

Appendix B: Data and Methodology

A. Centers for Medicare & Medicaid Services (“CMS”) Cost Reports

We use Centers for Medicare and Medicaid Services Cost Reports (“Medicare Cost Reports” or “Cost Reports”) to obtain, for each hospital, aggregate estimates of Medicare and non-Medicare average inpatient revenues per discharge. These reports also provide additional information on discharges, hospital characteristics, and Medicare payments. Below is a summary of the main information provided by the Cost Reports:

- Hospital information: type of hospital, type of control, urban/rural, teaching, DSH status, number of beds, interns and residents, solid organs transplants, etc.
- Discharge data: inpatient discharges and outpatient visits (total, Medicare, and Medicaid)
- Total hospital patient charges (inpatient and outpatient) and total revenue
- Total Medicare charges, total costs, inpatient revenues, and other Medicare payments (*e.g.*, DSH, IME, SCH/MDH)
- Hospital costs: average hourly salaries, full-time equivalent staff, uncompensated care, total hospital operating cost, and balance sheet data

One limitation of the Cost Reports is that they do not provide actual payments for inpatient services from each payor. They only provide payment data on total patient services (that is, for both inpatient and outpatient services) and Medicare. Another limitation is that they provide inpatient discharges for Medicare, Medicaid, and a total for all payors. Consequently, one can only calculate discharges for an “other” category that would include third party payors, managed Medicare and Medicaid, self pay (and bad debt), and other payors. The reports are based on each hospital’s fiscal year (approximately 90% of the hospitals have fiscal years ending in June, September, or December). Observations that reported time periods other than a full year were dropped.

The data used in the regression analyses are limited to general short-term and specialty hospitals under Medicare’s Prospective Payment System (PPS) in the 50 states and the District of Columbia.¹ For other hospitals, the Cost Reports do not provide detailed data to calculate non-Medicare prices, and CMI data are not available from the CMS website.

Construction of Price Variables²

The average hospital inpatient price per discharge can be estimated as:

$$\text{IP Price (All Payors)} = (\text{Total IP Charges} * \text{Discount Factor}) / \text{Total IP Discharges},$$

where Total IP Charges include general inpatient routine care charges, intensive care charges, and inpatient ancillary services charges; the discount factor is based on the ratio of total hospital

¹ The state of Maryland has a Medicare waiver which allows the state to set hospital rates for Medicare. These hospitals are also analyzed even though they are not reimbursed under PPS.

² Hospitals lacking necessary pieces of the price variables, *e.g.*, the ratio of payments to charges (or discount factor), are excluded. Depending on the price variable, this results in the loss of between 2-3% of all hospitals in the dataset. Most if not all Kaiser Foundation Hospitals in California are excluded as their Cost Report data are insufficient to calculate inpatient charges.

revenues to total charges (inpatient and outpatient); and discharges exclude swing beds, hospice, and skilled nursing facilities.

For Medicare, the Cost Reports provide data on IP charges, payments and discharges. The average Medicare IP price per discharge is calculated as:

IP Price (Medicare) = Medicare IP Payments / Medicare Discharges

In order to obtain a measure of non-Medicare prices, we follow the formula proposed by Dafny (2007).³ This formula excludes Medicare revenues and discharges from the average IP price calculation:

IP Price (Non-Medicare) =

(Tot. IP Charges * Discount Factor – Medicare IP Payments) / Non-Medicare Discharges,

where the “Medicare IP Payments” subtracted in the numerator include DSH, IME, GME and other revenues from Medicare. Hence, all Medicare payments are excluded from the calculation of non-Medicare prices.

Removal of Extreme Values

Several key variables affecting the estimated price measures (*e.g.*, total discharges, inpatient charges, etc.) and explanatory variables (*e.g.*, beds) are scanned for extreme values, or statistical outliers. For each variable, the mean value of that variable is calculated by hospital across the 5 years in the sample. If a given observation for that hospital departs from the mean by a percentage greater than a specified threshold, that observation is replaced as missing. The mean is then recalculated and the process iterated until there are no remaining extreme values. The threshold is determined by the nature of the variable – for example, the threshold for beds is lower than that for inpatient charges since it is more likely for there to be bigger changes year to year in inpatient charges than in beds. In addition, observations with revenues per discharge of less than \$500 or greater than \$35,000 were also dropped. Depending on the variable, this process affects anywhere from effectively 0% of the observations up to approximately 2% of the observations. In each case, the observations being replaced as missing were examined to make sure that the process was detecting likely data errors rather than removing substantive data.⁴

B. CMS Hospital Inpatient Prospective Payment System (“PPS”) Final Rule Impact Files⁵

The Final Rule Impact Files (“Impact Files”) contain information for hospitals reimbursed under PPS (and those in the state of Maryland). In addition to variables specifying the location of each hospital such as region and county, information is also provided on CMI (including transfer-adjusted), beds, resident to bed ratio, DSH adjustment, and various other factors that Medicare

³ Dafny, 2009, p. 531 and fn. 13.

⁴ We note that regression analyses without removing these extreme values lead to essentially similar results, albeit with lower R-squared values due to the additional noise these extreme values add to the model variables.

⁵ For more detail, *see*: http://www.cms.gov/AcuteInpatientPPS/01_overview.asp#TopOfPage

uses to adjust PPS payments. As noted above, hospitals with no CMI data (typically non-PPS hospitals) are excluded from the analysis.⁶ Key variables that are used from the Impact Files include CMI, Wage Index, and Outlier Payments, described in further detail below.

Case Mix Index (CMI): Individual inpatient discharges are associated with a Diagnosis Related Group (DRG) that classifies patients that use similar hospital resources into the same group (e.g., DRG 7 refers to lung transplants). Each DRG has an associated DRG-weight which reflects the estimated relative costliness of patients in that DRG compared with the average Medicare patient across the country. CMI provides the average DRG weight for the hospital, calculated as the sum of all DRG weights for Medicare discharges divided by the total number of discharges.

*Wage Index:*⁷ For each labor market area, where the areas are defined as Core Based Statistical Areas (CBSAs) and statewide rural areas outside of CBSAs, CMS calculates an average hourly wage as total wage costs divided by total hours for all hospitals in the area. Similarly, a national average hourly wage is calculated. The wage index is defined as the ratio of the area's average hourly wage to the national average hourly wage. Various adjustments to this ratio can be made by CMS for hospitals that, for example, lie near a CBSA border and are determined to face a labor market more appropriately defined by a higher-cost adjacent CBSA. Adjustments are also made for occupational mix.

*Outlier Payments:*⁸ Outlier payments exist to reimburse hospitals for cases with costs that exceed the fixed-loss cost threshold amount, which is determined yearly by CMS and then adjusted by case and hospital for DRG weight, wage index, IME, DSH, etc. For each dollar in cost exceeding the determined threshold, the hospital is reimbursed at a constant rate of between 80-90%, depending on the DRG. Payments are determined separately for operating and capital costs, which are affected by the hospital's operating and capital cost-to-charge ratios and several other adjustments. The Impact Files provide data on outlier payments as a percentage of provider operating/capital PPS payments.

⁶ Approximately 2% of the observations in the Medicare Cost Reports (for short-term general and specialty hospitals) do not have CMI data in the impact files.

⁷ For more detail, see: https://www.cms.gov/AcuteInpatientPPS/03_wageindex.asp. MedPAC has proposed an alternative wage index measure that smoothes the index between counties, uses wage data from BLS/Census surveys rather than hospital cost reports, and fixes the occupational mix rather than allowing it to vary by hospital. The MedPAC index does not appear to have been made publically available. For more detail, see: http://www.medpac.gov/documents/Jun07_EntireReport.pdf.

⁸ For more detail, see: https://www.cms.gov/AcuteInpatientPPS/04_outlier.asp#TopOfPage. While the CMS impact files typically contain CMI information for Maryland hospitals, outlier payments are not present (due to the different system of reimbursement). Therefore, all regressions including outlier payments exclude all Maryland hospitals.

C. Behavioral Risk Factor Surveillance System (“BRFSS”)⁹

The Centers for Disease Control and Prevention (CDC) collects data on a variety of questions related to health risk behaviors, healthcare access, and health status. The data are collected via telephone interview from approximately 350,000 adults a year. Each record in the data is an individual survey, which can be aggregated by county, state, or any areas composed of counties and/or states. Most variables are self reported, and respondents may refuse to answer questions. Data were analyzed for survey years 2000-2009, with each variable calculated across all available surveys over the period. When aggregating to the CBSA level, all non-CBSA counties in a given state are considered part of the “rural” area of the state. The variables used are:

- Current Smoker: defined as those who have smoked 100 cigarettes in their life and smoke either “every day” or “some days” at the time of the survey
- Heavy Drinker: defined as those males who drink an estimated 60 or more drinks every 30 days, and females who drink an estimated 30 or more drinks every 30 days
- Health Checkup: time since last visit to a doctor for a routine checkup (a general physical exam not for a specific injury or illness)

D. Area Resource File (“ARF”)¹⁰

The ARF is published by the Health Resources and Services Administration (HRSA). It is a collection of data from a wide range of sources at the county level. ARF compiles information from the American Medical Association and the American Hospital Association, among other sources. Population characteristics such as income and racial makeup are compiled from Census data. When aggregating to CBSA, all non-CBSA counties in a given state are considered part of the “rural” area of the state. The variables used in the analysis typically measure these characteristics as percentages of the total population in the area – *e.g.*, the percentage of population in the CBSA under age 65 that have health insurance.

E. Quality Data

Risk-adjusted 30-day readmission and mortality rates for heart attack, heart failure, and pneumonia for Medicare beneficiaries are obtained from Hospital Compare data.¹¹ To estimate these rates, CMS uses Medicare claims data and a risk-adjustment process that controls for factors such as age, gender, and comorbidities. Mortality and readmission rates, however, are not available for all hospitals in our sample. We also included in the dataset indicator variables marking hospitals that appear in the 2010-2011 US News & World Report “Best Hospitals.”¹²

⁹ For more information, *see*: <http://www.cdc.gov/brfss/>.

¹⁰ For more information, *see*: <http://arf.hrsa.gov/>.

¹¹ For more information, *see*: https://www.cms.gov/HospitalQualityInits/11_HospitalCompare.asp.

¹² For more information, *see*: <http://health.usnews.com/best-hospitals/rankings>.

Explanatory Variables

The table below summarizes the explanatory variables used and the source of each variable (the price variables are calculated from the Medicare Cost Reports as described above). Some variables are directly available from the specified source while others are calculated.

Explanatory Variable	Source	Explanatory Variable	Source
Case-Mix Index (CMI)	Impact Files	Average Inpatient Cost per Discharge	Cost Reports
Medicare Wage Index	Impact Files	% Population Black in Hospital CBSA	ARF
Operating Outlier Payments (%)	Impact Files	% Population Hispanic in Hospital CBSA	ARF
Share of Medicare Discharges	Cost Reports	% Population in Poverty in Hospital CBSA	ARF
Share of Medicaid Discharges	Cost Reports	% Population Male in Hospital CBSA	ARF
Teaching Intensity (Interns and Residents per Bed)	Cost Reports	% Population Under Age 65 Insured in Hospital CBSA	ARF
Rural Hospital Indicator	Cost Reports	Average Age in Hospital CBSA	ARF
Share of Outpatient Charges over Total Charges	Cost Reports	% Population Current Smoker in Hospital CBSA	BRFSS
Total Hospital Assets	Cost Reports	% Population Heavy Drinker in Hospital CBSA	BRFSS
Hospital Beds	Cost Reports	% Population with a Health Checkup in Last 2 Years in Hospital CBSA	BRFSS
Total Discharges per Bed	Cost Reports	Mean of 2008 30-day Mortality Rates for Heart Attack, Heart Failure, and Pneumonia	Hospital Compare
DSH amount	Cost Reports	Mean of 2008 30-day Readmission Rates for Heart Attack, Heart Failure, and Pneumonia	Hospital Compare
Organ Transplants per Bed	Cost Reports	2010 US News & World Report Best Hospitals in any of the 16 Specialties	US News
Uncompensated Charges per discharge	Cost Reports		

Conversion of Variables to Logarithmic Form

In the literature reviewed, some researchers convert variables to logarithmic form.¹³ For ease of exposition, we convert the price variables and the explanatory variables to logarithmic form. One advantage of this approach is that the coefficients from the regression models (*see* Appendix C) can be interpreted as the percentage change in prices from a 1% change in the explanatory variable.¹⁴

¹³ Logarithms are standard functions used in various scientific formulas. One advantage of converting a variable to logarithmic form is that extreme values are “smoothed out” into smaller scopes. This reduces the possibility that the regression results are driven by a few unusual observations.

¹⁴ The time trend variable and the indicator variables (rural and US News & World Report Best Hospitals indicators) are not converted to logarithmic form.

This transformation does not generally change the explanatory power of the models with respect to the alternative level regressions.¹⁵ In a few cases, the transformation to logarithmic form changes the statistical significance of the variables. For example, mortality and readmission rates show significant correlation with hospital prices in the logarithmic regressions but not in the level regressions. The reverse occurs with the teaching intensity variable and the US News & World Report hospital rankings, which show statistically significant correlation in the level regressions but not in the logarithmic regressions.

¹⁵ The conversion to logarithmic form changes the interpretation of the R-squared values of the regressions. However, considering that the R-squared values are similar to those from level regressions with the same variables, we refer to the R-squared values as the proportion of the price variation explained by the model.