MAINTAIN AND EXPAND
THE HEALTHCARE COST AND UTILIZATION PROJECT (HCUP)
Contract No. HHSA-290-2006-00009-C

RAPID CYCLE ESTIMATES TEST METHODS REPORT
DELIVERABLE #825.11B

July 31, 2009

Submitted to:
Jenny Schnaier, Project Officer
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850

Submitted by:
Thomson Reuters
5425 Hollister Ave, Suite 140
Santa Barbara, CA 93111

Table of Contents

Introduction 1

Historical experience for the receipt of state inpatient data 1

Outcome statistics and patient conditions 4

Estimation strategies 6

Use of historical data 6

Sampling designs and weighting methods 6

Statistical methods 6

Results 8

Conclusion 9

List of Tables

Table 1. Patient conditions and procedures 4

Table 2. Codes for Methicillin-resistant *Staphylococcus Aureus* (MRSA) 5

Table 3. Codes for Septicemia 5

Table 4. Comparison of unstratified early estimates 11

Table 5. Comparison of stratified early estimates 12

Table 6. ALOS correlations between state and NIS estimates, 2002 – 2006 17

Table 7. Mean charge correlations between state and NIS estimates, 2002 – 2006 18

Table 8. Mortality rate correlations between state and NIS estimates, 2002 – 2006 19

List of Figures

Figure 1: Timing of HCUP Data Delivery…… 2

Figure 2. HCUP States Providing Annual Data by August 1 Following Each Data Year (turquoise)…… 3

Figure 3. RMSE for ALOS, stratified estimates vs. previous NIS estimate…… 13

Figure 4. RMSE for Mean Charges, stratified estimates vs. previous NIS estimate…… 14

Figure 5. RMSE for Mortality Rates, stratified estimates vs. previous NIS estimate…… 15

Figure 6. RMSE for Total Discharges, stratified estimates vs. previous NIS estimate……16

Introduction

This report presents an abbreviated summary of a feasibility study to determine whether it is possible to generate accurate national estimates of selected statistics based solely on annual inpatient discharge data from a subset of "early" states. For these analyses, we defined two sets of early states: 1) those we anticipated would deliver their 2008 data by August 1, 2009, and 2) a slightly larger subset consisting of those we anticipated would deliver their 2008 data by September 1, 2009.

The broader task for Rapid-Cycle National Estimates is to develop more timely nationwide and regional estimates from inpatient (IP) and emergency department (ED) data. For both IP and ED, our initial focus was on developing estimates using existing data streams of annual inpatient data. Later efforts will address the possibility of obtaining data from HCUP Partners more frequently (e.g., quarterly).

We begin this report with a description of our historical experience concerning the timing of annual inpatient data. Based on this experience combined with our best judgment concerning the delivery of 2008 data, we identified two subsets of "early" states to use in these analyses.

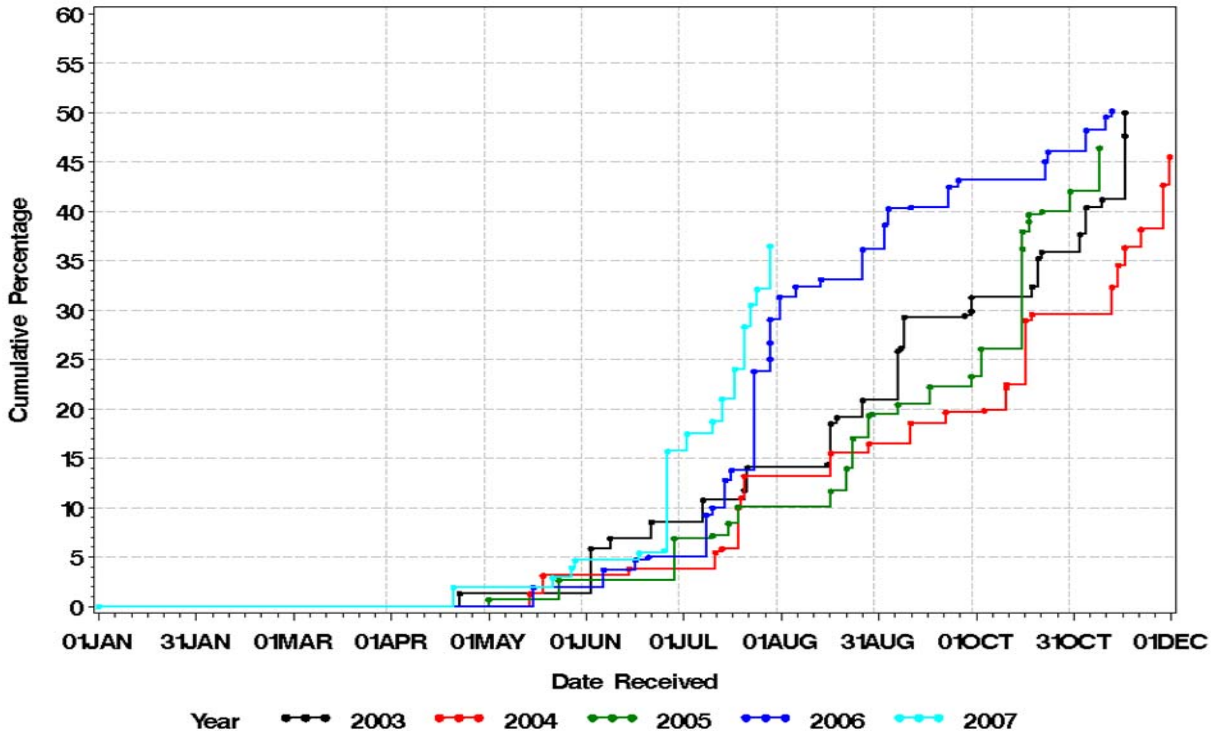
Using 2002 – 2006 state inpatient data (SID), we tested the ability of two methods for predicting full-sample Nationwide Inpatient Sample (NIS) statistics, and compared the performance of those methods to simple estimates that simply predicted each year's NIS value by carrying forward the previous year's NIS value. We reasoned that simply predicting each year's NIS value by the previous year's NIS value provided a minimal baseline performance for predictions, and that an acceptable prediction method should nearly always do much better than carrying the previous year value forward. We tested these methods on predicting average charges, average lengths of stay, average mortality rates, and the total number of discharges for 28 diagnoses and procedures. In addition, we predicted rates for four patient safety indicators.

We summarize the methods and the results of these analyses below. We determined that the performance of the two methods we tested using annual data were not enough of an improvement over carrying the previous NIS value forward, and we offer some recommendations for further research.

Historical experience for the receipt of state inpatient data

We began by summarizing the timing of inpatient data arrival over the past four years. Figure 1 shows the percentage of the total U.S. discharges covered by the receipt of 2003 - 2006 inpatient data, along with results to date (through August, 2008) for the 2007 data year.
In Figure 1, each dot represents the receipt of data from one state. The x-axis reflects the date that the state data file was received. The y-axis represents the cumulative percentage of U.S. discharges received. For example, for data year 2003 (black line), as of July 1, 2004, HCUP had received data from four states (4 black dots prior to July 1) representing about 8 percent of total U.S. discharges for 2003. For earlier data years, an even lower number of states and lower percentage of total discharges had been received by July 1. Looking one month later, the percentage of U.S. discharges collected by August 1 hit a high of around 30 percent for 2006, which is much higher than the 10 to 14 percent of U.S. discharges collected by August 1 for the 2003-2005 data years. We used these two cutoff dates, August 1 and September 1, for this study.

Figure 1: Timing of HCUP Data Delivery



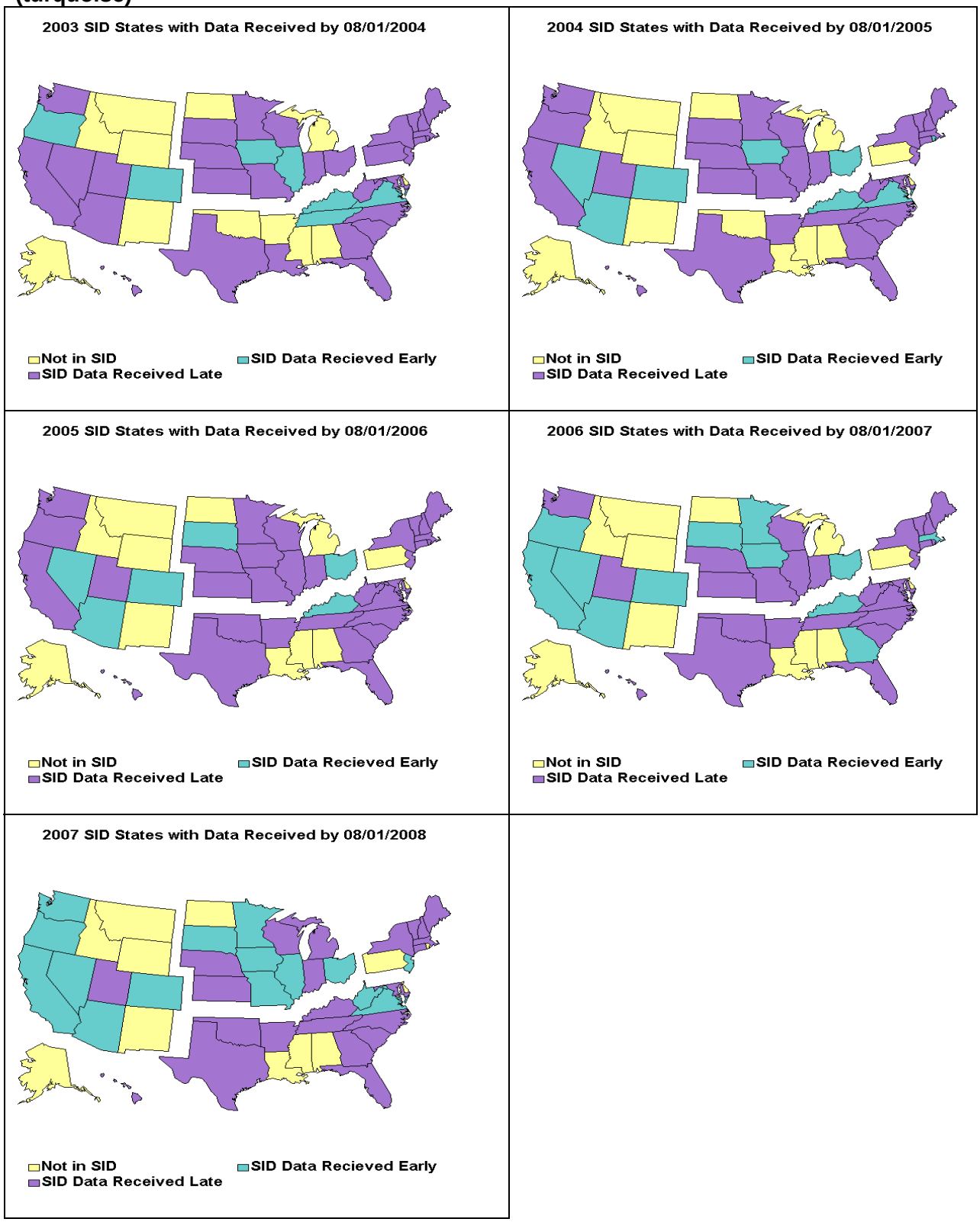
The choice of cutoff dates was also affected both by the number of discharges and by the geographic distribution of early states. The number of discharges is large. Even 10 percent of the U.S. total annual discharges translates into about 4 million discharges, which is about half of the size of the NIS. The geographic distribution of states that delivered data by August 1 is shown in Figure 2 for data years 2003 – 2007. For the data year 2007, only one Northeast state (NJ) delivered their inpatient files before August 1. The results were variable for the other three regions, although usually at least two states delivered early in each of those regions.

Based on these considerations, *and based on our expectations for the delivery of 2008 data*, we developed analytic files consisting of annual SID data for the period 2002 – 2006 from the following "early" states, by region:

- West: (August 1) AZ, OR, NV, CA, CO, and WA.
- Midwest: (August 1) IA, SD, MN, OH, IL. (September 1 add WI and MO).
- Northeast: (August 1) NJ.
- South: (August 1) WV (September 1 add MD).

We omitted IN, VA, and ME because there was not 2003 SID data for IN, 2005 or 2006 SID data for VA, and 2004 or 2005 SID data for ME. Clearly, the representation was very thin for the Northeast and the South regions. The September 1 subset of states added only three states to the August 1 subset: WI, MO, and MD.

Figure 2. HCUP States Providing Annual Data by August 1 Following Each Data Year (turquoise)



Outcome statistics and patient conditions

We generated estimates of nationwide discharge counts, mortality rates, mean total charges, and average lengths of stay (ALOS) for patients with selected principal diagnoses and procedures. Table 1 contains the list of 28 conditions and procedures employed in our analyses.

Table 1. Patient conditions and procedures

	Single-	Multi-	ICD-9-CM codes
Diagnoses			
1. Asthma	128		
2. Cancer - Breast			174.x, 175.x, 233.0
3. Cancer - Colon			153.x,159.0, 230.3
4. Cancer - Lung			162.2-162.9, 231.2
5. Cancer - Prostate			185, 233.4
6. Diabetes	49, 50		
7. Acute Myocardial Infarction	100		
8. Depression (MHSA Mood	657		
9. Decubitus Ulcers		12.3.1	
10. Stroke, ischemic			433.x1, 434.x1, 436
11. Dementia	653		
12. Methicillin-resistant			See
13. Pneumonia	122		In addition to CCS122, add code 487.0
14. Peptic Ulcers			531.00-533.91
15. Septicemia			See Table 3
16. Cellulitis			681.00-681.9
17. Renal Failure, Acute	157		
Procedures			
18. Hip Replacements	153		
19. Knee replacements	152		
20. Bariatric Procedures			44.31-44.39, 44.68, 44.95-
21. C-sections	134		
22. CABG	44		
23. PTCA	45		
24. Spinal fusion			81.00-81.09, 81.31-81.39,
Patient Safety Indicators			
25. Accidental puncture or laceration			#
26. Iatrogenic pneumothorax			#
27. Postoperative hemorrhage or			#
28. Postoperative respiratory failure.			#

PSI specifications are available at http://www.qualityindicators.ahrq.gov/psi_download.htm.

Table 2. Codes for Methicillin-resistant *Staphylococcus Aureus* (MRSA)

MRSA requires 2 codes:
V09.0 (infection with microorganisms resistant to penicillins) plus a code for Staph aureus Infection. Staph aureus infections:
482.41 Pneumonia due to Staphylococcus aureus
038.11 Staphylococcus aureus septicemia
041.11 Staphylococcus aureus infection in conditions classified elsewhere and of unspecified site

Table 3. Codes for Septicemia

Any of the following codes:
038.0 Streptococcal septicemia
038.10 Unspecified staphylococcal septicemia
038.11 Staphylococcus aureus septicemia
038.19 Other staphylococcal septicemia
038.2 Pneumococcal septicemia
038.3 Septicemia due to anaerobes
038.40 Septicemia due to unspecified gram-negative organism
038.41 Septicemia due to hemophilus influenzae (H. influenzae)
038.42 Septicemia due to Escherichia coli (E. coli)
038.43 Septicemia due to pseudomonas
038.44 Septicemia due to serratia
038.49 Other septicemia due to gram-negative organism
038.8 Other specified septicemia
038.9 Unspecified septicemia
003.1 Salmonella septicemia
054.5 Herpetic septicemia
036.2 Meningococcemia
022.3 Anthrax septicemia

Estimation strategies

We used historical SID data to simulate two estimation methods and assess the accuracy of those methods assuming that we had applied the methods in previous years, as we explain in this section.

Use of historical data

We developed and tested the accuracy of our methods using historical data. In particular, we generated estimates using only data from the subsets of "early" states and compared those estimates to the actual known national estimates using the entire NIS for each of four data years: 2003, 2004, 2005, and 2006. We employed two definitions of "early": states for which we forecast would deliver their 2008 data by 1) August 1, 2009 and, 2) by September 1, 2009. The later cutoff date included more data, potentially increasing the accuracy of estimates. This gave us an idea of the accuracy that we could expect to obtain for the 2008 data year rapid cycle estimates.

Sampling designs and weighting methods

Our objective was to generate early NIS estimates, not to create an early version of the NIS. Therefore, we did not sample hospitals as we do for the NIS. However, for one of the methods we tested, which we call "stratified estimates," we post-stratified the early SID hospitals and calculated sample weights using the same methods as those we used for the NIS (see NIS Documentation). That is, we weighted sample discharges within each hospital stratum to the total population of discharges estimated from AHA data in that stratum.

The other estimation method, which we call "unstratified estimates," simply weighted discharges by the frequency of discharges observed in the early states.

Statistical methods

While several years of historical SID data might have offered better estimates of long-term trends, considerable time and expense are associated with combining multiple years of historical data for estimation, and we wanted to generate estimates as quickly and as efficiently as possible. Consequently, we used only two years of data at a time to produce estimates for each year.

That is, we estimated statistics for each year 2003, 2004, 2005, and 2006 using only two years of SID data, the most current year and the previous year, combined with the previous year of NIS data. For example, for ALOS we estimated the *change* in ALOS between 2004 and 2005 from the early SID states. We assumed that the change that occurred for the early SID states was a good estimate of the change that would occur for the entire NIS. Therefore, we estimated the 2005 ALOS by multiplying the ALOS estimated for the 2004 NIS by the *estimated 2004-2005 change* that occurred for the early SID states. We then compared this *estimated* 2005 ALOS to the *actual* 2005 ALOS estimated from the full 2005 NIS. The next section of this report presents the formulas for these estimates.

Perhaps the simplest estimate for the next year is the previous year NIS estimate. For example, we could use the 2004 NIS ALOS as an estimate of the 2005 NIS ALOS. Our proposed methods should do at least as well as this "carry the last year forward" estimate. Therefore, we used these simple estimates as comparison benchmarks for the performance of the methods that used early SID data, rather than try to improve estimates based on annual data.

Below, we present the equations for the two methods used to produce estimates from the early SID data. Other, more sophisticated estimators were originally proposed for this study.

However, AHRQ and the Thomson Reuters team agreed that resources would be better spent exploring the potential and advantages of obtaining early quarterly data from a larger number of states than would be possible for early annual data.

Formula for unstratified estimates

The unstratified estimates used a weighted average of the state-specific changes in the SID between two years to estimate the overall change in the NIS between those same two years. The weights were simply the proportion of discharges in the early SID states.

Let's say that $Y_{K,T}$ is the estimated statistic (e.g., ALOS) based on the SID for state K in year T, and say that $Y_{NIS,T}$ is the estimated statistic based on the full NIS for the U.S. in year T. The unstratified national estimate for year T based on N early states was the product of the full NIS estimate for year T-1 and a weighted average of the state-specific changes from year T-1 to year T for the early states, according to the following formula:

$$\hat{Y}_{NIS,T} = Y_{NIS,T-1}\left\{\frac{\sum_{K=1}^{N} W_{K,T}\cdot Y_{K,T}/Y_{K,T-1}}{\sum_{K=1}^{N} W_{K,T}}\right\} = Y_{NIS,T-1}\cdot\sum_{K=1}^{N} P_{K,T}\cdot R_{K,T}$$

The weights $W_{K,T}$ were equal to the number of early NIS discharges represented by state K in year T. The ratios $R_{k,T}$ were the ratio of the statistics between the base year and the prediction year for state K, and the proportions $P_{K,T}$ represented the proportion early SID discharges from state K in year T.

Formula for stratified estimates

The stratified estimates used stratum-specific sample weights to the population of discharges. As explained earlier, we stratified hospitals and employed AHA data to estimate the total population number of discharges in each stratum. For J hospital strata, the estimate was given by the following formula:

$$\hat{Y}_{NIS,T} = Y_{NIS,T-1}\left\{\frac{\sum_{S=1}^{J} W_S\cdot Y_{S,T}/Y_{S,T-1}}{\sum_{S=1}^{J} W_S}\right\} = Y_{NIS,T-1}\cdot\sum_{S=1}^{J} P_S\cdot R_{S,T}$$

In this formula, W_S was the weight for discharges in stratum S (the ratio of the AHA estimate of total discharges to the number of observed discharges), and $Y_{S,T}$ was the estimated statistic in stratum S based on the early states for year T. We anticipated that the stratified estimates would be an improvement over the unstratified estimates to the extent that the year-to-year changes varied by hospital stratum.

Root mean squared error (RMSE)

For each outcome, we attempted to predict the "true" NIS estimate for the next data year. Therefore, we also calculated the root mean squared error (RMSE) of prediction:

$$RMSE = \sqrt{\frac{\sum_{T=1}^{4}\left(\hat{Y}_{NIS,T} - Y_{NIS,T}\right)^2}{4}}$$

This is a common measure of estimation accuracy, which we will use to compare the performance of the various methods. Smaller values of RMSE are better.

Results

As stated earlier, we nearly always expect the proposed methods to perform better than simply using the previous year NIS estimate. **Table 4** summarizes the performance of the unstratified estimate and **Table 5** summarizes the performance of the stratified estimate compared with the previous year estimate.

For each condition or procedure, we generated four early-state estimates, one for each year 2003-2006. For the four PSIs, we generated only two estimates, one for 2005 and another for 2006, because we lacked data for PSIs in earlier years. Therefore, we generated 104 estimates for each outcome (four estimates for each of the 28 conditions/procedures and two estimates for each of the four PSIs).

For each outcome, Table 4 shows the number of times for which

1. the August SID unstratified estimate was closer to the next year NIS value than was the previous-year NIS value;

2. the September SID unstratified estimate was closer to the next year NIS value than was the previous-year NIS value; and

3. the September SID unstratified estimate was closer to the next year NIS value than was the August SID unstratified estimate.

For example, for acute myocardial infarction (AMI), the unstratified ALOS estimate was closer than the previous-year NIS estimate for all four years whether using the August 1 cutoff (column 2) or the September 1 cutoff (column 3) for early states. Therefore, the unstratified estimates of AMI ALOS were always better than were the previous-year NIS estimates. However, the unstratified ALOS estimate for AMI based on the September 1 cutoff states was better than the estimate based on the August 1 cutoff states (column 4) for only one year out of four. This is somewhat surprising because more states are included in the September 1 subset of early states.

In Table 4, the row labeled “All” summarizes the results for unstratified estimates over all 104 estimates for the conditions, procedures, and PSIs. These overall results are disappointing. The estimates using early state data were better than the previous-year NIS estimate only about 60% of the time for ALOS, about 82% of the time for charges, about 60% of the time for mortality, and about 75% of the time for discharge counts.

Similarly, summarizes the results for the stratified estimates. The stratified estimates fared about the same as the unstratified estimates. This is surprising because we expected that year-to-year changes might vary systematically among different types of hospitals and regions of the country, and that the stratified estimates would weight these changes more appropriately than the unstratified estimates.

We also compared root-mean squared errors (RMSE) of prediction among 1) the previous year NIS estimate, 2) the stratified early state August estimates, and 3) the stratified early state September estimates. While the counts shown in Table 5 give us a sense of the *frequency* (the number of years) for which the stratified early estimates were better than the previous-year NIS estimate, the RMSE gives us a sense of the relative *magnitude* of the errors among estimators.

Smaller RMSE is better.

The RMSE results are shown in Figure 3 (ALOS), in Figure 4 (mean charges), in Figure 5 (mortality rates), and in Figure 6 (total discharges). These results mirror the results shown for stratified estimates in. For example in Figure 3, for asthma ALOS, the RMSE for the previous year NIS is 0.0102 (green bar), which is considerably smaller than the RMSE of 0.0193 for the August early states (blue bar) and the RMSE of 0.0187 for the September early states (red bar). Likewise, the results for asthma ALOS in Table 5 show that the previous year NIS is closer to the current year NIS estimate 4 out of 4 times compared with the August early state estimate and closer 3 out of 4 times compared with the September early state estimates.

Looking across Figures 3 through 6, we see that the green bars (previous year NIS estimates) are often shorter than or about the same length as the blue and red bars (early state estimates), indicating that the previous year NIS estimates are better—have smaller mean squared errors—or are nearly as good as the early state stratified estimates. Only for average charges (Figure 4) are the early estimates almost always substantially better than previous year NIS estimates, a result that is consistent with the results for average charges in Table 5.

Table 6, Table 7, Table 8 give us some insights into these results. These tables show the correlation between state estimates and the NIS estimate over the five years 2003-2006. We have only five estimates per state, but it is still useful to observe the general correlation patterns. Correlations of 70% or higher are dark green, correlations of 0 to 70% are light green, and negative correlations are red. A correlation above 70% corresponds to a state that explains at least half of the NIS variance (over the five years). A correlation of 99% corresponds to a state with a trend that is virtually identical to the NIS trend.

For example, for AMI ALOS, the correlation is 99% or more between the NIS and three states -- CA, NV, and WI. Indeed, the trends for CA, NV, and WI are virtually identical to the NIS trend over the five years (based on plotted trends not shown in this report). A large negative correlation corresponds to a state with estimates that are going in the opposite direction of the NIS estimates. For example, the AMI ALOS correlation is large and negative between the NIS and two states -- AZ and OR. The AZ and OR trends tend to move upward while the NIS trend tends to move downward. Overall, most of the state-specific trends correlate well with the NIS trend for AMI ALOS, consequently, the early state estimates tend to do better than the previous-year NIS estimate for all four years (consistent with the results shown in Table 4 and Table 5).

In aggregate, we can see that the correlations are mostly green for charges (Table 7), and that is where the early state estimates do best compared with the previous-year NIS estimate. However, there are many negative correlations mixed with positive correlations for ALOS and mortality, explaining the relatively poor results for those outcomes (Table 4 and Table 5).

Conclusion

The two methods for early estimates that we tested in this study proved somewhat disappointing using annual discharge data. Perhaps the most important outcome for which we would like good forecasts is the total number of discharges. For total discharges, about 75% of the time the stratified estimates based on early data were closer to the full NIS estimates than were the previous year NIS estimates (Table 5). In addition, the RMSE for total discharges was *substantially* smaller for about half of the conditions (Figure 6). However, the standard of comparison—the previous year NIS estimate—represents a minimal threshold that requires little effort to produce and that lacks one piece of information that we desire: the direction of the trend (e.g., is diabetes ALOS going up or down?). Moreover, the results for ALOS and mortality were

more disappointing still.

Therefore, rather than expend more effort on improving estimates using annual data, AHRQ and Thomson Reuters researchers concluded that we should shift our efforts to developing estimates using early quarterly data. Preliminary investigation has shown the potential for obtaining early quarterly data from many more states than those for which we can obtain early annual data. We anticipate that a broader array of states will be more representative of the overall NIS trends and therefore lead to better early estimates across the entire panel of outcomes and conditions.

We recommend that AHRQ evaluate the feasibility of generating early national estimates from quarterly data in much the same vein as we have evaluated the use of annual data. One should start by identifying the states for which it is possible to obtain early quarterly data (one, two, three, or four quarters of early data) by a designated cutoff date, such as August 1. We should then use the historical SID annual data, dividing it up into quarters, to simulate the arrival of early SID quarterly data and attempt to estimate national outcome statistics for the same list of diagnoses, procedures, and PSIs that we used in this study. One additional complexity to the estimation process will be to account for the within-year trends for some diseases. However, we do not expect that to pose a significant challenge.

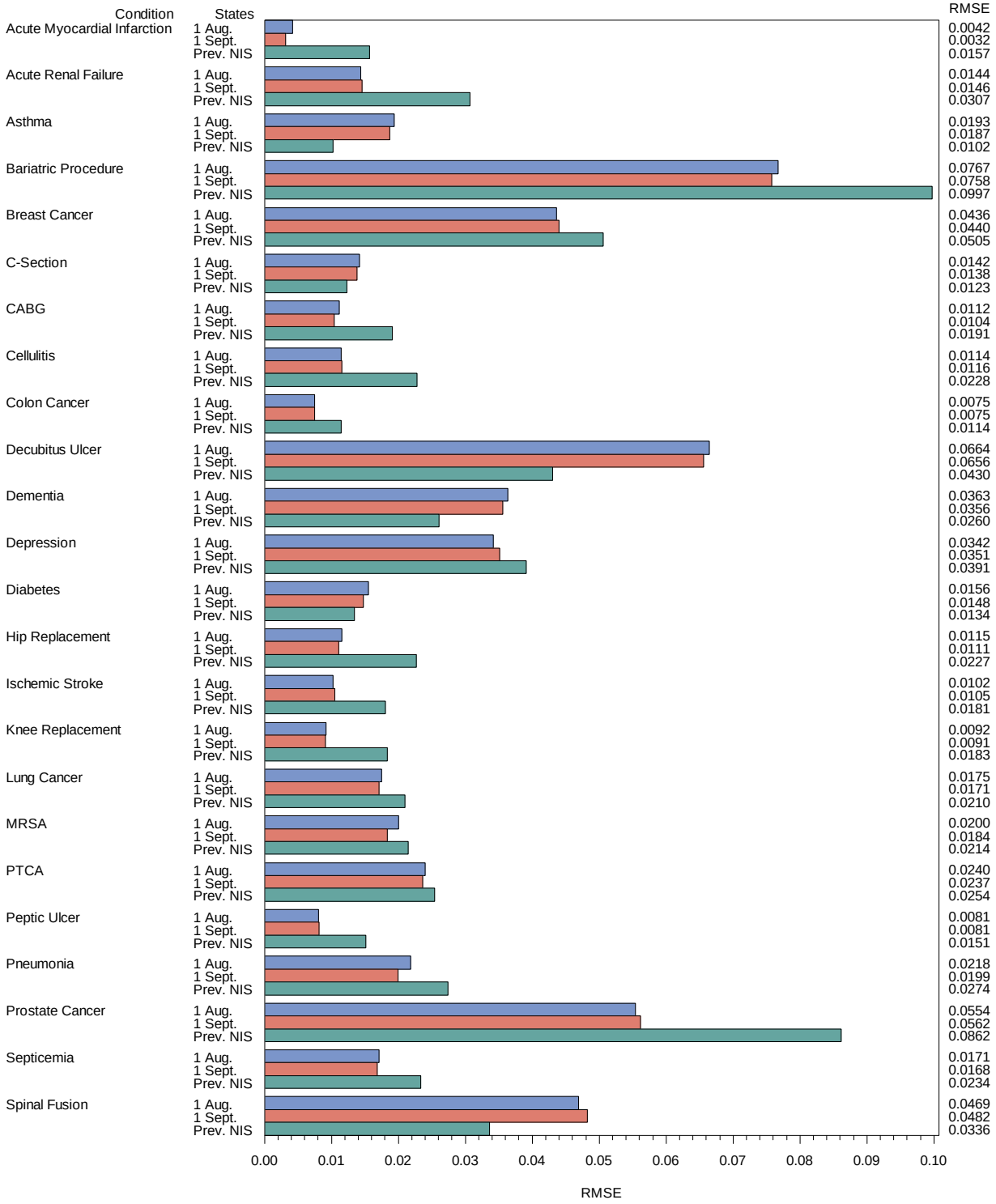
Table 4. Comparison of unstratified early estimates

	Statistic											
	ALOS			Charges			Mortality			Discharges		
	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.
	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better
All	62 (57 %)	63 (61 %)	50 (48 %)	85 (82 %)	85 (82 %0	51 (49 %)	64 (62 %)	61 (59 %)	60 (58 %)	76 (73 %)	79 (76 %)	61 (59 %)
Accidental Puncture or Laceration (PSI)	1	1	1	2	2	0	0	0	2	1	1	1
Acute Myocardial Infarction	4	4	1	4	4	2	4	4	1	3	3	3
Acute Renal Failure	4	4	1	3	3	1	4	4	2	4	4	3
Asthma	1	1	1	3	3	2	1	1	2	4	4	3
Bariatric Procedure	3	3	3	3	3	3	2	2	3	2	2	3
Breast Cancer	1	2	3	4	4	2	2	2	3	3	3	3
C-Section	2	2	3	3	3	3	2	0	1	3	3	1
CABG	4	4	2	2	2	1	3	2	2	3	3	2
Cellulitis	3	3	3	3	3	2	3	3	4	4	3	2
Colon Cancer	3	3	3	4	4	2	2	2	2	3	4	3
Decubitus Ulcer	0	0	1	3	3	4	2	2	3	3	4	3
Dementia	3	2	1	4	4	1	3	3	1	2	2	3
Depression	1	1	1	3	3	0	0	0	2	3	3	1
Diabetes	3	3	2	3	3	3	3	3	4	3	3	2
Hip Replacement	2	2	1	4	4	2	2	2	4	2	2	1
Iatrogenic Pneumothorax (PSI)	2	2	2	2	2	1	0	0	2	0	0	0
Ischemic Stroke	3	3	2	3	3	3	4	4	2	2	2	4
Knee Replacement	4	4	1	4	4	2	3	3	3	3	3	1
Lung Cancer	2	1	3	3	3	2	2	2	2	1	3	3
MRSA	2	2	2	3	3	2	4	4	3	4	4	3
PTCA	2	2	2	3	3	2	2	1	1	4	4	2
Peptic Ulcer	2	2	2	3	3	3	3	3	2	4	4	2
Pneumonia	2	3	2	2	2	3	4	4	2	4	4	4
Postoperative Hemorrhage or Hematoma (PSI)	2	1	0	2	2	0	2	1	0	1	1	0
Postoperative Respiratory Failure (PSI)	0	1	2	1	1	1	1	2	1	0	0	1
Prostate Cancer	3	2	2	3	3	2	0	0	1	3	3	3
Septicemia	3	3	0	4	4	1	4	4	3	4	4	2
Spinal Fusion	0	2	3	4	4	1	2	3	2	3	3	2

Table 5. Comparison of stratified early estimates

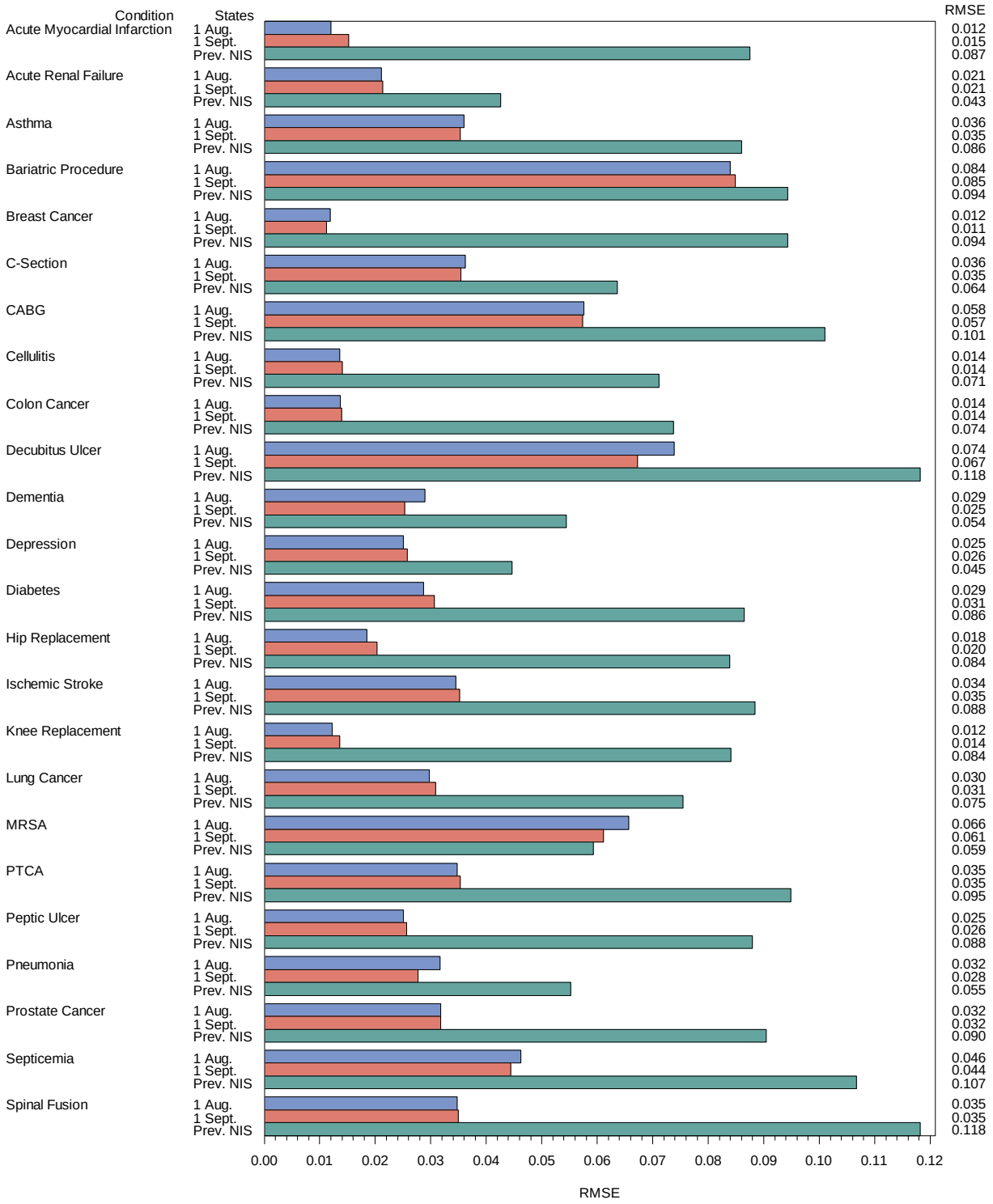
	Statistic											
	ALOS			Charges			Mortality			Discharges		
	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.	Aug. vs. NIS	Sept. vs. NIS	Sept. vs. Aug.
	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better	# Better
All	54 (52 %)	58 (56 %)	54 (52 %)	88 (85 %)	87 (84 %)	51 (49 %)	62 (57 %)	61 (59 %)	52 (50 %)	75 (72 %)	79 (76%)	51 (49 %)
Accidental Puncture or Laceration	1	1	0	2	2	1	1	1	1	0	1	1
Acute Myocardial Infarction	3	3	4	4	4	1	4	4	1	3	3	2
Acute Renal Failure	4	4	1	3	3	1	4	4	1	4	4	3
Asthma	0	1	2	4	4	3	1	1	2	4	4	2
Bariatric Procedure	3	3	3	3	3	2	1	1	3	1	2	2
Breast Cancer	1	1	1	4	4	2	3	3	1	3	3	2
C-Section	2	2	3	3	3	4	3	2	2	3	3	2
CABG	2	4	2	2	2	2	2	1	2	4	3	0
Cellulitis	3	3	2	4	3	2	3	4	2	3	3	3
Colon Cancer	3	3	2	4	4	1	2	2	3	4	4	3
Decubitus Ulcer	0	0	2	3	3	4	2	3	3	3	4	1
Dementia	1	1	2	4	4	4	3	3	1	3	4	2
Depression	1	1	2	3	3	2	0	0	3	3	3	2
Diabetes	3	3	2	3	3	1	4	4	3	2	3	2
Hip Replacement	2	2	2	4	4	0	2	2	3	2	2	0
Iatrogenic Pneumothorax	2	2	0	2	2	0	0	0	0	1	0	0
Ischemic Stroke	3	3	3	3	3	2	4	4	1	2	3	2
Knee Replacement	4	4	1	4	4	1	3	2	1	3	3	2
Lung Cancer	1	1	3	3	3	1	2	2	1	2	2	4
MRSA	2	2	3	3	3	2	4	4	1	3	3	0
PTCA	2	2	2	3	3	2	0	0	1	3	3	1
Peptic Ulcer	2	2	2	3	3	1	2	2	4	4	4	2
Pneumonia	2	3	4	2	2	4	4	4	3	4	4	2
Postoperative Hemorrhage or Hematoma	1	1	0	2	2	0	1	1	2	1	1	1
Postoperative Respiratory Failure	0	0	2	1	1	1	2	2	1	0	0	2
Prostate Cancer	2	2	2	4	4	1	0	0	2	3	3	2
Septicemia	3	3	2	4	4	4	4	4	3	4	4	3
Spinal Fusion	1	1	0	4	4	2	1	1	1	3	3	3

Figure 3. RMSE for ALOS, stratified estimates vs. previous NIS estimate



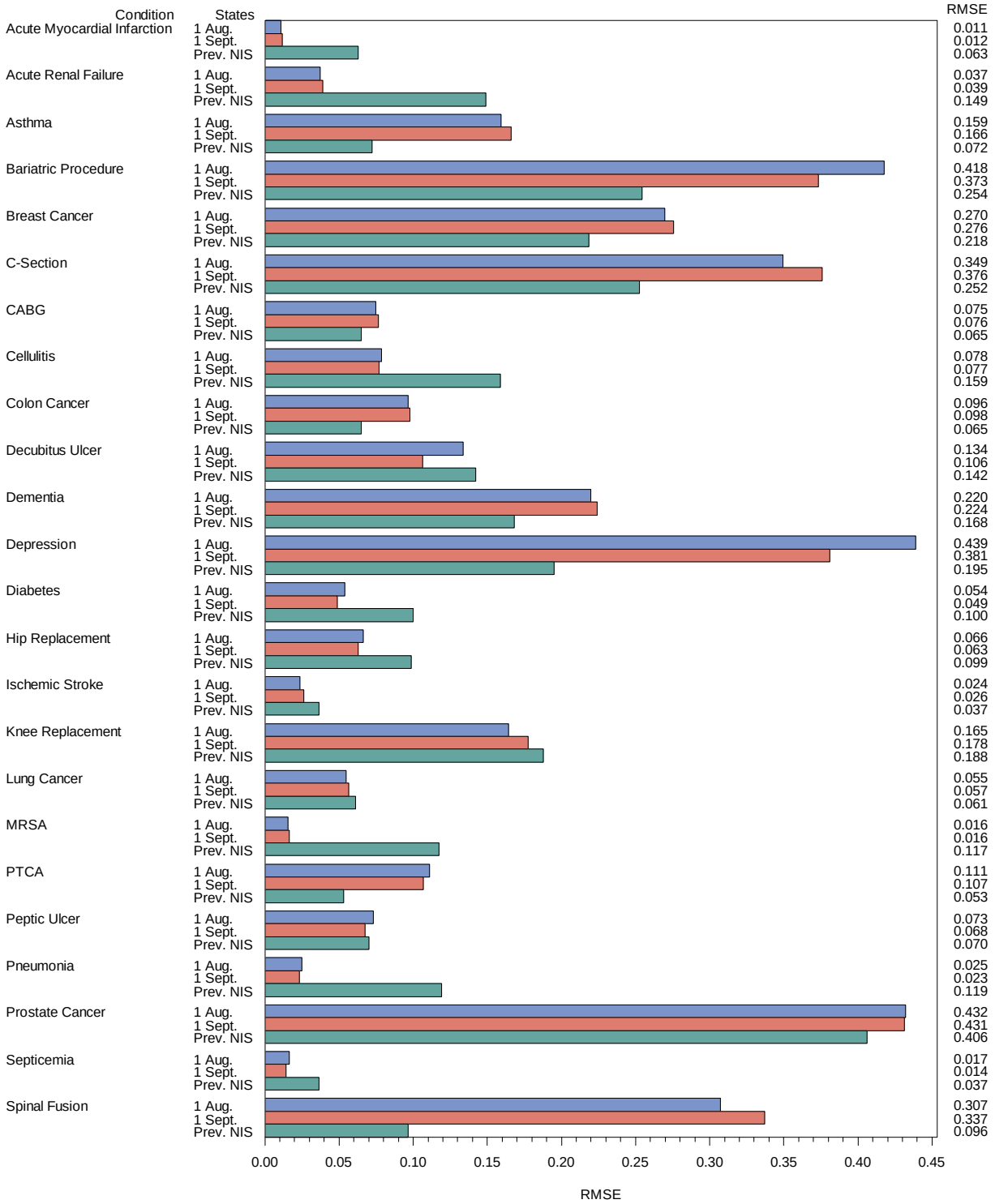
Note: Smaller RMSE is Better

Figure 4. RMSE for Mean Charges, stratified estimates vs. previous NIS estimate



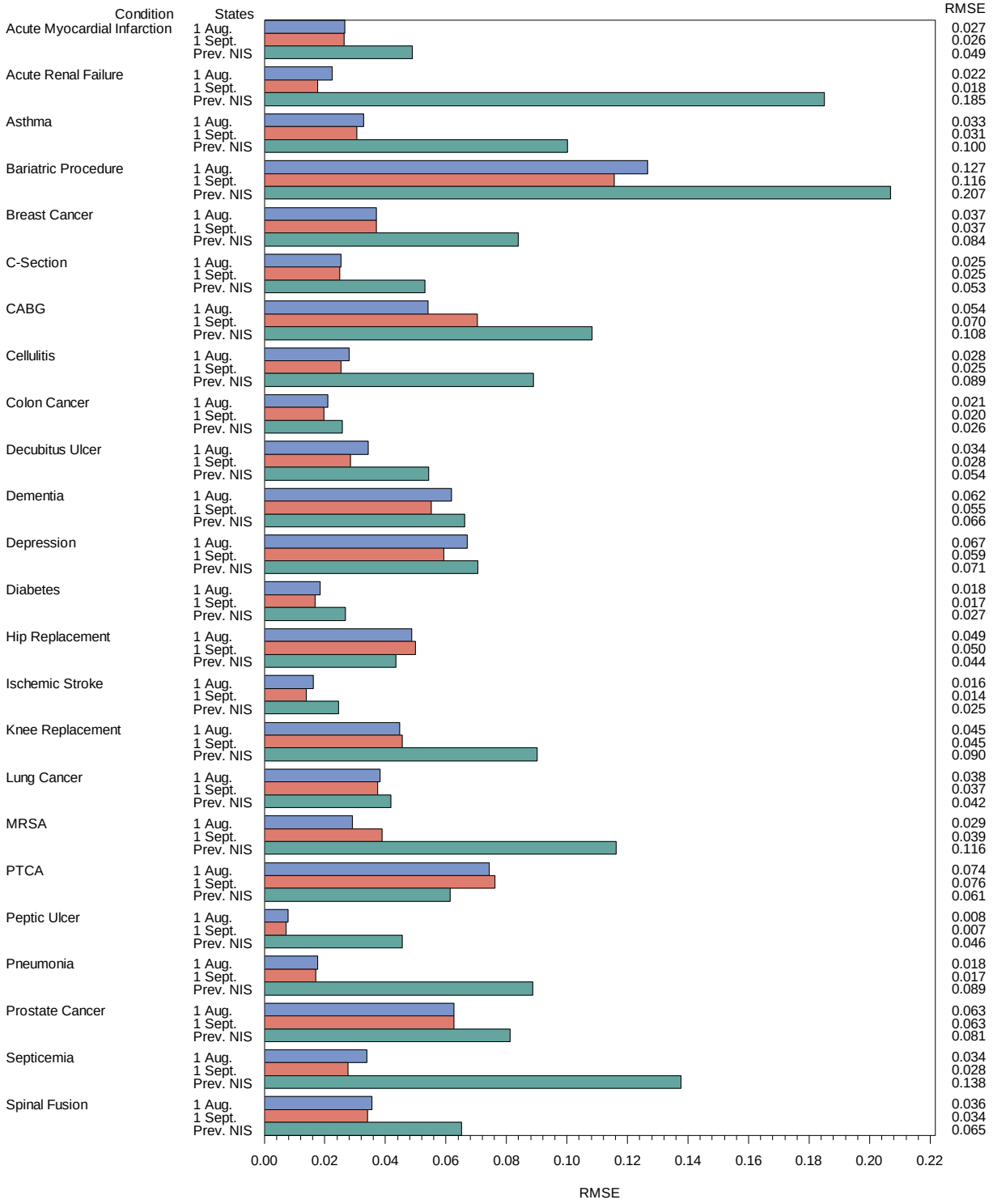
Note: Smaller RMSE is Better

Figure 5. RMSE for Mortality Rates, stratified estimates vs. previous NIS estimate



Note: Smaller RMSE is Better

Figure 6. RMSE for Total Discharges, stratified estimates vs. previous NIS estimate



Note: Smaller RMSE is Better

Table 6. ALOS correlations between state and NIS estimates, 2002 – 2006

	State															
	AZ	CA	CO	IA	IL	MD	MN	MO	NJ	NV	OH	OR	SD	WA	WI	WV
Acute Myocardial Infarction	-0.867	0.999	0.750	0.871	0.943	0.958	0.856	0.983	0.933	0.990	0.989	-0.703	0.742	0.886	0.997	-0.025
Acute Renal Failure	0.926	0.987	0.623	0.924	0.999	0.945	0.942	0.929	0.989	0.853	0.989	0.901	0.793	0.961	0.948	0.992
Asthma	0.635	0.694	0.560	0.800	0.694	0.500	-0.611	0.890	0.608	-0.622	0.811	0.239	0.925	-0.781	-0.869	0.988
Bariatric Procedure	-0.314	0.886	0.969	0.873	-0.582	0.896	0.975	0.919	0.912	0.944	0.015	0.813	0.188	0.404	0.937	0.283
Breast Cancer	0.602	0.257	0.873	-0.041	0.682	-0.024	0.895	0.330	-0.286	0.198	0.330	-0.407	-0.067	-0.765	0.681	-0.428
C-Section	0.771	0.806	0.809	-0.839	0.755	0.943	0.829	0.953	0.908	0.694	0.907	0.882	0.629	0.746	0.905	0.917
CABG	0.984	0.939	-0.011	-0.498	0.860	0.909	-0.266	0.746	0.943	0.981	0.890	0.605	-0.620	0.958	-0.976	0.965
Cellulitis	0.533	0.973	-0.203	0.894	0.906	0.920	0.953	0.944	0.957	0.694	0.972	0.935	-0.775	0.932	0.952	0.627
Colon Cancer	0.521	0.933	0.309	0.747	0.917	0.749	0.684	0.539	0.816	-0.490	0.593	0.531	0.907	0.468	0.915	0.317
Decubitus Ulcer	0.887	-0.555	-0.293	-0.242	-0.614	-0.177	-0.952	0.059	-0.701	-0.536	-0.762	0.262	0.228	-0.964	-0.371	0.091
Dementia	0.156	0.138	0.355	0.104	0.312	0.373	0.519	-0.140	-0.058	-0.009	-0.044	0.787	-0.074	-0.109	-0.366	0.673
Depression	-0.734	0.706	-0.191	-0.056	0.104	0.691	0.189	0.468	0.804	-0.429	0.660	-0.587	0.885	-0.176	0.799	-0.435
Diabetes	-0.980	0.781	0.877	0.878	0.710	0.934	0.690	0.881	0.901	0.856	0.838	-0.661	-0.745	0.889	0.814	0.979
Hip Replacement	0.956	0.994	0.953	0.929	0.942	0.988	0.912	0.969	0.994	0.224	0.804	0.913	0.602	0.944	0.978	0.960
Ischemic Stroke	-0.766	0.970	0.090	0.923	0.930	0.970	0.899	0.927	0.979	0.072	0.895	-0.959	0.438	0.724	0.920	0.907
Knee Replacement	0.989	0.986	0.937	0.976	0.994	0.599	0.965	0.935	0.863	-0.125	0.880	0.849	0.848	0.993	0.989	0.977
Lung Cancer	-0.537	0.707	0.101	0.850	0.645	0.873	0.262	0.296	0.771	0.599	0.541	-0.078	0.706	0.660	0.824	-0.278
MRSA	0.722	0.977	0.655	0.656	0.946	0.976	0.924	0.121	0.686	0.869	0.757	0.293	0.930	0.961	0.947	0.874
PTCA	-0.810	0.877	0.453	0.778	0.793	0.823	0.626	0.874	0.171	0.441	0.705	0.661	0.647	0.801	0.791	0.911
Peptic Ulcer	-0.430	0.834	0.481	-0.516	0.089	0.593	0.718	0.038	0.072	0.863	0.701	0.025	-0.082	0.809	0.923	0.412
Pneumonia	0.125	0.695	0.875	0.893	0.988	0.919	0.961	0.930	0.877	0.927	0.963	0.612	0.835	0.763	0.900	0.933
Prostate Cancer	0.927	0.702	0.946	0.836	0.831	0.778	0.790	0.692	0.943	0.658	0.982	0.844	-0.096	0.861	0.718	0.237
Septicemia	0.927	0.890	0.634	0.767	0.962	0.951	0.843	0.564	-0.367	0.911	0.945	0.903	0.817	0.914	0.346	0.560
Spinal Fusion	0.246	-0.733	-0.129	-0.767	-0.422	-0.280	-0.451	-0.524	-0.149	0.292	0.361	0.465	-0.101	-0.208	-0.347	-0.909

Table 7. Mean charge correlations between state and NIS estimates, 2002 – 2006

	State															
	AZ	CA	CO	IA	IL	MD	MN	MO	NJ	NV	OH	OR	SD	WA	WI	WV
Acute Myocardial Infarction	0.992	0.997	0.990	0.974	0.986	0.974	0.990	0.992	0.973	0.964	0.993	0.985	0.970	0.989	0.993	0.985
Acute Renal Failure	0.981	0.968	0.901	0.832	0.961	0.501	0.964	0.779	0.463	0.609	0.985	0.819	0.849	0.966	0.802	0.973
Asthma	0.990	0.965	0.944	0.991	0.962	0.972	0.999	0.963	0.937	0.996	0.953	0.945	0.981	0.989	0.971	0.983
Bariatric Procedure	0.845	0.887	0.915	0.783	0.828	0.795	0.809	-0.409	0.758	-0.744	0.812	0.652	0.733	0.865	0.635	0.772
Breast Cancer	0.996	0.998	0.999	0.979	0.994	0.988	0.988	0.984	0.988	0.970	0.999	0.991	0.970	0.993	0.999	0.991
C-Section	0.959	0.978	0.970	0.947	0.921	0.964	0.967	0.969	0.984	0.908	0.946	0.975	0.963	0.958	0.965	0.929
CABG	0.966	0.969	0.948	0.974	0.975	0.941	0.961	0.964	0.985	0.984	0.962	0.978	0.986	0.966	0.944	0.981
Cellulitis	0.978	0.992	0.982	0.965	0.995	0.940	0.949	0.987	0.868	0.956	0.985	0.989	0.989	0.975	0.993	0.908
Colon Cancer	0.990	0.993	0.997	0.956	0.994	0.978	0.920	0.999	0.987	0.879	0.997	0.965	0.999	0.980	0.987	0.966
Decubitus Ulcer	0.934	0.989	0.736	0.385	0.837	0.784	0.215	0.934	0.846	0.430	0.911	0.903	0.839	0.757	0.898	0.913
Dementia	0.966	0.962	0.993	0.831	0.991	0.994	0.942	0.975	0.637	0.252	0.983	0.963	0.875	0.957	0.878	0.979
Depression	0.998	0.996	0.994	0.258	0.983	0.972	0.985	0.990	0.937	0.928	0.980	0.945	0.822	0.984	0.978	0.963
Diabetes	0.972	0.990	0.966	0.966	0.991	0.932	0.929	0.981	0.956	0.626	0.959	0.972	0.945	0.947	0.927	0.944
Hip Replacement	0.981	0.998	0.993	0.983	0.992	0.965	0.988	0.991	0.992	0.974	0.993	0.998	0.963	0.986	0.989	0.997
Ischemic Stroke	0.959	0.983	0.960	0.967	0.983	0.949	0.985	0.976	0.963	0.966	0.976	0.968	0.937	0.974	0.974	0.949
Knee Replacement	0.993	0.995	0.999	0.993	0.990	0.986	0.999	0.997	1.000	0.962	0.994	0.997	0.993	0.992	0.996	0.975
Lung Cancer	0.992	0.974	0.990	0.986	0.981	0.885	0.969	0.988	0.880	0.908	0.988	0.995	0.982	0.977	0.983	0.949
MRSA	0.840	0.899	0.617	0.506	0.901	0.750	0.890	0.956	0.813	-0.267	0.976	0.784	-0.554	0.941	0.956	0.778
PTCA	0.992	0.987	0.968	0.969	0.988	0.971	0.997	0.973	0.977	0.958	0.983	0.992	0.762	0.989	0.989	0.996
Peptic Ulcer	0.953	0.991	0.993	0.942	0.971	0.982	0.969	0.978	0.987	0.851	0.990	0.947	0.901	0.994	0.987	0.970
Pneumonia	0.966	0.936	0.971	0.978	0.998	0.893	0.972	0.989	0.935	0.781	0.998	0.986	0.897	0.969	0.961	0.995
Prostate Cancer	0.982	0.993	0.981	0.990	0.977	0.984	0.984	0.988	0.946	0.951	0.992	0.981	0.972	0.957	0.987	0.896
Septicemia	0.986	0.987	0.976	0.997	0.991	0.991	0.965	0.974	0.939	0.983	0.989	0.979	0.997	0.992	0.984	0.993
Spinal Fusion	0.992	0.996	0.990	0.977	0.982	0.993	0.987	0.980	0.995	0.990	0.991	0.989	0.991	0.964	0.982	0.965

Table 8. Mortality rate correlations between state and NIS estimates, 2002 – 2006

	State															
	AZ	CA	CO	IA	IL	MD	MN	MO	NJ	NV	OH	OR	SD	WA	WI	WV
Acute Myocardial Infarction	0.817	0.999	0.869	0.563	0.959	0.952	0.746	0.969	0.988	0.897	0.987	0.824	0.346	0.859	0.957	0.832
Acute Renal Failure	0.985	0.991	0.971	0.884	0.982	0.894	0.995	0.981	0.995	0.975	0.999	0.991	0.696	0.906	0.979	0.913
Asthma	0.270	-0.411	-0.268	-0.546	-0.189	-0.733	0.340	-0.531	-0.878	-0.426	0.482	-0.135	0.689	-0.624	0.214	-0.928
Bariatric Procedure	-0.264	-0.072	0.366	-0.040	0.345	-0.140	0.559	0.652	0.550	-0.364	-0.816	-0.126	0.486	0.208	0.949	-0.170
Breast Cancer	0.881	0.629	0.521	0.546	0.501	0.024	0.657	0.500	0.101	0.278	-0.475	-0.174	-0.224	-0.445	0.319	0.255
C-Section	-0.784	0.807	0.587	0.396	0.622	0.567	-0.087	-0.317	-0.070	0.900	0.585	-0.866	-0.659	-0.190	-0.063	0.636
CABG	0.423	0.747	0.886	0.765	0.798	0.322	0.086	0.261	0.075	-0.226	0.913	0.367	0.880	0.920	0.672	-0.658
Cellulitis	0.796	0.937	0.445	0.776	0.943	0.972	0.529	0.880	0.950	0.974	0.955	0.640	0.416	0.930	0.772	0.560
Colon Cancer	0.845	-0.976	-0.809	-0.063	-0.029	0.888	0.013	-0.659	-0.564	0.443	0.477	-0.231	-0.407	0.533	-0.168	0.265
Decubitus Ulcer	0.091	0.484	0.194	0.276	-0.089	0.828	0.540	0.809	-0.594	-0.554	0.544	0.137	-0.414	-0.482	0.169	0.619
Dementia	0.444	0.594	0.488	0.549	0.583	0.678	-0.013	0.057	0.378	-0.161	0.668	-0.121	-0.299	0.454	-0.120	-0.420
Depression	-0.075	0.576	-0.174	-0.971	0.564	0.727	-0.513	0.495	0.947	0.002	0.154	-0.559	0.076	-0.047	0.186	0.598
Diabetes	0.854	0.931	0.876	0.039	0.940	0.936	0.543	0.882	0.983	0.956	0.979	0.920	-0.053	0.934	0.966	0.842
Hip Replacement	0.522	0.902	0.799	0.302	0.759	-0.042	0.674	0.981	0.698	0.823	0.878	0.960	-0.147	0.539	0.955	0.986
Ischemic Stroke	0.778	0.995	0.643	0.353	0.782	-0.011	0.943	0.432	0.886	0.366	0.918	0.958	-0.219	0.349	0.930	-0.117
Knee Replacement	0.909	0.839	0.666	-0.525	0.595	0.438	0.827	0.480	0.986	0.262	-0.152	-0.705	0.729	0.600	-0.349	0.378
Lung Cancer	0.879	0.624	0.742	-0.583	0.571	0.190	-0.567	0.046	0.927	0.529	0.787	0.786	-0.221	0.488	0.739	0.688
MRSA	0.949	0.979	0.965	0.778	0.917	0.934	0.677	0.787	0.890	0.422	0.900	0.789	0.878	0.968	0.851	0.614
PTCA	0.340	-0.222	0.482	0.910	0.492	0.002	0.781	-0.462	-0.269	-0.045	0.256	0.040	-0.371	-0.306	0.504	-0.427
Peptic Ulcer	-0.467	0.638	0.491	-0.397	0.514	0.708	-0.012	0.688	0.601	0.352	0.924	-0.162	0.343	0.458	0.409	0.086
Pneumonia	0.980	0.988	0.964	0.952	0.982	0.980	0.998	0.976	0.964	0.944	0.990	0.982	0.026	0.981	0.976	0.970
Prostate Cancer	-0.288	-0.619	-0.393	0.221	0.389	-0.804	-0.171	-0.677	0.861	0.033	0.083	0.424	0.325	0.893	0.722	0.455
Septicemia	0.142	0.509	0.493	0.898	0.779	0.978	0.225	0.556	0.647	0.540	-0.059	0.488	0.685	0.724	0.626	0.420
Spinal Fusion	0.083	0.625	-0.543	0.817	0.895	-0.610	0.780	-0.201	0.373	-0.352	0.464	-0.101	0.088	0.227	-0.150	0.034