ECONOMETRICA, INC.

REPORT

Appendix C – Comparison Hospital Methodology

Evaluation and Monitoring of the Bundled Payments for Care Improvement Model 1 Initiative

Contract No.:
HHSM-500-2011-00015I

Order No.:
HHSM-500-T0008

Project No.:
2248-000

Submitted To:

Centers for Medicare & Medicaid Services
Attn.: Arpit Misra
Contracting Officer's Representative
Center for Medicare & Medicaid Innovation
7205 Windsor Boulevard, Mail Stop C3-21-28
Baltimore, MD 21244

Submitted By:

Econometrica, Inc.
7475 Wisconsin Avenue
Suite 1000
Bethesda, MD 20814
www.EconometricaInc.com

July 9, 2015

ECONOMETRICA, INC.

July 9, 2015

Mr. Arpit Misra
Contracting Officer's Representative
Centers for Medicare & Medicaid Services
Center for Medicare & Medicaid Innovation
7205 Windsor Boulevard, Mail Stop C3-21-28
Baltimore, MD 21244-1850

Reference: Contract No.: HHSM-500-2011-00015I; Order No.: HHSM-500-T0008; "Evaluation and Monitoring of the Bundled Payments for Care Improvement Model 1 Initiative" (Project No.: 2248-000).

Dear Mr. Misra:

Econometrica is pleased to submit this Appendix C – Comparison Hospital Methodology as part of the Annual Report to the Centers for Medicare & Medicaid Services, Center for Medicare & Medicaid Innovation, regarding work being conducted under the above-referenced contract.

Appendixes A, B, D, and E are being submitted as separate files.

If you wish to discuss any aspect of this submission, please feel free to contact me at (301) 395-2281.

Sincerely,

Econometrica, Inc.

Monique Sheppard, Ph.D.
Project Director

cc: Contract File

Table of Contents

List of Tables

List of Figures

Appendix C: Comparison Hospital Methodology

The impact of Bundled Payments for Care Improvement (BPCI) Model 1 participation is identified by comparing two different hospital groups:

- *BPCI Awardees* include all Awardees that participate in the BPCI Model 1 initiative for the corresponding quarter.
- *BPCI Comparison hospitals* include hospitals that are not participating in BPCI Model 1 but are otherwise comparable to the BPCI Awardees in terms of measureable contextual factors that are likely to impact the effectiveness of the policy (confounding contextual factors).

Finding a cohort of (statistically) similar Comparison hospitals allows for an effectively created counterfactual scenario that, by comparing the two cohorts, enables the beginning of isolating impacts that may be attributed to participation in BPCI Model 1. The remainder of this section details (data) issues that drove the methods, rationale for criteria used to define the comparability of a comparison hospital, and the results.

C.1. Selection of an Appropriate Comparison Cohort

The matching design accounts for three distinct issues. *First,* the evaluation includes a large number of outcome variables across five domains:[1]

- *Case Mix and Patient Characteristics* (e.g., age, CMS-HCC scores).
- *Care Process* (e.g., heart failure composite).
- *Health Outcome and Resource Utilization* (e.g., mortality, readmissions).
- *Medicare Payments and Internal Costs* (e.g., total Medicare payments and Awardee cost savings).
- *Patient Care Experience* (e.g., patient satisfaction with care).

Second, the evaluation focuses on a small group of Awardee hospitals. At the time of writing this report, 24 Awardee ("treated") hospitals were considered.[2] *Third,* all but one of these treated hospitals are located in one State: New Jersey; the one exception is located in Kansas. The relatively large pool of potential Comparison hospitals is spread across the country.

The large number of potential outcome measures required for this evaluation[3] and the small sample size create a methodological challenge. For these reasons, Mahalanobis matching is utilized, which uses a distance measure that does not rely on an econometric model or distributional assumptions and is consequently better suited to deal with small sample sizes.

[1] These are all domains considered throughout the demonstration period of performance; not all are presented in this report.
[2] This includes all 24 BPCI Model 1 Awardee hospitals.
[3] Approximately 134 measures were initially considered and that number was reduced for this report (see Section III.A).

The Mahalanobis distance $D_{i,j}$ between units (i.e., hospitals) *i* and *j* is defined as:

$$D_{i,j} = \sqrt{(X_i - X_j)'\Sigma^{-1}(X_i - X_j)}$$

where X_i and X_j represent the set of matching covariates for units *i* and *j*, respectively, and Σ represents the variance-covariance matrix of X in the full comparison group when the study focuses on the average treatment effect on the treated (ATT),[4] or in the pooled treatment and full comparison groups when it focuses on the average treatment effect (ATE)[5] (Stuart, 2010).

ATE may include the effect on individuals or units for which the program was not intended, and ATT is considered to be better suited (Heckman, 1997) for this reason. Accordingly, ATT is used. Once the distance between all possible pairs of treated units is calculated, one or more Comparison hospitals are selected for each treated unit. Different matching procedures are available to select Comparison hospitals. The most common and frequently used is *k:1* nearest neighbor, which consists of selecting for each treated unit *i* the *k* comparison units with the smallest distance from *i* (Stuart, 2010; Stuart and Rubin, 2007). The simplest case is *k=1*, which implies that only one comparison hospital is selected per treated unit. Although choosing *k>1* in theory may lead to higher power in the calculations, (in practice) the gain is often minimal (Stuart 2010) for two reasons. *First,* in a comparison of two samples, the precision is largely driven by the smaller group size (Cohen, 1988). This means that as long as the number of treated units remains fixed, there would be little advantage in expanding the group of selected Comparison hospitals. *Second*, the statistical power increases when the two groups are more similar because of higher precision (Snedecor and Cochran, 1980). Because the first matched hospital is by construction at least as good[6] as an eventual second matched hospital, *k>1* may lower the precision. As discussed below, however, 4:1 matching was used for two reasons. It was found that additional matches did not substantially degrade the matching quality. This may occur in cases where there is a large universe of potential Comparison hospitals.[7] At the same time, future reports will be generated that focus on each BPCI Awardee separately. In that context, it will be desirable to provide comparison information based on several (in this case, four) matched hospitals rather than just a single matched hospital.

Another important issue is whether potential Comparison hospitals can be used as matches for more than one treated unit—often called matching "with replacement," as opposed to matching "without replacement" (Stuart and Rubin, 2007). A first advantage of matching with replacement is that the order of matching (i.e., which hospital is matched first) does not matter. A second advantage is that matching with replacement often yields better matches, because Comparison hospitals that are similar to treated units can be used more than once. A potential disadvantage is that only a few unique comparison units may end up being selected as matches. For this reason, monitoring the number of times a comparison unit is matched is considered a best practice (Stuart and Rubin, 2007).

---

[4] ATT is the average gain from treatment for those who were treated.
[5] ATE is the average gain from treatment for a unit randomly selected from the population.
[6] The first comparison hospital matched to a treatment hospital will be the best match, given predefined criteria and the potential pool for comparison matches.
[7] Stuart, E. A., & Rubin, D. B. (2007). "Matching methods for causal inference: Designing observational studies." Accessed November 20, 2014, from http://www.biostat.jhsph.edu/~estuart/StuRub_MatchingChapter_07.pdf.

Based on these considerations, the Mahalanobis metric was opted to be used with 4:1 nearest neighbor matching, with replacement, on a reduced number of covariates whose effects span all outcome domains.[8] Following the advice of matching methodologists (Rubin, 2001; Stuart, 2010), the selection of covariates is based on scientific understanding of the different confounding factors that drive changes in the outcome variables (primarily) and program participation. Table C.1 describes the variables included in the analysis and their sources.[9] Data were matched up from the first quarter of 2011 through the first quarter of 2013.

Table C.1: Variable Names, Specifications, and Sources

Variable Name	Technical Specification	Source
Pre-Match Variables		
Urban	Equal to 1 if the hospital is located in an urban area; 0 otherwise.	FY 2014 Final Rule
Provider type	Nine types of acute care hospitals: • Inpatient Prospective Payment System (IPPS) • Rural Referral Center (RRC) • Indian Health Service • Medicare Dependent Hospital (MDH) • MDH/RRC • Sole Community Hospital (SCH) • SCH/RRC • Essential Access Community Hospital (EACH) • EACH/RRC	FY 2014 Final Rule
Hospital Characteristics		
Indicator for general hospital	Equal to 1 for general hospital; 0 otherwise.	AHA Annual Survey Database FY 2011
Indicator for teaching hospital	Equal to 1 for teaching hospital; 0 otherwise.	AHA Annual Survey Database FY 2011
Indicator for presence of Emergency Department (ED)	Equal to 1 if the hospital has an ED; 0 otherwise.	AHA Annual Survey Database FY 2011
Number of beds	Number of hospital beds.	FY 2014 Final Rule
Urban type	Equal to 1 if hospital is located in a large urban area; equal to 2 if hospital is located in another type of urban area.	FY 2014 Final Rule
Patient and Inpatient Episode Characteristics		
Indicator for high surgical Medicare Severity-Diagnosis-Related Group (MS-DRG) percentage	Equal to 1 if hospital has more than 90 percent surgical MS-DRGs.	2008 Inpatient 5% Base Claims Database
Percent of hospital days paid by Medicare	Medicare days as a proportion of total.	FY 2014 Final Rule

[8] Caliper matching was also experimented with, which specifies a "not to exceed" distance for all matches.
[9] Several other matching specifications were considered, using 1:1, 2:1, and 3:1 nearest neighbor matching as well as additional covariates, such as change in readmission and mortality rates, case mix index, average MS-DRG weight, disproportionate share percentage, and proportion of patients who are dually eligible.

Variable Name	Technical Specification	Source
Average Hierarchical Condition Category (HCC) score	Calculated from inpatient claims and HCC files.	Inpatient claims HCC files
Length of stay	Calculated from inpatient claims file.	Inpatient claims
Change in length of stay between: • Q1 2011 and Q2 2011 • Q3 2011 and Q4 2011 • Q1 2012 and Q2 2012 • Q3 2012 and Q4 2012	Calculated from inpatient claims file.	Inpatient claims
All-cause mortality	Calculated from inpatient claims and Beneficiary Summary File.	Inpatient claims Beneficiary Summary File
All-cause readmissions	Calculated from inpatient claims file.	Inpatient claims

The BPCI Model 1 Awardees are all located in urban areas and are all subject to the IPPS. Because these two characteristics are fundamental to hospital functioning and payments, the universe of potential Comparison hospitals were excluded from any that are not located in urban areas or do not receive payments according to IPPS. This results in an “exact match” between the treatment and comparison hospital for these two dimensions. The two variables listed under the “Pre-Match Variables” category in Table C.1 were used for this purpose.

Once the universe of potential Comparison hospitals was reduced, two categories of variables were matched: hospital characteristics and patient characteristics. The hospital characteristics include indicators for whether the hospital is a general hospital, whether it is a teaching hospital, and whether it has an ED. Bed size was also matched and a variable capturing whether the hospital is located in a large urban or other urban area. These variables capture characteristics that fundamentally affect how hospitals function. For example, general and specialty hospitals are likely to be affected differently by new payment models because specialty hospitals are likely to serve more complex patients with different baseline quality outcomes and to treat varying DRGs. In addition, general hospitals may be more likely to participate in initiatives such as BPCI Model 1.

Similarly, teaching hospitals operate differently from non-teaching hospitals. In particular, hospitals with a graduate medical education program receive additional indirect medical education payments “to reflect the higher patient care costs of teaching hospitals relative to non-teaching hospitals.”[10] Two variables were also matched capturing hospital capacity—presence of an ED and number of hospital beds. Hospitals with different values for these variables may also treat different types of patients.

Among urban hospitals, large urban areas were differentiated between those located in all other urban areas. This is important in the specific context of this evaluation, because most treated hospitals are located in New Jersey. Unlike most States, New Jersey is predominantly urban with very few rural areas. However, New Jersey does present multiple suburban areas whose hospitals

---

[10] CMS. (2014, August 8). *Indirect medical education.* Retrieved from http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Indirect-Medical-Education-IME.html.

are likely to behave differently compared with those in large urban centers. In an approach that merely distinguishes between urban and rural settings, these suburban areas would be categorized together with large urban centers. Urban area size may also be associated with quality outcomes such as readmissions and care experience.

The matching procedure includes a number of patient and inpatient episode characteristics as well. The model includes an indicator for hospitals with a high (greater than 90 percent) proportion of inpatient stays falling within surgical MS-DRGs. This facilitates matching for the Kansas Surgery and Recovery Center, which is a surgical specialty hospital with a very small proportion of non-surgical inpatient stays. The distinction is important because the types of services provided to medical patients are very different from those provided to surgical patients. In addition, medical and surgical patients may have different outcomes and different types of complications and rely on different provider processes and payment streams. The characteristics of patients admitted for medical and surgical care can often differ as well. For example, some hospitals may have many private-payer elective surgeries.

The percent of hospital days paid by Medicare captures information that could affect hospitals' participation decisions. BPCI Model 1 is based on inpatient stays paid by Medicare. The larger the proportion of hospital days paid by Medicare, the larger the incentive provided by the model. In addition, matching on proportion of Medicare days ensures that treatment and Comparison hospitals face similar incentives such as those provided by the Inpatient Quality Reporting program, which reduces the annual payment update for Medicare hospitals that fail to report certain quality measures.

The average HCC score for patients with an inpatient stay at the hospital summarizes the clinical complexity of the population served by the hospital. Chronic condition burden, which the HCC score captures, is correlated with patient outcomes[11] and expenditures. Patient acuity could also be linked to ability and/or appetite to take on additional risk by participating in the BPCI initiative. For example, hospitals serving patients with higher HCC scores may have more potential savings opportunities.

Finally, baseline metrics are included for outcome variables in the matching process: length of stay, mortality, and readmissions. While HCC scores capture patients' chronic disease burden, length of stay proxies for comorbidities, complications, and expenditures for a specific inpatient stay. Furthermore, an important channel through which hospitals may reduce internal costs is by reducing length of stay. In addition to matching on the average length of stay during the baseline period, the change in length of stay is also matched over four intervals: Q1 2011–Q2 2011, Q3 2011–Q4 2011, Q1 2012–Q2 2012, and Q3 2012–Q4 2012. Matching on *changes* in this variable helps ensure that the assumption necessary for a valid difference in differences (DiD) analysis (i.e., that pre-period trends in key outcome variables are parallel) is satisfied.

[11] Li, P., Kim, M. M., & Doshi, J. A. (2010). Comparison of the performance of the CMS Hierarchical Condition Category (CMS-HCC) risk adjuster with the Charlson and Elixhauser comorbidity measures in predicting mortality. *BMC Health Services Research, 10*(245). Accessed November 20, 2014, from http://www.biomedcentral.com/1472-6963/10/245.

Readmission and mortality were selected because they are important indicators of baseline quality-of-care and coordination-of-care performance. In addition, these variables are strongly linked to patient demographic and community factors (for which it can be difficult to control).

Table C.2 shows the number of BPCI Model 1 treated hospitals, the number of potential Comparison hospitals, the number of matched Comparison hospitals, the number of unique matched Comparison hospitals, and the number of replacements. Table C.3 lists the Comparison hospitals matched to each treated hospital and the State in which each hospital is located.[12]

Table C.2: Number of Treated and Comparison Hospitals

Number of BPCI Model 1 Treated Hospitals	Number of Potential Comparison Hospitals	Number of Matched Comparison Hospitals	Number of Unique Matched Comparison Hospitals	Number of Replacements
24	1,851	96	82	30

[12] Unlike propensity score matching, Mahalanobis does not base the computation of the distance on a propensity score estimated through regression analysis. Hence, there are no regression analysis intermediate results.

Table C.3: BPCI Model 1 Treated Hospitals and Comparison Group Hospitals

Treated Hospital		Comparison Hospital 1		Comparison Hospital 2		Comparison Hospital 3		Comparison Hospital 4	
Name	State	Name	State	Name	State	Name	State	Name	State
Overlook Medical Center	NJ	Baptist Hospital East	KY	Sentara Leigh Hospital	VA	Sentara Virginia Beach General Hospital	VA	Northwest Hospital & Medical Center	WA
Capital Health Medical Center – Regional	NJ	St. Vincent's Medical Center	CT	Baton Rouge General Medical Center	LA	ProMedica Flower Hospital	OH	East Texas Medical Center Tyler	TX
Deborah Heart and Lung	NJ	Heart Hospital of Lafayette	LA	Bakersfield Heart Hospital	CA	St. Vincent Heart Center of Indiana	IN	Massachusetts Eye and Ear Infirmary	MA
Morristown Medical Center	NJ	Hartford Hospital	CT	Saint Francis Hospital and Medical Center	CT	Emory University Hospital	GA	Texas Health Presbyterian Hospital Dallas	TX
Robert Wood Johnson University Hospital	NJ	North Shore University Hospital	NY	Baptist Hospital of Miami	FL	Saint Barnabas Medical Center	NJ	Clara Maass Medical Center	NJ
University Medical Center of Princeton at Plainsboro	NJ	Norwalk Hospital	CT	Danbury Hospital	CT	Saint Mary's Hospital	CT	Sacred Heart Hospital	WI
Saint Michael's Medical Center	NJ	Sinai-Grace Hospital	MI	Saint Louis University Hospital	MO	Louis A. Weiss Memorial Hospital	IL	Botsford Hospital	MI
CentraState Medical Center	NJ	Sentara Virginia Beach General Hospital	VA	Swedish Covenant Hospital	IL	Adventist La Grange Memorial Hospital	IL	Middlesex Hospital	CT
Capital Health Medical Center – Hopewell	NJ	St. Vincent's Medical Center	CT	Sacred Heart Hospital	WI	Norwalk Hospital	CT	Creighton University Medical Center	NE
Cooper Hospital	NJ	Staten Island University Hospital	NY	St. Luke's-Roosevelt Hospital Center	NY	DeKalb Medical at North Decatur	GA	Rhode Island Hospital	RI

Treated Hospital		Comparison Hospital 1		Comparison Hospital 2		Comparison Hospital 3		Comparison Hospital 4	
Name	State	Name	State	Name	State	Name	State	Name	State
Hunterdon Medical Center	NJ	Bon Secours-DePaul Medical Center	VA	Sentara Princess Anne Hospital	VA	Mayo Clinic Hospital	AZ	Bon Secours Maryview Medical Center	VA
Robert Wood Johnson University Hospital - Rahway	NJ	St. Joseph Hospital	NY	Regional Medical Center of San Jose	CA	Mercy Hospital Western Hills*	OH	Columbia St. Mary's Ozaukee Hospital	WI
St. Joseph's Regional Medical Center	NJ	Hackensack University Medical Center	NJ	St. Luke's-Roosevelt Hospital Center	NY	Emory University Hospital	GA	Rhode Island Hospital	RI
Saint Peter's University Hospital	NJ	Tristar Centennial Medical Center	TN	Brookwood Medical Center	AL	Monmouth Medical Center	NJ	UPMC St. Margaret	PA
St. Mary's Hospital Passaic	NJ	Regional Medical Center of San Jose	CA	San Jacinto Methodist Hospital	TX	Holy Name Medical Center	NJ	Palisades Medical Center	NJ
Jersey Shore University Medical Center	NJ	Hartford Hospital	CT	Sentara Norfolk General Hospital	VA	South Nassau Communities Hospital	NY	Florida Hospital Tampa	FL
Inspira Medical Center – Woodbury	NJ	Suburban Hospital	MD	Clara Maass Medical Center	NJ	McLaren Macomb	MI	St. Mary Mercy Hospital	MI
Robert Wood Johnson University Hospital – Hamilton	NJ	Florida Hospital Memorial Medical Center	FL	Sparks Regional Medical Center	AR	Capital Regional Medical Center	FL	Fort Walton Beach Medical Center	FL
The Valley Hospital	NJ	Northwest Community Hospital	IL	Oak Hill Hospital	FL	Baptist Memorial Hospital-Desoto	MS	Gaston Memorial Hospital	NC
Inspira Medical Center – Elmer	NJ	Duke Raleigh Hospital	NC	St. Elizabeth Medical Center	NY	Sebastian River Medical Center	FL	Longview Regional Medical Center	TX

Treated Hospital		Comparison Hospital 1		Comparison Hospital 2		Comparison Hospital 3		Comparison Hospital 4	
Name	State	Name	State	Name	State	Name	State	Name	State
Inspira Medical Center – Vineland	NJ	Texoma Medical Center	TX	Peace River Regional Medical Center	FL	Shore Medical Center	NJ	Fort Sanders Regional Medical Center	TN
Saint Clare's Hospital	NJ	Southwest General Health Center	OH	Gaston Memorial Hospital	NC	Atrium Medical Center	OH	Baptist Memorial Hospital-Desoto	MS
Kansas Surgery and Recovery	KS	North Carolina Specialty Hospital	NC	Fresno Surgical Hospital	CA	Oakleaf Surgical Hospital	WI	Patients' Hospital of Redding	CA
JFK Medical Center	NJ	Sentara CarePlex Hospital	VA	Mercy Suburban Hospital	PA	Sentara Virginia Beach General Hospital	VA	South Nassau Communities Hospital	NY

*This comparison hospital was ultimately dropped from analyses as it was considered a medical center with no inpatient beds in 2014.

C.2. Quality of Matched Comparison Cohort

The quality of the match is assessed by examining the bias in the covariates before and after matching. A high-quality match (i.e., low bias) would result in small standardized differences between BPCI Model 1 hospitals ("treatment group") and selected Comparison hospitals ("comparison group") across observable characteristics. Statistically, bias is defined using the formula provided in the seminal paper by Rosenbaum and Rubin (1985):

$$\text{Bias} = \frac{\bar{X}_T - \bar{X}_C}{\left(\frac{(\sigma_T^2 + \sigma_C^2)}{2}\right)^{1/2}}, \text{ (Equation 1)}$$

where $\bar{X}_T$ and $\bar{X}_C$ represent the sample means in the matched treatment and comparison groups, respectively, for a given covariate, and σ_T^2 and σ_C^2 represent the variances in the full (also known as "reservoir" or pre-matched) treatment and full comparison groups, respectively. Note that, because the differences are standardized using the treatment and comparison group variances (in the denominator), it is possible to observe large biases, even when the sample means of the treatment and comparison groups (in the numerator) are very similar. In particular, the lower the variance of the treatment(σ_T^2) and the comparison(σ_C^2) groups—or in other words, the more uniform each of these two groups—the larger the estimated bias, holding everything else constant.

Rosenbaum and Rubin (1985) suggest a 10-percent (in absolute value) threshold for the standardized difference after matching; Stuart (2010) suggests a 5-percent threshold. In addition to the absolute level of bias after matching, it is relevant to monitor the percentage reduction in bias after matching (Rosenbaum and Rubin, 1985). An effective matching procedure will lead to a large bias reduction. In addition, for the BPCI evaluation, a DiD approach is being used. Therefore, it is also essential to judge match quality based on whether the pretreatment trends in key outcomes for treatment and comparison groups are similar. This is discussed below.

Table C.4 shows overall bias before matching, i.e., the bias between treatment group and the *potential* comparison group (1,851 hospitals, Table C.2) and overall bias after matching, which compares the treatment group to actual, matched Comparison hospitals. The overall bias is the average of the percentage biases across individual variables matched on. The overall bias of 30.1 percent before matching is reduced to 9.3 percent after matching. Note that the after-matching bias exceeds the 5-percent rule of thumb threshold but falls below the 10-percent threshold.

Table C.4: Overall Matching Quality

Mean Bias Before Matching	Mean Bias After Matching
30.1%	9.3%

Table C.5 provides several statistics to assess the quality of the match at the covariate level. Column 1 indicates whether the statistics that follow correspond to the matched or the unmatched groups, and Column 4 (titled "% Bias") shows the results of applying equation (1) for each covariate. The statistic displayed in Column 4 summarizes the imbalance between the full treated and full potential comparison groups. For example, for the covariate "teaching hospital," the difference in means in the unmatched population (0.67 for the treated and 0.36 for the comparison) leads to a standardized bias of 64.7 percent. This value is well above the 5- and 10-percent thresholds already discussed. The t-test for the equality of these two means has a value of 3.15 and an associated *p value* of 0.002, which indicates that the means of treatment and comparison groups for the unmatched population are statistically different at the 5-percent level. This is evidence that, for the unmatched population, the treatment and comparison groups are imbalanced in terms of teaching status. In contrast, the difference can be seen in means between the treatment and comparison groups for the matched population for this variable leads to a standardized bias of 6.5 percent. Also, t-tests suggest that there is no longer a statistical difference between the means of the treated and comparison groups for the matched population.

Overall, Table C.5 indicates that, out of 15 variables, 6 presented biases larger than 10 percent and 7 have biases greater than 5 percent after matching. It is also important to note that these remaining biases will be at least partially removed later in this study, when multiple covariates are adjusted for in the DiD models.

Table C.5: Matching Quality Statistics

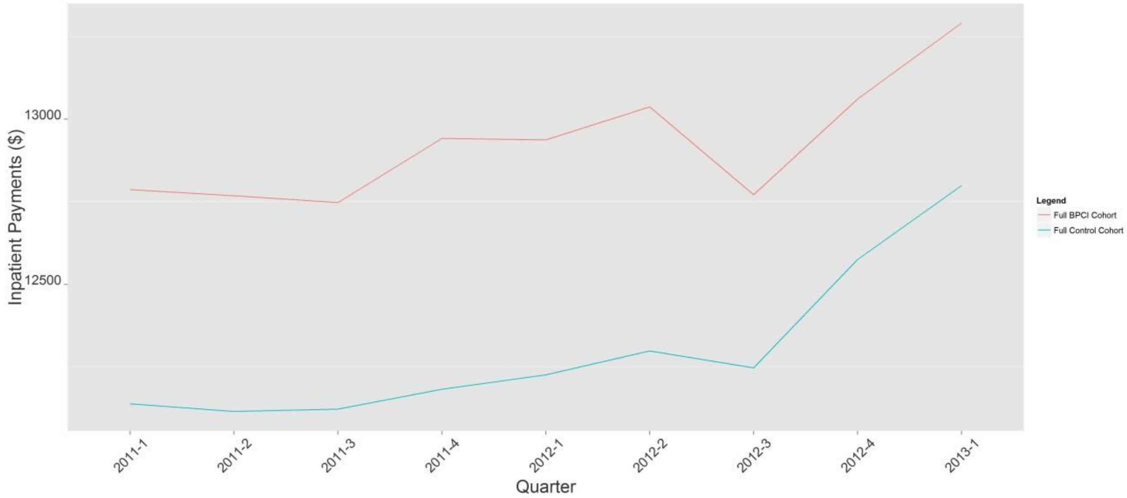
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Unmatched/ Matched	Mean Treated	Mean Comparison	% Bias	% Reduction in |Bias|	T-Test	p>|t|
Indicator for high surgical MS-DRG percentage	Unmatched	0.04167	0.02664	8.2	100	0.45	0.652
	Matched	0.04167	0.04167	0	100	0	1
Percent of hospital days paid by Medicare	Unmatched	0.44664	0.38124	59.5	78.9	2.51	0.012
	Matched	0.44664	0.43281	12.6	78.9	0.5	0.617
Number of beds	Unmatched	301.79	260.23	21.6	96.7	0.96	0.337
	Matched	301.79	300.44	0.7	96.7	0.03	0.979
Urban type	Unmatched	1.2917	1.4386	-30.6	85.8	-1.44	0.15
	Matched	1.2917	1.3125	-4.3	85.8	-0.15	0.878
Average HCC score	Unmatched	1.5567	1.4447	49	44.3	2.22	0.027
	Matched	1.5567	1.4943	27.3	44.3	1.06	0.296
Indicator for teaching hospital	Unmatched	0.66667	0.35608	64.7	89.9	3.15	0.002
	Matched	0.66667	0.63542	6.5	89.9	0.22	0.825
Indicator for general hospital	Unmatched	0.95833	0.97336	-8.2	100	-0.45	0.652
	Matched	0.95833	0.95833	0	100	0	1
Indicator for presence of ED	Unmatched	0.95833	0.96231	-2	100	-0.1	0.919
	Matched	0.95833	0.95833	0	100	0	1
Length of stay	Unmatched	6.394	5.8545	63.3	63.5	2.74	0.006
	Matched	6.394	6.1973	23.1	63.5	0.84	0.403

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Unmatched/ Matched	Mean Treated	Mean Comparison	% Bias	% Reduction in |Bias|	T-Test	p>|t|
All-cause mortality	Unmatched	0.06772	0.06554	11	99.3	0.54	0.586
	Matched	0.06772	0.06774	-0.1	99.3	0	0.998
All-cause readmissions	Unmatched	0.18638	0.17445	30.9	21.3	1.55	0.122
	Matched	0.18638	0.17699	24.3	21.3	0.86	0.394
Change in LOS, Q1 2011–Q2 2011	Unmatched	-0.20635	-0.17325	-10.1	9.9	-0.42	0.675
	Matched	-0.20635	-0.17652	-9.1	9.9	-0.45	0.655
Change in LOS, Q3 2011–Q4 2011	Unmatched	0.04924	0.1118	-15	83.9	-0.63	0.526
	Matched	0.04924	0.05929	-2.4	83.9	-0.11	0.911
Change in LOS, Q1 2012–Q2 2012	Unmatched	0.02448	-0.11547	37.3	52.5	1.46	0.144
	Matched	0.02448	-0.04195	17.7	52.5	0.9	0.372
Change in LOS, Q3 2012–Q4 2012	Unmatched	0.23087	0.08177	40.5	72.1	1.65	0.1
	Matched	0.23087	0.18928	11.3	72.1	0.56	0.581

The bias statistics presented above are useful measures of match quality and indicate, on the whole, that the matched hospitals constitute a good comparison group for the BPCI Awardees. However, in this evaluation of BPCI Model 1, the suitability of Comparison hospitals further depends on whether Comparison hospitals exhibit outcome variable trends that are similar to those of the Awardee hospitals. This additional qualification of a quality matched stems from satisfying the fundamental assumption for the DiD estimator to produce unbiased results: The trajectories of the dependent variables must differ for the two cohorts only because of the introduction of the intervention. In other words, pre-model (treatment) trends should be parallel.

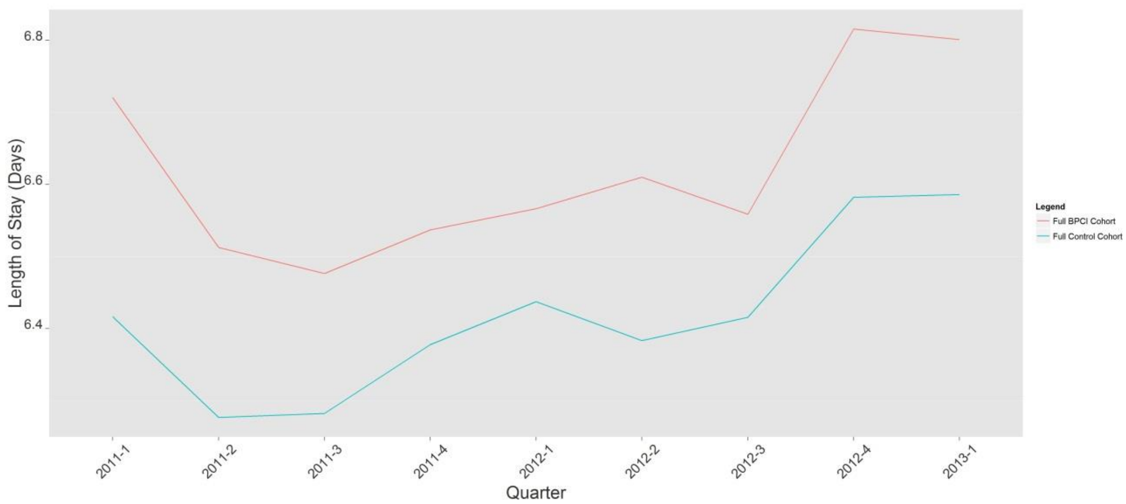
The typical method of assessing the validity of the assumption is to graph the patterns of relevant outcome variables for the treatment and comparison groups. Figures C.1 through C.4 show the pre-intervention trends for four outcome variables for the BPCI Model 1 Awardees and Comparison hospitals that include a patient’s length of stay, 30-day all-cause mortality, 30-day all-cause readmissions, and 30-day total Medicare payments over an inpatient stay. The horizontal axis indicates the quarters prior to the start of the BPCI program and run from first quarter of 2011 through the first quarter of 2013. For average inpatient payments (Figure C.1), average length of stay (Figure C.2), 30-day all-cause mortality rate (Figure C.3), and 30-day all-cause readmission rate (Figure C.4), the paths are quite parallel, and for Figure C.3 they are almost identical. This provides evidence that the parallel trend assumption is satisfied. These figures are based on unadjusted averages; however, the DiD model will control for several covariates (i.e., additional patient or hospital characteristics) and will therefore account for remaining differences between the Awardee and comparison cohorts. Indeed, within Section IV, each measure presented shows trends that are adjusted to mirror the DiD models.

Figure C.1: Average Inpatient Payments by Quarter for BPCI Model 1 Hospitals and Selected Comparison Hospitals (in Dollars)



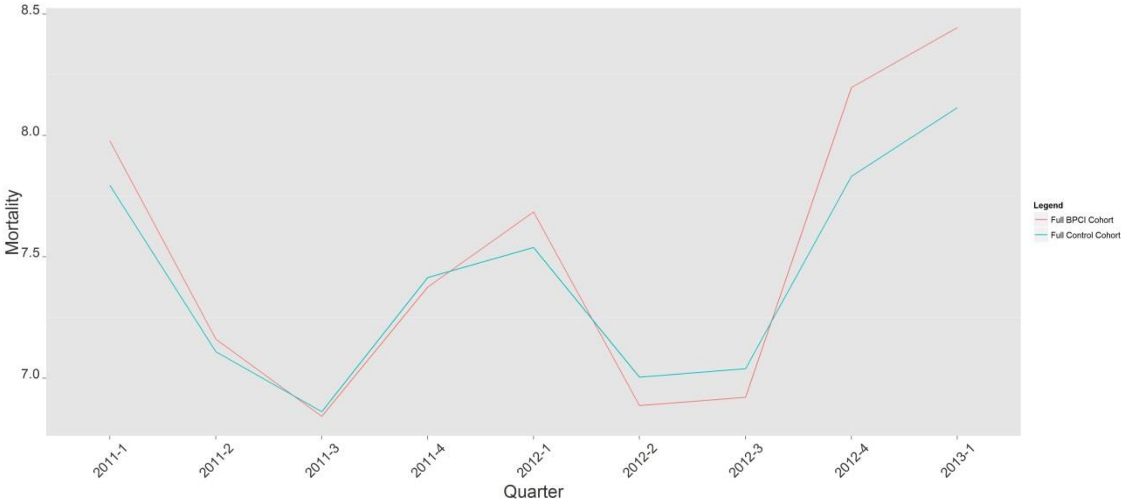
Data source: Inpatient, outpatient, carrier, home health, skilled nursing facility, hospice, and durable medical equipment claims.

Figure C.2: Average Length of Stay by Quarter for BPCI Model 1 Hospitals and Selected Comparison Hospitals (in Days)



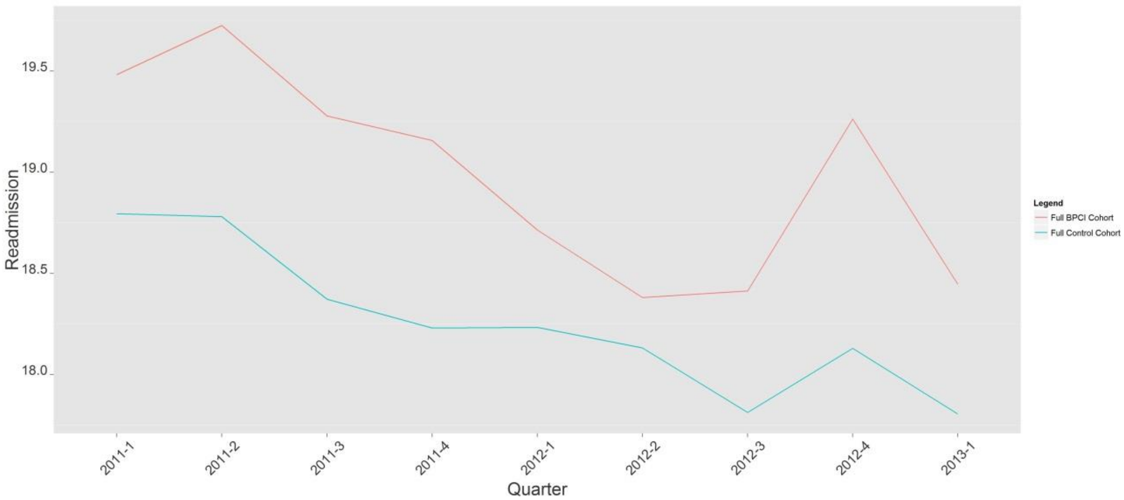
Data source: Inpatient, outpatient, carrier, home health, skilled nursing facility, hospice, and durable medical equipment claims.

Figure C.3: 30-Day All-Cause Mortality by Quarter for BPCI Model 1 Hospitals and Selected Comparison Hospitals (Rate)



Data source: Inpatient, outpatient, carrier, home health, skilled nursing facility, hospice, and durable medical equipment claims.

Figure C.4: 30-Day All-Cause Readmissions by Quarter for BPCI Model 1 Hospitals and Selected Comparison Hospitals (Rate)



Data source: Inpatient, outpatient, carrier, home health, skilled nursing facility, hospice, and durable medical equipment claims.