**Title**
Comparing Resource Use in Medical Admissions of Children With Complex Chronic Conditions



**Journal**
Medical Care, 57(8)

**ISSN**
0025-7079

**Authors**
Silber, Jeffrey H
Rosenbaum, Paul R
Pimentel, Samuel D
et al.

**Publication Date**
2019-08-01

**DOI**
10.1097/mlr.0000000000001149

Peer reviewed

HHS Public Access
Author manuscript
*Med Care*. Author manuscript; available in PMC 2020 August 01.

Published in final edited form as:
*Med Care*. 2019 August ; 57(8): 615–624. doi:10.1097/MLR.0000000000001149.

# Comparing Resource Use in Medical Admissions of Children with Complex Chronic Conditions

**Jeffrey H. Silber, MD, PhD**[1,2,3,4,5], **Paul R. Rosenbaum, PhD**[5,6], **Samuel D. Pimentel, PhD**[7], **Shawna Calhoun, MPH**[1], **Wei Wang, PhD**[1], **James E. Sharpe, MS**[1], **Joseph G. Reiter, MS**[1], **Shivani A. Shah, BA**[1], **Lauren L. Hochman, BA**[1], **Orit Even-Shoshan, MS**[1]

[1]Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, PA

[2]Departments of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

[3]Department of Anesthesiology and Critical Care, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

[4]Department of Health Care Management, The Wharton School, University of Pennsylvania, Philadelphia, PA

[5]The Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA

[6]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA

[7]Department of Statistics, University of California, Berkeley, CA

## Abstract

**Background:** Children with Complex Chronic Conditions (CCCs) utilize a disproportionate share of hospital resources.

**Objectives:** We asked whether some hospitals display a significantly different pattern of resource utilization than others when caring for similar children with CCCs admitted for medical diagnoses.

**Research Design:** Using Pediatric Health Information System data from 2009–2013, we constructed an inpatient Template of 300 children with CCCs, matching these to 300 patients at each hospital, thereby performing a type of direct standardization.

**Subjects:** Children with CCCs were drawn from a list of the 40 most common medical principal diagnoses, then matched to patients across 40 Children's Hospitals.

**Measures:** Rate of ICU admission, length of stay, resource cost.

**Results:** For the Template-matched patients, when comparing resource-use at the lower 12.5%-ile and upper 87.5%-ile of hospitals, we found: ICU utilization was 111% higher (6.6% vs. 13.9%, $P < 0.001$); hospital length of stay was 25% higher (2.4 vs. 3.0 days per admission, $P < 0.001$); and

**Address Correspondence to:** Dr. Jeffrey H. Silber, Center for Outcomes Research, Children's Hospital of Philadelphia, Roberts Building, 2716 South Street, Room 5123, Philadelphia, PA 19146-2305, (T) 215-590-5635, (F) 215-590-2378, silber@email.chop.edu.

**Conflict of Interest**: The authors have no conflicts of interest to disclose.

finally, total cost per patient varied by 47% ($6,856 vs. $10,047, P < 0.001). Furthermore, some hospitals, compared to their peers, were more efficient with low risk patients and less efficient with high risk patients, while other hospitals displayed the opposite pattern.

**Conclusions:** Hospitals treating similar patients with CCCs admitted for similar medical diagnoses, varied greatly in resource utilization. Template Matching can aid chief quality officers benchmarking their hospitals to peer institutions, and can help determine types of their patients having the most aberrant outcomes, facilitating quality initiatives to target these patients.

## BRIEF SUMMARY:

Resource utilization greatly varies in Template matched children with Complex Chronic Conditions admitted across Children's Hospitals

### Keywords

Direct Standardization; Hospital Variation; Resource Utilization

## INTRODUCTION

Rates of hospitalizations among children with Complex Chronic Conditions (CCCs)[1] have been rising[2–7] and these patients account for a disproportionate share of expenditures at Children's Hospitals.[2] In 2006, such children accounted for only 10% of admissions at Children's Hospitals, yet represented 25% of pediatric hospital days and 40% of charges at those hospitals,[2] and have been estimated to account for one third of all childhood medical spending.[8] Medically complex children average 1.5 hospitalizations per year,[3] and children with CCCs have been found to have inpatient expenses 17.3 times those of healthy children.[9] Furthermore, children with medical complexity (CMC) have been shown to have increased length of stay and expenditures.[10–12] Assessing the differences in resource use between hospitals on their CCC patients can be challenging due to differences in case-mix between hospitals. When studies of resource variation across hospitals treating CCC or CMC patients use case-mix adjustment through regression, this does not assure that CCC or CMC patients across hospitals are similar. For example, in a recent study of asthma across Children's Hospitals, we observed that a random sample of asthma patients across hospitals produced great variation in patient characteristics, which would need to be adjusted for in order to compare quality of care.[13] Therefore, when observing variation in resource use or outcomes across hospitals one does not know whether to attribute the variation to differences in case-mix or to hospital practice style.

We developed "Template Matching"[13–16] to more transparently compare resource utilization across hospitals in this diverse population of CCC children encompassing many different acute diagnoses coupled with many different patient comorbidities. Template Matching, a form of direct standardization,[14,17] allows for a close comparison of similar patients across hospitals by selecting a reference population of patients (the "Template") and finding patients similar to the Template at each hospital using multivariate matching.[14,15,18,19] In so doing, we aim to better compare variation in practice style through the assessment of resource use for children with CCCs admitted to medical services in Children's Hospitals to facilitate quality of care initiatives for this vulnerable and complex population. In a previous study,[13] we examined variation in resource utilization in children admitted for one disease, asthma, at Children's Hospitals through the use of Template Matching. We now examine variation in resource utilization by hospital for children with medically diverse CCCs admitted for 40 different principal diagnoses.

## METHODS

This study was approved by the Children's Hospital of Philadelphia Institutional Review Board.

### Patient Population and Definitions

Data for this study were obtained from the Pediatric Health Information System (PHIS), an administrative database containing inpatient, emergency department, ambulatory surgery and observation unit data from 45 not-for-profit, tertiary care pediatric hospitals in the United States. These hospitals are affiliated with the Children's Hospital Association (Overland Park, KS). The PHIS hospitals are 45 of the largest and most advanced children's hospitals in America, and constitute the most demanding standards of pediatric service in America.

We examined all patients with Complex Chronic Conditions (CCC), and studied those admitted for 40 of the most frequent medical conditions as found in the PHIS database between October 1, 2009 and June 30, 2013. Encounters were classified as medical as defined by MS-DRG version 25.[20] The diagnosis codes were grouped using the ICD-9 clinical classification definitions. Of the 45 PHIS hospitals, four had incomplete data needed for this study and therefore were excluded, leaving 41 available for matching.

All non-transfer inpatient and non-research observational unit were considered if the patient was between 1 and 18 years old and had a positive indicator for a complex chronic condition. The CCC definition was based on Feudtner et al.[1] with some modifications (see Table 1 and Supplemental Tables 1 and 2). Hospitals were included in the analysis if they had a minimum of 350 CCC patients for the match.

### Statistical Methods

**Constructing the Template—**The Template consisted of 300 patients, and served to create strata of similar patients across the study hospitals. Various power calculations described in Supplemental Section 1 showed this sample size had sufficient power to detect important differences across hospitals.[14,15,21,22] The desired properties of the Template reference population were that it represent study conditions at the rates they appear in the overall study population, that it include events that occur at a 1 in 300 rate, but that it exclude very rare events (like a 1 in 2,000 rate). Therefore, we created 2,000 random samples of 300 patients from the PHIS data set of study eligible patients and chose the best random sample to be the Template. The best random sample was defined as the single random sample of 300 patients which had the smallest Mahalanobis distance to the patients admitted for a top 40 principal diagnosis.

**Matching Methodology—**Variables considered for matching to the Template included age in years, sex, broad CCC type, CCC-sub type, principal diagnosis, Medicaid as the patient's primary insurance, and additional clinical characteristics. See Supplemental Section 1 for the full list of matched variables. Table 1 describes these CCC group definitions and also provides the ICD-9 principal diagnosis utilized for medical admissions. We used the first medical admission in the data set for each patient. When possible, we matched patients exactly on principal diagnosis. If this was not possible, we attempted to match patients with a principal diagnosis that was clinically similar utilizing the digit groupings of the ICD-9 codes. We describe this process in detail in Supplemental Table 2 and give a list of "backup" ICD-9 codes used for four rare conditions in Supplemental Table 3. Supplemental Figure 1A displays this near exact hierarchy for matching principal diagnoses.

We performed our matches in R (R Development Core Team 2013) using Pimentel's *rcbalance* package.[23] We matched to minimize the Mahalanobis distance[14,15,24] between individual matched patients, subject to various constraints that force balance on covariates and interactions in matched groups. Details concerning the elements of this medical distance are provided in the Supplemental Sections 1 and 2.

To improve the quality of the matches between the Template and the specific hospital, we used "fine balance"[25–29] within hospitals (see Supplemental Section 1). We further required the matches to achieve "refined covariate balance" on a hierarchy of nominal covariates given in decreasing order of importance.[23] This type of balance requires the first covariate in the hierarchy to satisfy optimal "near fine balance", the second to satisfy optimal near fine balance subject to the balance constraint of the first variable, and so on down the hierarchy. An illustration of the process of Template formation and matching is provided in Figure 1. The full hierarchy of variables we used to perform the match is outlined in Supplemental Figures 1A and 1B.

**Assessing the Quality of the Match—**We assessed the quality of the match at each hospital by comparing the matched samples to each other using statistical tests that take as a benchmark the covariate balance that would be produced by allocating patients at random to hospitals. We were satisfied with the control of measured covariates if they were as balanced, or better balanced, than they would be in a vast randomized trial. Unlike randomization, however, matching can only balance measured covariates and not unmeasured covariates. For continuous variables we used the Kruskal-Wallis test, a non-parametric version of the one-way ANOVA test,[30] to assess whether the measured covariates were as balanced among the hospitals as they would be by random allocation. Similarly, the Pearson $\chi^2$ test was done for each binary variable in a 2xN hospitals table. The chi-square statistic for the Kruskal-Wallis and Pearson tests divided by its degrees of freedom, $\chi^2$/df, is also reported. Random assignment of patients to hospitals would lead these chi-squares divided by degrees of freedom $\chi^2$/df to have asymptotic expectation equal to 1. Values of $\chi^2$/df > 1 indicate worse balance, and values of $\chi^2$/df < 1 indicate better balance than expected by random assignment. Variables successfully controlled by matching have $\chi^2$/df values of 1 or less and an insignificant P-value.

In addition to checking the overall balance among all hospitals, we also checked the quality of the match one hospital at a time, one variable at a time, and one category of a variable at a time, emphasizing important variables. This process of checking the match quality required an enormous number of tests, where we would expect one P-value less than 0.05 per 20 tests even if patients had been allocated at random to hospitals. When looking at an enormous number of P-values for one hospital, we benchmarked the size of these P-values with reference to the Simes procedure[31] often used to control the false discovery rate. Additionally, we performed one overall test, the cross-match test[32,33] using all variables at once in effort to distinguish each hospital from the Template and identify imbalances in combinations of variables.

Individual hospitals were required to have no significant difference (no $P \leq 0.05$) on the most important (Level 1) variables comprised of CCC or principal diagnosis groupings, based on the Kruskal-Wallis and the Pearson $\chi^2$ tests. At refined levels beyond Level 1, with many more variables and categories, (125 p-values per hospital) the match quality was further tested using the Simes procedure as discussed above.[31] Finally, we performed a check on the balance between each hospital and the Template by examining total variation distance (the sum of the absolute Template minus index hospital differences in category percentages)[23] comparing each hospital match to 10,000 simulated randomized experiments. See Supplemental Section 3 for details. Supplemental Figure 2 illustrates the decision model for whether to include a hospital in the final analysis based on the quality of the match.

All matching was completed without knowledge of outcomes, as suggested by Rubin.[34,35]

**Outcome and Cost Variables—**Once hospital matches were complete and we were sure the matches found comparable patients, the Template was discarded and hospitals were then compared to each other on the following outcome variables: rate of ICU admission; hospital length of stay; and resource cost. The annual unit costs for each billing code were determined using methodology published previously,[13] a modification of Keren et al.[36] and details are provided in the Supplemental Section 1.

**Differentiating Overall Main Effect from Hospital-Patient Interactions (Risk Synergy)—**We first test whether each matched sample at a hospital had outcomes that differed significantly from the matched controls at other hospitals. For continuous variables, we used quantile tests[37,38] that determined whether each patient exceeded the median or 90$^{th}$ percentile value of all matched patients and used the Mantel-Haenszel statistic[17] to test the equality of each hospital with the others in exceeding this value. To adjust for multiple testing, we used the Bonferroni correction with the cutoff of $P < 0.05/k$, with $k$ indicating the number of tests (in this case, the number of hospitals being compared).

Specific advantage was defined as observing better patient outcomes in a specific hospital compared to the outcomes of matched control patients from other facilities.[13,39] Risk synergy described a situation during which, as patient risk at admission increases or decreases, the specific advantage changes in a systematic way.[13,39,40] For example, as patient risk increases, the focal hospital's patient may have increasingly better outcomes than matched controls from other hospitals. We defined risk using predicted length of stay from a model fit to PHIS patients not included in this study (Supplemental Table 4). The longer the predicted length of stay, the greater the risk. Each patient received a predicted length of stay based on the model that we developed, and each of the Template patients was assigned to a stratum and matched to patients of similar risk. At each hospital we tested for an interaction between the predicted-patient risk stratum and the main effect of the focal hospital. To obtain the assigned risk of each stratum, we computed the mean risk of the matched controls from all the hospitals. For binary outcomes we tested synergy using a conditional logistic regression evaluating the interaction for admission to the focal hospital and a linear term for average patient risk in the matched set, conditioning on the 300 patient strata in the logistic model. For continuous outcomes we tested for risk synergy by robust regression by fitting a linear model using m-estimation, as implemented in the rlm function in the MASS package in R with the default settings.[41–43] evaluating the interaction between an indicator for admission to the focal hospital and a linear term for average patient risk in the matched set (while also adjusting for the 300 patient strata in the Template). If either regression found a significant interaction, the hospital was considered to demonstrate risk synergy.

When testing if outcomes differed across all hospitals, we utilized tests that account for the pairing based on the Template. We used Cochran's Q test[44] for binary outcomes such as ICU utilization, and Friedman's Test[45] for continuous outcomes such as cost and length of stay.

## RESULTS

There were 735,591 patients in the PHIS data set at study eligible hospitals. Of these patients, 324,777 patients (or 44.2%) were defined as having a CCC, of which 208,575 (or 65.5%) were defined as medical patients per MS-DRG version 25. Of these, 148,625 (or 71.3%) had an admission for a study-relevant principal diagnosis and non-outlier costs. Excluding transfer patients left 133,040 patients eligible for the match. The cohort, Template sample, and Template are described in Supplemental Table 6.

### Quality of the Matches

All 41 hospitals met the minimum volume requirement for matching and 40 passed the match quality criteria. Table 2 (and Supplemental Table 7) describes the 40 hospitals in the dataset that were successfully matched to the Template and available for analysis. One hospital did not pass the Level 1 variable test (see Supplemental Figure 2). The table describes the variability of patient characteristics at the hospitals using the Template matched sample of 300 patients. All covariates describing the patients' initial health after matching were very similar across hospitals. For example, after matching, the distribution of chronic disease groups in each hospital was almost exactly the same, as was the distribution of principal diagnoses. Other patient characteristics, such as age and Medicaid status, were also similar across hospitals.

### Variation in Outcomes

The bottom of Table 2 describes the differences in patient outcomes across hospitals. While the Template match produced groups of patients with little variation in characteristics at admission, the patient outcomes differed greatly by hospital, and all outcomes displayed statistically significant variation. For the Template-matched population, comparing the lower eighth (12.5%-ile)[46] and upper eighth (87.5%-ile) of hospitals on each outcome, we found: ICU utilization was 111% higher (6.6% vs. 13.9%, p <0.001); patient length of stay was 25% higher (2.4 vs. 3.0 days per admission, p<0.001); finally, total cost per patient varied by 47% ($6,856 vs. $10,047, p<0.001). See Supplemental Table 8 for variation within subsets of principal diagnoses.

Having established that each hospital contributed patients very similar to the Template at the time of admission, Table 3 ranks in practice style across hospitals, using median and 90$^{th}$ percentile cost, 90$^{th}$ percentile length of stay, and ICU utilization rates (Supplemental Tables 9–11 include details). Hospital M is on par with the other hospitals for median cost (Hospital M Median $6,030 vs Overall $6,417) but has 13.5$^{th}$ smallest 90$^{th}$ percentile LOS and 2$^{nd}$ smallest ICU utilization, as well as the 36$^{th}$ smallest 90$^{th}$ percentile cost. Hospital B is a high-level performer across all outcomes, while Hospital MM does relatively poorly across all outcomes.

### Observing Comparative Advantage Risk Synergy

Figure 2 illustrates how hospitals differ in their ability to care for patients of varied levels of risk. The figure shows cost results for the smoothed plots of two hospitals (AA and O), and ICU utilization results for two hospitals (S and E) using locally weighted scatterplot smoothing (LOWESS).[47] Both hospitals AA and O had significant risk synergy for cost. Hospital AA displayed increasing cost as patients were admitted with more complex problems (as identified by a higher expected LOS based on conditions on admission). The opposite was true for Hospital O, where it did relatively better as initial patient risk increased. Hospitals S and E both also displayed significant risk synergy related to ICU utilization, again in opposite directions. Results for all hospitals for cost and ICU utilization can be found in Supplemental Table 12.

### Variation in Mortality

Deaths were rare in the Template, with 1 death in 300 (0.33%) and 23 deaths in the 12,000 (=300×40) matched patients or 0.19%. Twenty-four of the 40 hospitals had at least one death, and the highest mortality rate in any hospital was 2/300. Consequently, we did not report mortality as a primary outcome in this analysis, though in clinical settings, this may be of interest.

## DISCUSSION

We focused our study on the examination of children with complex chronic conditions because they are always of major concern to any healthcare system. Indeed, they are a population that many have focused on when aiming to reduce unnecessary hospital expenditures.[2,8,48] While it is well known that children with complex chronic conditions often require more ICU care, often stay longer and utilize a great deal of financial resources when hospitalized, it is not so clear that any hospital really knows how well they perform when compared to other hospitals. To understand how a hospital is performing relative to other hospitals, one needs some sort of benchmarking process. This is the first study that looks at a diverse group of children with CCCs using direct standardization, providing a unique view of resource consumption across Children's Hospitals for this population.

Template Matching allows a Chief Quality Officer (CQO) to directly compare the performance of their hospital to others that were also matched to the template reference population. The overall hospital main effect-their average performance over the 300 template matched patients relative to the performance of other hospitals-can be observed. If a hospital has a problem, the CQO can do more than just know there is a general problem. By observing outcome synergy, the CQO can now focus on patient groups along the risk continuum needing improvement.

Template Matching provides three important advantages over standard regression approaches when comparing hospital resource utilization: (1) the successfully matched template patients are very similar across hospitals, even when the case-mix across hospitals can be very different; (2) each patient in the template is mapped to corresponding patients across all hospitals being evaluated, allowing us to characterize in detail the patients being used for the comparison, and this knowledge can be used to better study the style of practice at hospitals for specific types of patients of interest; and (3) once matches are found for a hospital, analyses can be supplemented with actual detailed chart review of the specific patients in the template—something not possible with regression analyses. As data systems improve, and sharing across hospitals is expanded, the use of a template may provide increasing benefits to those seeking to improve patient care.

There are a number of limitations to Template Matching. First, the Template can be thought of as a standardized exam administered to each hospital. The exam questions are the patients in the Template, and the grades are determined by how each hospital's matched patients fared compared to the other hospitals in the analysis. However, some might argue that the standardized exam may not tap into specific strengths at various hospitals. One solution to this is to construct a hospital specific "boutique" template[15] that uses a single hospital of interest as the new reference population (say a hospital that did not perform well on the Template exam). The hospital's boutique Template can then be matched to the other hospitals and quality and outcomes can be compared. The advantage of this approach is that the new directly standardized exam reflects the patients seen at the hospital of interest, though possibly fewer hospitals will be able to match to the boutique Template.

Finally, a weakness in Template matching rests in data quality. Like any method using claims data, poor coding or inconsistency in coding may lead to bias, just as in regression approaches. We also lacked detailed information on socioeconomic status. While we did match on Medicaid status, we did not have information such as family income or occupation, which have been shown to be predictive of childhood medical resource use.[49,50]

## CONCLUSION

We observe large variation across Children's Hospitals in ICU usage as well as length of stay and cost when treating very similar patients with complex chronic disease. Using

Template Matching allows caregivers to examine their utilization with respect to other hospitals in a way that **transparently** compares similar patients, and allows for identification of differential patterns of practice compared to other hospitals in the data set. Through the knowledge gained about each individual patient in the hospital's own template-matched sample, and the matched controls at the other hospitals, differences may become apparent hospital-wide or over select patient groups. Using Template Matching not only informs a hospital about its overall treatment style and intensity, but it makes transparent where outcomes deviate from the norm, and allows closer examination for causes of these deviations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Funding Source:

AHRQ U18HS020508 and DOD FA9550–11-C-0028; Jeffrey H. Silber, MD, PhD, Principal Investigator

## REFERENCES

1. Feudtner C, Hays RM, Haynes G, et al. Deaths attributed to pediatric complex chronic conditions: National trends and implications for supportive care services. Pediatrics. 2001;107:E99. [PubMed: 11389297]
2. Simon TD, Berry J, Feudtner C, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. Pediatrics. 2010;126:647–655. [PubMed: 20855394]
3. Berry JG, Agrawal RK, Cohen E, et al. The landscape of medical care for children with medical complexity. 6 2013 Available at: https://www.childrenshospitals.org/issues-and-advocacy/children-with-medical-complexity/issue-briefs-and-reports/the-landscape-of-medical-care-for-children-with-medical-complexity. Accessed: August 17, 2016.
4. Berry JG, Hall M, Hall DE, et al. Inpatient growth and resource use in 28 children's hospitals: A longitudinal, multi-institutional study. JAMA Pediatr. 2013;167:170–177. [PubMed: 23266509]
5. O'Mahony L, O'Mahony DS, Simon TD, et al. Medical complexity and pediatric emergency department and inpatient utilization. Pediatrics. 2013;131:e559–565. [PubMed: 23319525]
6. Berry JG, Hall DE, Kuo DZ, et al. Hospital utilization and characteristics of patients experiencing recurrent readmissions within children's hospitals. JAMA. 2011;305:682–690. [PubMed: 21325184]
7. Berry JG, Toomey SL, Zaslavsky AM, et al. Pediatric readmission prevalence and variability across hospitals. JAMA. 2013;309:372–380. [PubMed: 23340639]
8. Cohen E, Berry JG, Camacho X, et al. Patterns and costs of health care use of children with medical complexity. Pediatrics. 2012;130:e1463–1470. [PubMed: 23184117]
9. Kuo DZ, Melguizo-Castro M, Goudie A, et al. Variation in child health care utilization by medical complexity. Matern Child Health J. 2015;19:40–48. [PubMed: 24740726]
10. Simon TD, Cawthon ML, Stanford S, et al. Pediatric medical complexity algorithm: A new method to stratify children by medical complexity. Pediatrics. 2014;133:e1647–1654. [PubMed: 24819580]
11. Agrawal R, Hall M, Cohen E, et al. Trends in health care spending for children in Medicaid with high resource use. Pediatrics. 2016;138. [PubMed: 27544347]
12. Gold JM, Hall M, Shah SS, et al. Long length of hospital stay in children with medical complexity. J Hosp Med. 2016;11:750–756. [PubMed: 27378587]
13. Silber JH, Rosenbaum PR, Wang W, et al. Auditing practice style variation in pediatric inpatient asthma care. JAMA Pediatr. 2016;170:878–886. [PubMed: 27398908]

14. Silber JH, Rosenbaum PR, Ross RN, et al. Template matching for auditing hospital cost and quality. Health Serv Res. 2014;49:1446–1474. [PubMed: 24588413]

15. Silber JH, Rosenbaum PR, Ross RN, et al. A hospital-specific template for benchmarking its cost and quality. Health Serv Res. 2014;49:1475–1497. [PubMed: 25201167]

16. Berry JG. This is my hospital's performance on quality of care-no excuses. JAMA Pediatr. 2016;170:837–838. [PubMed: 27400256]

17. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. 3rd ed. New York: John Wiley & Sons; 2003.

18. Silber JH, Rosenbaum PR, Trudeau ME, et al. Multivariate matching and bias reduction in the surgical outcomes study. Med Care. 2001;39:1048–1064. [PubMed: 11567168]

19. Rosenbaum PR. Part II: Matching. In: Design of Observational Studies. New York: Springer; 2010:153–253.

20. Centers for Medicare & Medicaid Services. Details for title: Files for FY 2008 Final Rule and Correction Notice. Available at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download-Items/CMS1247844.html. Accessed: February 27, 2019.

21. Miettinen OS. Individual matching with multiple controls in the case of all-or-none responses. Biometrics. 1969;25:339–355. [PubMed: 5794104]

22. Fleiss JL, Levin B, Paik MC. Chapter 4. Determining Sample Sizes Needed to Detect a Difference between Two Proportions: Unequal Sample Sizes In: Statistical Methods for Rates and Proportions. 3rd ed New York: John Wiley & Sons; 2003:75–76.

23. Pimentel SD, Kelz RR, Silber JH, et al. Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. J Am Stat Assoc. 2015;110:515–527. [PubMed: 26273117]

24. Mahalanobis PC. On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India. 1936;2:49–55.

25. Rosenbaum PR. Chapter 10: Fine Balance In: Design of Observational Studies. New York: Springer; 2010:197–206.

26. Rosenbaum PR, Ross RN, Silber JH. Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. J Am Stat Assoc. 2007;102:75–83.

27. Silber JH, Rosenbaum PR, Polsky D, et al. Does ovarian cancer treatment and survival differ by the specialty providing chemotherapy? J Clin Oncol. 2007;25:1169–1175. [PubMed: 17401005]

28. Silber JH, Rosenbaum PR, Kelz RR, et al. Medical and financial risks associated with surgery in the elderly obese. Ann Surg. 2012;256:79–86. [PubMed: 22566017]

29. Yang D, Small DS, Silber JH, et al. Optimal matching with minimal deviation from fine balance in a study of obesity and surgical outcomes. Biometrics. 2012;68:628–636. [PubMed: 22008180]

30. Kruskal W, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47:583–621.

31. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika. 1986;73:751–754.

32. Heller R, Rosenbaum PR, Small D. Using the cross-match test to appraise covariate balance in matched pairs. Am Stat. 2010;64:299–309.

33. Rosenbaum PR. An exact distribution-free test comparing two multivariate distributions based on adjacency. J R Statist Soc Series B. 2005;67:515–530.

34. Rubin DB. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. Stat Med. 2007;26:20–36. [PubMed: 17072897]

35. Rubin DB. For objective causal inference, design trumps analysis. Ann Appl Stat. 2008;2:808–840.

36. Keren R, Luan X, Localio R, et al. Prioritization of comparative effectiveness research topics in hospital pediatrics. Arch Pediatr Adolesc Med. 2012;166:1155–1164. [PubMed: 23027409]

37. Rosenbaum PR. Reduced sensitivity to hidden bias at upper quantiles in observational studies with dilated treatment effects. Biometrics. 1999;55:560–564. [PubMed: 11318214]

38. Rosenbaum PR. Chapter 5: Models for treatment effects. 5.3: Dilated effects In: Observational Studies. 2nd ed New York: Springer-Verlag; 2002:173–179.

39. Silber JH, Rosenbaum PR, Ross RN, et al. Indirect standardization matching: Assessing specific advantage and risk synergy. Health Serv Res. 2016;51:2330–2357. [PubMed: 26927625]
40. Silber JH, Rosenbaum PR, McHugh MD, et al. Comparison of the value of nursing work environments in hospitals across different levels of patient risk. JAMA Surg. 2016;151:527–536. [PubMed: 26791112]
41. Ripley B, Venables B, Bates DM, et al. Support Functions and Datasets for Venables and Ripley's MASS. MASS package version 7.3–40. 2015.
42. Hampel FR, Ronchett EM, Rousseeuw PJ, et al. Chapter 6. Linear Models: Robust Estimation In: Robust Statistics. The Approach Based on Influence Functions. New York, NY: John Wiley & Sons; 1986:307–341.
43. Huber PJ. Robust Statistics. Hoboken, NJ: John Wiley & Sons; 1981.
44. Cochran WG. The comparison of percentages in matched samples. Biometrika. 1950;37:256–266. [PubMed: 14801052]
45. Friedman M The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc. 1937;32:675–701.
46. Tukey JW. Chapter 2. Schematic Summaries (pictures and numbers). Section 2G: Eighths, sixteenths, etcet al. In: Exploratory Data Analysis. Reading, MA: Addison-Wesley Publishing Company Inc; 1977:53–54.
47. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74:829–836.
48. Lindley LC, Lyon ME. A profile of children with complex chronic conditions at end of life among Medicaid beneficiaries: Implications for health care reform. J Palliat Med. 2013;16:1388–1393. [PubMed: 24102460]
49. Petrou S, Kupek E. Socioeconomic differences in childhood hospital inpatient service utilisation and costs: prospective cohort study. J Epidemiol Community Health. 2005;59:591–597. [PubMed: 15965144]
50. Hofer TP, Wolfe RA, Tedeschi PJ, et al. Use of community versus individual socioeconomic data in predicting variation in hospital use. Health Serv Res. 1998;33:243–259. [PubMed: 9618670]

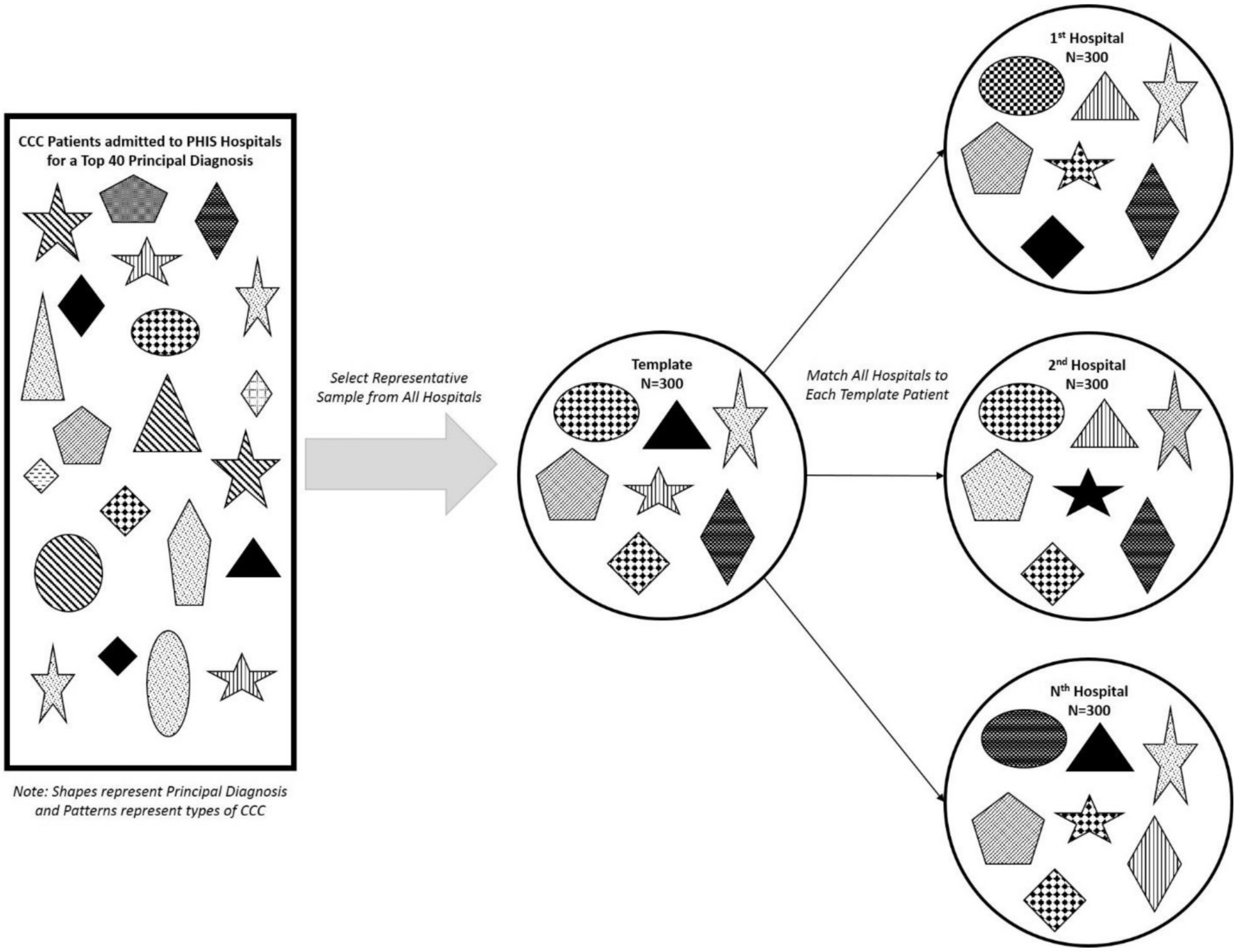


**Figure 1. Template Creation and Matching Process:**

The template was constructed by creating 2,000 random samples of 300 patients from the PHIS data set of study eligible patients, and selecting that template that had the smallest Mahalanobis distance to the patients admitted for a top 40 principal diagnosis. This created a representative template of CCC patients across study hospitals. The template was used to create strata of similar patients across the study hospitals.

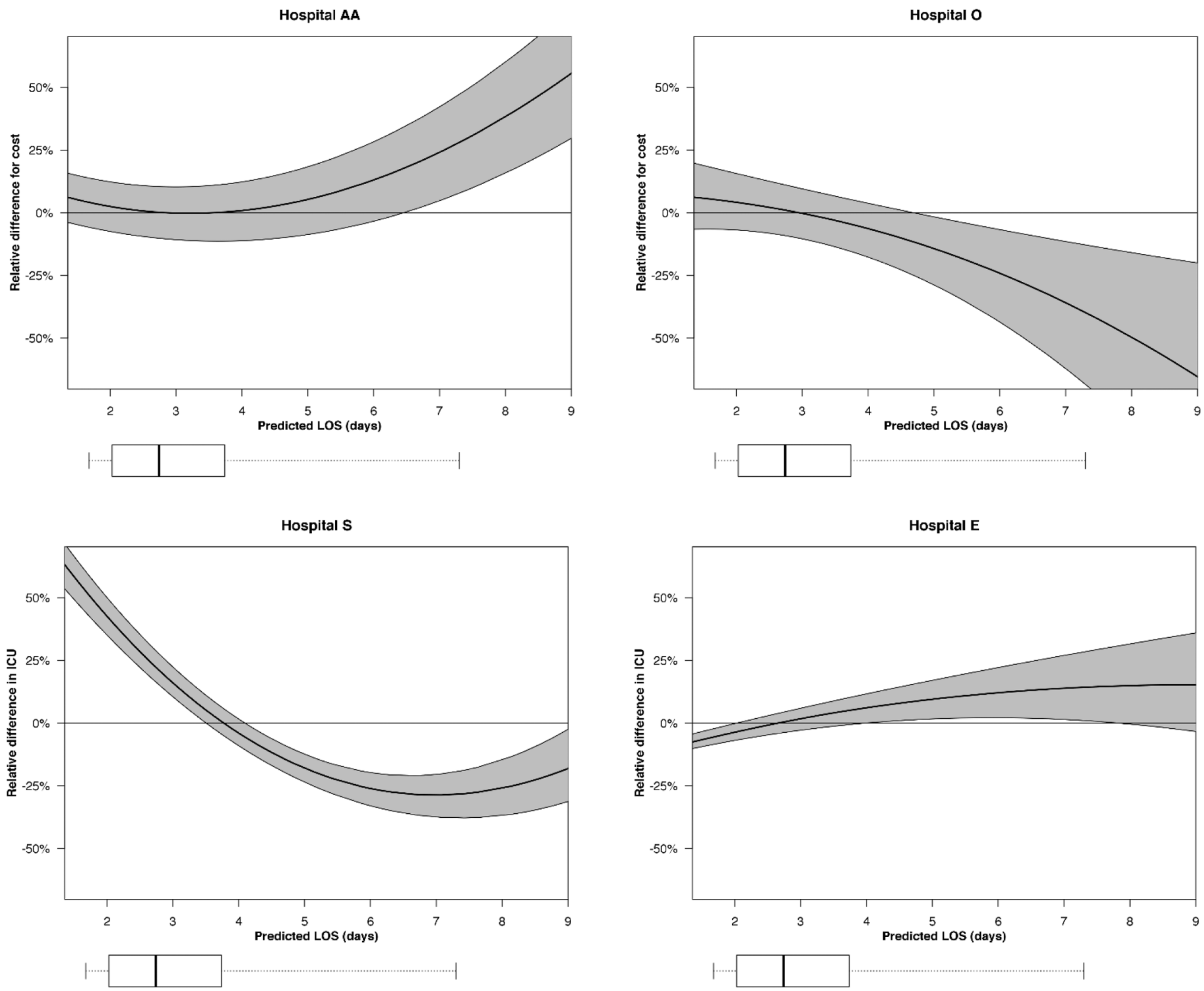


**Figure 2. Risk Synergy Plots for Selected Hospitals:**

The x-axis of each graph represents the risk, estimated by predicted length of stay on admission, for each template patient strata. The y-axis for the Hospitals **AA** and O represents the relative difference in log cost (focal minus control) inside each matched pair. The y-axis for Hospitals S and E represents the difference in ICU utilization (focal minus control). A point falling on the horizontal line at 0 represents no difference in cost (or ICU utilization) between the 2 patients in the matched pair, and a point falling below the line suggests a lower cost (or ICU utilization) for the focal vs control patient. The solid lines represent the locally weighted scatterplot smoothing (LOWESS) line.[47] LOWESS 95% CI bands (shaded areas) for the central tendency line were produced using the bootstrap method. A box plot at the bottom of each graph denotes the 5%, 25%, 50%, 75%, and 95% values of the predicted risk over all strata. Each graph illustrates an individual hospital. Note, all 4 slopes (risk synergy plots) were statistically significant (see Supplemental Table 11).

## Table 1.

### ICD-9 Codes for Complex Chronic Condition Variables.

The following table presents codes used to define the Complex Chronic Conditions (CCCs) and principal diagnoses used to define the study cohort. The CCCs are based on Feudtner et al.[1] The principal diagnoses represent the 40 most common principal diagnoses for inpatient medical conditions in the PHIS database. We excluded codes for external causes of injury and supplemental information (E/V) codes. See Supplemental Tables 1, 2, and 3 for additional details.

| Complex Chronic Conditions | | |
|---|---|---|
| **Level 1- Coarse Groups** | **Level 2- Intermediate Groups** | **Level 3- Most Refined Groupings (ICD-9)** |
| Respiratory & Cardiovascular | Cardiovascular | 745.0–747.4; 425.0–425.4; 429.1; 426.0–427.4; 427.6–427.9 |
| | Respiratory | 748.0–748.9; 770.7; 277.0x |
| Malignancy & Hematologic/Immunologic | Hematologic/Immunologic | 282.5–282.6; 282.0–282.4; 279.00–279.9; 288.1–288.2; 446.1; 042 |
| | Malignancy | 140–209.x |
| Other | Gastrointestinal | 750.3; 751.1–751.3; 751.6–751.9; 571.4–571.9; 555.0–556.9 |
| | Metabolic | 270.0–270.9; 271.0–271.9; 272.0–272.9; 277.3; 277.5; 275.0–275.3; 277.2; 277.4; 277.6; 277.8–277.9 |
| | Neuromuscular | 740.0–742.9; 318.0–319.0; 330.0–337.9; 343.0–343.9; 345.0–345.9; 359.0–359.3 |
| | Other Congenital or Genetic Defect | 758.0–758.9; 259.4; 737.3; 756.0–756.5; 553.3; 756.6–756.7; 759.7–759.9 |
| | Renal | 753.0–753.9; 585 |
| **Principal Diagnoses** | | |
| **Level 1- Coarse Groups** | **Level 2- Intermediate Groups** | **Level 3- Most Refined Groupings (ICD-9)** |
| Neoplasm/ Blood | Neoplasms (140–239) | 204.00 |
| | Blood and blood-forming organs (280–289) | 282.61; 282.62; 282.64; 288.00; 288.03 |
| Circulatory/ Respiratory | Circulatory system (390–459) | 427.89; 446.1 |
| | Respiratory system (460–519) | 464.10; 465.9; 466.11; 466.19; 486 ; 487.1; 493.91; 493.92; 507.0; 518.81 |
| Digestive/ Genitourinary | Digestive system (520–579) | 555.9; 558.9; 564.00 |
| | Genitourinary system (580–629) | 599.0 |
| Endocrine/ Ill-defined | Endocrine, nutritional/metabolic, immunity (240–279) | 276.51; 277.02 |
| | Ill-defined conditions (780–799) | 780.39; 780.60; 783.41; 784.0 |
| Other | Infectious/parasitic diseases (001–139) | 008.8; 038.9; 079.99 |
| | Nervous system (320–359) | 345.10; 345.11; 345.3; 345.40; 345.41; 345.50; 345.90; 345.91 |
| | Medical/Surgical Complications (996–999) | 999.31 |

Note: If patients were only flagged as a CCC because of a benign tumor, they were excluded from the study.

**Table 2.**

**Assessing if Patient Covariates and Outcomes Vary Significantly Across Hospitals (N = 40 hospitals, 300 patients per hospital)**

In this table we present the distribution of each covariate across the 40 hospitals. Note that for patient demographic, CCC, and principal diagnosis covariates matched through the template, the $\chi^{2}$/df is less than 1 (indicating better balance than would be achieved by random allocation of patients to hospitals) and the p-values are insignificant. Despite finding clinically similar patients, we see large variation in outcomes.

| | Percentile Range for Template Matched Hospitals (N=40 Hosps) | | | | | | |
|---|---|---|---|---|---|---|---|
| **Patient Covariates* (Percent unless noted)** | **Lower Eighth (12.5% Hosps)** | **Lower Quartile (25% Hosps)** | **Median (50% Hosps)** | **Upper Quartile (75% Hosps)** | **Upper Eighth (87.5% Hosps)** | **$\chi^2$ / 39**** | **P-value** |
| **Demographic Covariates** | | | | | | | |
| Age (years) | 6.8 | 6.9 | 7.2 | 7.3 | 7.4 | 0.70 | 0.92 |
| Sex (% female) | 47.0 | 47.9 | 49.5 | 50.7 | 51.3 | 0.46 | 1.00 |
| Medicaid | 50.7 | 52.0 | 52.7 | 53.3 | 53.7 | 0.47 | 1.00 |
| History of Tracheotomy | 2.7 | 3.0 | 3.0 | 3.7 | 4.0 | 0.25 | 1.00 |
| History of GE Tube | 11.7 | 11.7 | 12.0 | 12.0 | 12.3 | 0.04 | 1.00 |
| History of Neurological Impairment | 53.3 | 53.3 | 53.3 | 53.7 | 54.0 | 0.02 | 1.00 |
| Mean Predicted Length of Stay | 2.9 | 2.9 | 2.9 | 3.0 | 3.0 | 0.12 | 1.00 |
| Mean Predicted Days in the ICU | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.10 | 1.00 |
| **CCC ICD9 Clinical Groupings** | | | | | | | |
| Cardiovascular | 12.0 | 12.3 | 12.8 | 13.0 | 13.3 | 0.09 | 1.00 |
| Respiratory | 6.7 | 7.0 | 7.3 | 7.7 | 7.7 | 0.14 | 1.00 |
| Hematologic or Immunologic | 14.7 | 15.0 | 15.3 | 15.7 | 15.7 | 0.04 | 1.00 |
| Malignancy | 10.0 | 10.3 | 10.7 | 10.7 | 11.0 | 0.05 | 1.00 |
| Gastrointestinal | 4.3 | 4.3 | 4.7 | 5.0 | 5.0 | 0.10 | 1.00 |
| Metabolic | 4.3 | 4.7 | 5.0 | 5.3 | 5.7 | 0.21 | 1.00 |
| Neuromuscular | 47.3 | 47.7 | 48.0 | 48.3 | 48.4 | 0.05 | 1.00 |
| Other Congenital or Genetic Defect | 11.7 | 12.3 | 12.5 | 13.0 | 13.3 | 0.15 | 1.00 |
| Renal | 1.7 | 2.0 | 2.3 | 2.7 | 3.0 | 0.42 | 1.00 |
| **Principal Diagnosis ICD9 Clinical Groupings** | | | | | | | |
| Neoplasms | 3.3 | 3.3 | 3.3 | 3.3 | 3.3 | 0.00 | 1.00 |
| Blood and Blood-forming organs | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 0.00 | 1.00 |
| Circulatory system | 7.7 | 7.7 | 7.7 | 7.7 | 7.7 | 0.00 | 1.00 |

| Patient Covariates* (Percent unless noted) | Percentile Range for Template Matched Hospitals (N=40 Hosps) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Lower Eighth (12.5% Hosps) | Lower Quartile (25% Hosps) | Median (50% Hosps) | Upper Quartile (75% Hosps) | Upper Eighth (87.5% Hosps) | $\chi^2 / 39$** | P-value |
| Respiratory system | 18.3 | 18.3 | 18.3 | 18.3 | 18.3 | 0.00 | 1.00 |
| Digestive system | 7.0 | 7.0 | 7.0 | 7.0 | 7.0 | 0.00 | 1.00 |
| Genitourinary system | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 | 0.00 | 1.00 |
| Endo/nutritional/metabolic/immunity | 6.0 | 6.0 | 6.0 | 6.0 | 6.0 | 0.00 | 1.00 |
| Ill-defined conditions | 11.3 | 11.3 | 11.3 | 11.3 | 11.3 | 0.00 | 1.00 |
| Infectious/parasitic diseases | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 0.00 | 1.00 |
| Nervous system | 32.3 | 32.3 | 32.3 | 32.3 | 32.3 | 0.00 | 1.00 |
| Medical/Surgical Complications | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 0.00 | 1.00 |
| ***Outcomes*** | | | | | | | |
| **Patients in ICU (%)** | 6.6 | 7.7 | 9.7 | 11.8 | 13.9 | 8.12 | <0.001 |
| **Length of Stay (Days)**† | 2.4 | 2.5 | 2.8 | 2.9 | 3.0 | 6.49 | <0.001 |
| **Cost ($)**† | 6,856 | 7,741 | 8,283 | 9,636 | 10,047 | 10.05 | <0.001 |

*Each row represents the distribution range of hospitals for a specific patient covariate or outcome. Hospitals may have different percentiles in different rows.

**For covariates, we compare the variation among individual hospitals in patient covariates to the variation that would have been expected had patients been randomly assigned to hospitals. For the $\chi^2$ statistics, a $\chi^2$/degrees of freedom (=39 for 40 hospitals) that is greater than 1 suggests more variation than random, and less than 1 suggests less variation than random. For patient covariates, we checked the balance using $\chi^2$ tests for binary variables and the Kruskal-Wallis test for continuous variables.

†For outcome variables, we tested for differences within matched sets using Cochran's Q for binary variables and Friedman's test for continuous variables. For Length of Stay and Cost, we display the m-estimates, and report the statistic and P-value using the Friedman's Test.

‡Note that patient characteristics are very similar across hospitals whereas hospital outcomes are not.

**Table 3.**

**Practice Style Results Across 40 PHIS Hospitals.**

Here we display the overall results for the 12,000 matched patients, and for each hospital's 300 matched patients individually. The hospitals are ranked in four ways: on their median resource utilization, their 90th percentile resource utilization, 90th percentile length of stay, and rate of ICU admission among their template matched patients.

| | Resource Utilization | | | | Length of Stay | | ICU Utilization | |
|---|---|---|---|---|---|---|---|---|
| **Hospital** | **Median** | **Rank** | **90th** | **Rank** | **90th** | **Rank** | **%** | **Rank** |
| All Hospitals | $6,417 | NA | $27,438 | NA | 8 | NA | 10.5 | NA |
| A | $4,736[d] | 1 | $18,885 | 3 | 7.0 | 13.5 | 10.0 | 21 |
| B | $4,886[d] | 2 | $15,278[d] | 1 | 5.0[b] | 1.5 | 6.3[a] | 4.5 |
| C | $4,895[d] | 3 | $17,841[b] | 2 | 6.0[a] | 5 | 11.3 | 25.5 |
| D | $5,205[d] | 4 | $23,512 | 12 | 8.0 | 24 | 5.0[c] | 1 |
| E | $5,325[b] | 5 | $24,822 | 18 | 7.0 | 13.5 | 11.7 | 29 |
| F | $5,486 | 6 | $40,676[b] | 38 | 11.0[b] | 39.5 | 11.0 | 23 |
| G | $5,643[a] | 7 | $22,192 | 8 | 6.0 | 5 | 8.0 | 13.5 |
| H | $5,720 | 8 | $29,757 | 25 | 8.0 | 24 | 15.7[b] | 36 |
| I | $5,854 | 9 | $22,476 | 9 | 7.0 | 13.5 | 7.3 | 8.5 |
| J | $5,888 | 10 | $32,430[a] | 29 | 7.0 | 13.5 | 8.3 | 16 |
| K | $5,918 | 11 | $24,878 | 19 | 8.1[a] | 30.5 | 7.3 | 8.5 |
| L | $5,934 | 12 | $24,154 | 13 | 8.0 | 24 | 11.3 | 25.5 |
| M | $6,030 | 13 | $37,425 | 36 | 7.0 | 13.5 | 5.7[b] | 2 |
| N | $6,103 | 14 | $30,953 | 27 | 9.0 | 35 | 12.0 | 31 |
| O | $6,276 | 15 | $20,704 | 5 | 6.0[b] | 5 | 6.7[a] | 6.5 |
| P | $6,309 | 16 | $23,297 | 11 | 7.0 | 13.5 | 13.0 | 33 |
| Q | $6,323 | 17 | $19,735[b] | 4 | 7.0 | 13.5 | 10.3 | 22 |
| R | $6,351 | 18 | $24,646 | 16 | 8.0 | 24 | 12.3 | 32 |
| S | $6,399 | 19 | $21,612[c] | 7 | 7.0[a] | 13.5 | 30.7[d] | 40 |
| T | $6,422 | 20 | $24,270 | 15 | 8.0 | 24 | 8.7 | 17.5 |
| U | $6,446 | 21 | $23,110 | 10 | 8.0 | 24 | 17.3[d] | 39 |
| V | $6,472 | 22 | $24,268 | 14 | 7.0 | 13.5 | 16.7[c] | 38 |
| W | $6,541 | 23 | $26,550 | 22 | 9.0[a] | 35 | 16.0[c] | 37 |
| X | $6,849 | 24 | $21,171[b] | 6 | 8.0 | 24 | 13.3 | 34 |
| Y | $6,919 | 25 | $41,961[a] | 39 | 8.0 | 24 | 13.7 | 35 |

| | Resource Utilization | | | | Length of Stay | | ICU Utilization | |
|---|---|---|---|---|---|---|---|---|
| **Hospital** | **Median** | **Rank** | **90th** | **Rank** | **90th** | **Rank** | **%** | **Rank** |
| Z | $6,920 | 26 | $34,133[b] | 31 | 9.0 | 35 | 6.0[b] | 3 |
| AA | $6,923 | 27 | $29,400 | 24 | 8.1 | 30.5 | 8.0 | 13.5 |
| BB | $6,966 | 28 | $30,760 | 26 | 9.0[b] | 35 | 8.7 | 17.5 |
| CC | $6,983 | 29 | $35,011 | 33 | 7.0 | 13.5 | 9.0 | 19 |
| DD | $7,017 | 30 | $25,614 | 21 | 6.0[b] | 5 | 11.3 | 25.5 |
| EE | $7,083[a] | 31 | $36,812[b] | 35 | 9.0[a] | 35 | 11.3 | 25.5 |
| FF | $7,130[a] | 32 | $33,147 | 30 | 11.0[a] | 39.5 | 9.3 | 20 |
| GG | $7,409 | 33 | $34,856[b] | 32 | 6.1 | 8 | 11.7 | 29 |
| HH | $7,412 | 34 | $24,815 | 17 | 8.1 | 30.5 | 8.0 | 13.5 |
| II | $7,441[b] | 35 | $51,875[d] | 40 | 7.1 | 19 | 8.0 | 13.5 |
| JJ | $7,494[b] | 36 | $27,739 | 23 | 10.1[b] | 38 | 7.7 | 10.5 |
| KK | $7,592[c] | 37 | $32,107 | 28 | 5.0[b] | 1.5 | 6.7[a] | 6.5 |
| LL | $7,654[a] | 38 | $25,510 | 20 | 8.0 | 24 | 7.7 | 10.5 |
| MM | $7,884[b] | 39 | $35,843[a] | 34 | 8.1 | 30.5 | 11.7 | 29 |
| NN | $8,427[d] | 40 | $38,714[b] | 37 | 6.0[a] | 5 | 6.3[a] | 4.5 |

[a] p<0.05,

[b] p<0.01,

[c] p<0.001,

[d] p<0.0001

[*] All c and d meet the criteria for the Bonferroni correction (p<= 0.001 < 0.05/40 = 0.00125)

[†] The given P-values for Cost, Length of Stay, ICU Days and ICU utilization are calculated using a Mantel-Haenszel stratified test for conditional independence.[37]

[‡] When ranks of hospitals were tied, we report the average rank.