organ transplantation, at low-volume centers. Despite this, these data suggest that even the combination of a valid
benchmark and a powerful incentive (accreditation and reimbursement in this case) may
be insufficient to ensure uniform hospital/provider compliance.”

4. The authors should also clarify why compliance with volume standards should be
higher for kidney transplantation versus the other transplantation procedures. I
am guessing that this has to do with CMS reimburse(ement) for all care for ESRD
patients, but I couldn’t quite reason this through myself.

This is now clarified in the third paragraph of our background, where we state, “Under
the 1972 Social Security Act Amendments, Medicare guaranteed coverage to all patients
with chronic renal failure requiring hemodialysis. These benefits extend to kidney
transplant services, as well. Thus, Medicare’s stake in kidney transplantation is
significant. However, no such entitlement presently exists for patients with other chronic
disease states that commonly lead to organ transplantation; therefore, Medicare’s payer
market leverage is expected to be lower for other organ sites.”

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term,
which the author can be trusted to correct)

5. In the volume compliance regressions, the authors should also acknowledge that
there may be other demand factors that might influence caseload size for
hospitals, such as the number of patients with ESRD in the nearby area, the
number of smokers (for lung transplants), and the number of other providers. I
realize that it is difficult or impossible to obtain measures of these variables as
controls, but the shortcoming should be noted.

We now acknowledge this limitation in the fifth paragraph of our discussion, where we
now write, “Second, in this study, high Medicare market share was independently
associated with higher odds of compliance, even after adjustment for regional donor
organ supply. However, our adjustment for supply alone may be imperfect, as
compliance among hospitals performing liver transplantation, for which Medicare has the
smallest market share, was similar to that of kidney transplantation rather than heart or
lung transplantation. This indicates that there may be other factors influencing the
caseload size for hospitals (e.g., the number of patients with end stage renal disease in the
nearby area).”

6. The authors should also acknowledge that what they are interpreting as a leverage
effect may instead be a selection effect—hospitals that achieve lower mortality
rates may be able to attract more patients, particularly those covered by Medicare.

We acknowledge this limitation in the sixth paragraph of our discussion, where we state,
“A third limitation concerns this study’s observational design. Given the inherent
endogeneity of the data, the directionality of the association between Medicare market
share and hospital compliance cannot be directly discerned. For example, what we have
interpreted as a leverage effect may instead be a selection effect (e.g., the number of
patients with end stage renal disease in the nearby area) … However, we are reassured by
the results from our secondary analysis in which we examined the relationship between compliance and Medicare market share for lung transplants in the absence of a minimum threshold.”

Reviewer 2:
Major Essential Revisions (that the author must respond to before a decision on publication can be reached)

1. In the conclusions it is claimed that the data suggest that at least 30% of patients need to be at risk for motivation. However, only 20%-limit is used in the analyses, so it can be stated that 20%-limit seemed not to be enough for motivation, but the analyses do not tell if 30% is enough. As it should be quite straightforward to test different limits. It would be interesting to see how the effect varies by the cut-off; see e.g. figure 1 in Christian CK, Gustafson ML, Betensky RA, Dailey J, Zinner MJ: The Leapfrog volume criteria may fall short in identifying high-quality surgical centers. Ann Surg 2003, 238 (4):447-455.

The statement to which the reviewer is referring, “(our) findings would indicate that for infrequent diagnoses and procedures, more than 30% of a provider’s patients would need to be impacted by an incentive in order to motivate change,” was based on those data revealing that Medicare’s transplant market share varied by organ [57%, 28%, 27%, and 18% for kidney, lung, heart, and liver transplants, respectively (P<0.001)]; and volume-based benchmark compliance varied by transplant type [85%, 75%, 44%, and 39% for kidney, liver, heart, and lung transplants, respectively (P<0.001)]. Specifically, we were referring to heart and lung transplants.

Further, to examine the odds of hospital compliance with Medicare’s volume criteria as a function of Medicare’s market share, we constructed models in which our exposure (independent) variable was high Medicare leverage (a binary variable), indicating those hospitals where 20% of transplants, by type, were reimbursed through Medicare.

Controlling for organ supply and hospital case mix, we found that high market leverage was independently associated with compliance at hospitals transplanting kidneys (OR, 143.43; 95% CI, 18.66 – 1102.56), hearts (OR, 2.81; 95% CI, 1.51 – 5.25), and lungs (OR, 3.32; 95% CI, 1.63 – 6.76).

As requested by the reviewer, we did re-run these models of high Medicare penetration (controlling for organ supply and case mix), varying the threshold used from 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 to 100%. Unfortunately, we ran into computational problems with the models for those hospitals performing liver, heart, and lung transplants. As we raised the threshold above 20%, the numbers of hospitals in these cells got smaller and smaller. Consequently, our point estimates became unstable, and eventually, the models just failed.

For those hospitals performing kidney transplants, we noted that the odds of volume compliance lowered with increasing Medicare leverage, but there was still a strong association observed up to the 50% threshold. This may indicate that the relationship
between odds of volume compliance and Medicare leverage is a non-linear function, such that after a certain point, payer penetration does not really matter. Additionally, as we climb over the 50% threshold for kidney transplants, we are likely running into some potential selection issues, and those numbers may not be overly useful.

As such, we do not believe that these data are sufficient to establish a true cut point. Hence, we have smoothed our conclusions to read, “While others have suggested that an incentive applying to less than 15% to 20% of a provider’s patient panel is unlikely to induce provider response, these findings would indicate that for infrequent diagnoses and procedures, more than 30% of provider’s patients might need to be impacted by an incentive in order to motivate change.”

2. Testing in the tables 1 and 2 is not adequately reported. Please report n for each population. There should be a column for P-values and/or significance-stars. Why (is the standard deviation) of age is about three-fold (higher) in non-compliant hospitals? Is that fact in concordance with the assumptions of a T-test? For categorical variables with several classes summing up to 100% the equivalence of distributions should be tested instead of class-wise testing. Why race-other class in liver transplantation has no stars even though the difference seems to be rather large? Confidence intervals would be nice.

As requested by the reviewer, we have now included in Tables 1 and 2 the population totals (n) and separate columns for the significance stars. Given the large differences between the standard deviations of age at compliant vs. non-compliant hospitals, we also re-ran our T-test calculations for mean patient age, assuming unequal variances. For those categorical variables in Tables 1 and 2, chi-square analyses were performed.

3. It is difficult to follow the description of the model for Medicare market leverage. What is your dependent variable? What are your independent variables? How many observations do you have in the actual model? What are the mentioned potential cluster effects in this case? Is the count of donor organ supply hospital specific? Is it reasonable to use counts if hospitals are of different size? OR of magnitude 143 is pretty high; what does it tell? A table for the results (including estimates also for variables used in adjusting) would be nice. Have you considered alternatives for the use of categorized variables (see the cut-off comment above)?

We now clarify these models for the reader in the third paragraph of the statistical analyses subsection of our methods. Here we state, “For these models, the unit of analysis was the hospital-year. The outcome was CMS volume benchmark compliance. The exposure was high Medicare market leverage (the binary variable described above). Given the well-documented regional variation in transplantation, these models were adjusted for donor organ supply, introduced as a continuous count. The count of donor organ supply was UNOS-region specific, into which each hospital was sorted. These models were also adjusted for hospital case mix. For this, patient-level data were aggregated to the hospital level in order to derive a mean count of the Elixhauser codes.
specific for the four different transplant procedures, which described each hospital’s “typical” patient.” The concern for cluster effects arose because of the potential for a hospital to be included in a multiple number of years, which could induce a correlation and would require adjustment.

As for the reviewer’s concern about the magnitude of our ORs, we would refer him to an article by Zhang et al: **What’s the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA 1998, 280:1690-1691.** When the incidence of an outcome of interest (volume compliance in our analysis) is common in the study population (>10%), the adjusted odds ratio derived from the logistic regression can no longer approximate the risk ratio. The more frequent the outcome, the more the odds ratio overestimates the risk ratio when it is more than 1. The authors recommend a way of approximating a risk ratio (RR) from the adjusted odds ratio, using the formula below.

In a cohort study, \( P_0 \) (low Medicare penetration), indicates the incidence of the outcome of interest in the non-exposed group and \( P_1 \) in the exposed group (high Medicare penetration).

\[
RR = \frac{OR}{(1 - P_0) + (P_0 \times OR)}
\]

Through this approach, the adjusted risk ratio for volume compliance at hospitals performing kidney transplants in which Medicare has greater than 20% penetration is 18.42 (95% CI, 10.56 – 20.99).

4. Was the GEE-approach used to accommodate the panel structure or the hierarchical patient-volume structure in the NIS data? Correct the text to be more precise and change the reference, if the longitudinal analysis was not used. Report also the odds ratios for the variables used in the adjusting in the table 3 (for example, use a new table for each transplant type or (transposed) tables for unadjusted and adjusted odds ratios). Even though these variables are not of your primary interest, they offer interesting information about the operative mortality and make it easier to evaluate the adequacy of the model.

We now clarify our rationale for the GEE approach. In the fifth and sixth paragraphs of the statistical analyses subsection of our methods, we state, “Next, the relationship between patient-level operative mortality and hospital volume compliance was examined using generalized estimating equations (GEE). Given that not all of the hospitals were represented in multiple years of our sampling frame, the data did not represent a true panel structure. However, there was potential correlation between observations as patients were clustered within hospitals. Therefore, GEE models with a logit link and an exchangeable correlation matrix were fitted to produce a population-averaged odds ratio (OR) and a robust estimator to correct the standard errors for the potential correlation of observations within hospitals. For our analysis of operative mortality (outcome), our unit of analysis was at the patient level. Our exposure was hospital-level volume benchmark compliance. Each model was adjusted for patient age, gender, race, comorbidity, and
insurance type, as well as treatment year, teaching status, for-profit status, ownership, hospital bed capacity and census region.” Correspondingly, the associated reference is now Chao EC: *Structured correlation in models for clustered data*. *Stat Med* 2006, 25:2450-2468.

As requested by the reviewer, we have also broken Table 3 out into four separate tables (Tables 3-6) for each transplant type, in which each of the covariates are listed along with their adjusted odds ratios and 95% confidence intervals.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

5. Add a comment (and references) on the quality of NIS data.

In the data source subsection of our methods, we comment on the quality of the NIS data. We state, “(The NIS) is a useful dataset for multi-year longitudinal studies. To date, more than 25 manuscripts have been published using multiple years of the NIS that cover a breadth of healthcare topics, including solid organ transplantation.”

Included references follow below:


7. What is the coverage of the UNOS data (sample or total data)? How was the UNOS data “standardized” to be compatible with the NIS-sample? Discuss about the possible shortcomings.

The Organ Procurement and Transplantation Network (OPTN), is the unified transplant network established by the United States Congress under the National Organ Transplant Act of 1984 to be operated by a private, non-profit organization under federal contract. UNOS was awarded the very first OPTN contract on September 30, 1986, and has continued to administer the OPTN under contract with the Health Resources and Services Administration of the United States Department of Health and Human Services for more than 16 years and four successive contract renewals. As part of the OPTN contract, UNOS has established a system for the collection, storage, analysis and publication of
data pertaining to the patient waiting list, organ matching, and transplants. The data from UNOS on the number of donor organs recovered are total data, not a sample. We assigned each hospital in our study cohort to one of the 11 regions created by UNOS, based on the hospital’s two-letter state postal code (this is provided in the NIS). UNOS data were then merged with the study population files on UNOS region and year.

8. Explain more carefully how many hospitals were included in multiple years in NIS. Has there been any structural changes between 1993 and 2003 which may have had some impact on the association between volume compliance and market share?

In the last paragraph of the statistical analyses subsection of our methods, we clarify the number of hospitals sampled in multiple years. We state, “111, 54, 60, and 40 hospitals performing kidney, liver, heart, and lung transplants were sampled in multiple years.”

In 1998, the sampling method for the NIS changed to better reflect the cross-sectional population of hospitals. The hospital stratification variables were redefined, rehabilitation facilities were dropped from the target universe, and sampling preference was no longer given to prior year NIS hospitals. According to the HCUP Methods series, the changes to strata definitions have little effect on estimates of trends. Additionally, there was a change from the use of total discharges to the use of hospital discharges to estimate NIS discharge weights using data from the American Hospital Association (AHA) annual surveys of hospitals. However, NIS discharge weights were not employed in this analysis, as our sample constituted nearly the entire universe of transplant hospitals in the U.S.

9. The sentence “Based on the hospital-year volume for a given discharge record, a binary volume compliance variable was assigned at the hospital level and served as the unit of exposure” is unclear. Isn’t the discharge record patient specific? Unit of exposure of what?

We clarify this for the reader in our ascertainment of hospital volume compliance subsection of the methods. We now state, “The number of transplant procedures performed at each hospital was ascertained using a unique hospital identification code. Based on Medicare’s coverage announcement for lung transplants, combined heart-lung transplants were counted toward both the heart and lung case volume totals for an individual hospital. Each hospital was assigned a volume for each year it participated. Each individual hospital-year was considered independently. Binary variables were then constructed indicating whether the hospital-year was compliant with Medicare’s volume benchmark for each transplant type. The hospital-year volumes and binary volume compliance variables were then assigned to the discharge records that emanated from them. This was done for each transplant type, using those minimum volume requirements specified in the Federal Register.” For our analysis of operative mortality (outcome), our unit of analysis was at the patient level. Our exposure was then hospital-level volume benchmark compliance.
1. Are the volume requirements evidence-based or consensus statements? Move/copy the reference justifying volume requirements in discussion also to the methods-section. Discuss the use of other outcomes than operative mortality.

As requested by the reviewer, we clarify the origin of Medicare’s volume requirements in the second paragraph of our discussion. We state, “Medicare’s policy for transplant programs is predicated on the assumption that hospital compliance with volume standards will translate into better patient outcomes. Along this line of thinking, Medicare has linked both transplant program accreditation and procedural reimbursement to the attainment of its benchmarks. While Medicare’s volume thresholds were initially set without any evidence base, substantial literature has since been published documenting higher mortality after selected high risk procedures, including organ transplantation, at low-volume centers. Despite this, these data suggest that even the combination of a valid benchmark and a powerful incentive (accreditation and reimbursement in this case) may be insufficient to ensure uniform hospital/provider compliance.”

We also acknowledge the lack of additional outcome measures available in our data source. In the seventh paragraph of our discussion we state, “(T)ransplant-relevant clinical endpoints are limited within these administrative data. While alternative transplant outcomes (e.g., incidence of delayed graft function and graft survival) would have made this analysis richer, we could only examine the clinical and resource use information included in a typical discharge abstract. Moreover, our primary measure, operative mortality, is an infrequent event, particularly following kidney transplant. Though a significant difference in patient mortality for liver transplantation was demonstrated at volume compliant vs. non-compliant hospitals, the ability to detect a difference for kidney, heart, and lung transplants was perhaps limited by sample size. Further investigation is, therefore, warranted to examine the effect of volume compliance on more common, long-term outcomes using clinical data.”

10. Give a new reference for the robust variance estimates (for instance, some tutorial paper pointing out the importance of the technique in a similar context would be nice).


11. Please state to which hypotheses the P-values in the first and third sentence of results-section refer to.

The hypotheses tested here are 1) Does the proportion of transplants for which Medicare is the primary payer differ by transplant type? and 2) Does the average volume benchmark compliance differ by transplant type?

12. Smooth your interpretations and wordings to be in concordance with the data and the statistical evidence (e.g., the “overgeneralizing” text referring to tables 1 and 2 in the results section, lower likelihood interpretation in discussion sentence 2,
and the statement about the ultimate determinants of compliance in discussion page 12 beginning).

As requested by the reviewer, these sections of text have been smoothed to be in concordance with the data presented.

**Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)**

13. The introduction is very USA-specific. Some motivation and discussion about the implications beyond the referred debate would be of interest to an international reader.

While the issue of transplant program compliance with Medicare’s volume benchmarks pertains to the U.S. alone, the topic of payer market leverage and the insights gained from our analysis are applicable to other healthcare systems that lack a single payer.

14. Give a short introduction to Medicare system (or references to such information). Are there some sanctions for hospitals, if Medicare’s minimum standards are not fulfilled? Are the “small” hospitals really ready for dropping out their potentially profitable transplantation operations (and the expertise of their surgeons on this respect) because of some more or less arbitrary volume cut-off?

Medicare is the United States’ largest health insurance program, covering nearly 40 million Americans. It is a government-run health insurance program for people 65 years of age and older, some disabled people under 65 years of age, and people with end-stage renal disease. A complete discussion of Medicare is outside the scope of the manuscript. Sanctions for non-compliant hospitals with Medicare’s minimum standards include loss of transplant program accreditation, which is tied to procedure reimbursement. As stated in the last paragraph of our discussion, “Medicare requires that transplant programs undergo review every three years; however, … the Office of the Inspector General raised questions about the rigor of this oversight (Department of Health and Human Services: Medicare approved heart transplant centers. OIG Report No. OEI-01-02-00520, 2005.). A new rule change has been proposed to address these concerns [Hospital conditions of participation: Requirements for approval and re-approval of transplant centers to perform organ transplants. 42 CFR Parts 405, 482, and 488 (proposed), 2005.], but current lack of enforcement may contribute to the relatively low compliance among heart and lung transplant programs.”

15. Report which Elixhauser comorbidities were considered to be related to each type of transplantation.

The specific Elixhauser comorbidities used are now displayed in Tables 3 to 6.

On behalf of all the authors, I would like to thank you for your time and consideration. We are happy to clarify any additional issues that you may have. We continue to believe that *BMC Health Services Research* is the ideal venue for this manuscript.
Best wishes,

John M. Hollingsworth
Department of Urology
University of Michigan
1500 East Medical Center Drive
3876 Taubman Center
Ann Arbor, Michigan 48109-0330
USA
Email: kinks@med.umich.edu