

# Measuring Hospital Ownership Transitions in the United States, 2001-2014\*

Zack Cooper  
Stuart V. Craig  
Martin Gaynor  
John Van Reenen

This document: January 8, 2019  
The data: January 8, 2019

## 1 Description of the Data

The data accompanying this document,

- HC\_ext\_mergerdata\_public.csv
- HC\_ext\_mergerdata\_public.dta,

provide the user with a time varying series of unique hospital identifiers that allow for the tracking of hospital ownership changes (e.g. mergers) from 2001–2014. We identify each hospital system that is registered with the AHA and the ID of each site within that system. We then update system and site IDs when, for example, a hospital site is sold to another system or two hospital systems merge to form an entirely new system. These files were constructed to estimate the effect of horizontal hospital mergers on negotiated prices in

**Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. “The Price Ain’t Right: Hospital Prices and Health Spending on the Privately Insured,” *Quarterly Journal of Economics*, forthcoming.**

---

\*This document was prepared by Stuart Craig ([stucraig@upenn.edu](mailto:stucraig@upenn.edu)). We thank Charles Gray for excellent research assistance.

The data provide a roster of most of the hospital mergers that occurred from 2001 to 2014. We have included latitudes and longitudes of hospital sites, so that individuals can calculate the distance between merging entities.

All citations of the data should be made to Cooper et al. (forthcoming). For the most up to date version of the data, visit <https://healthcarepricingproject.org/>.

## 2 Tracking Hospital Ownership Over Time and Identifying Hospital Mergers

The American Hospital Association Annual Survey contains data on respondent hospitals in the United States, and is often treated as a census of hospitals. While the AHA data are an invaluable source of information regarding hospital characteristics and geography, they provide an incomplete picture of hospital ownership transitions for multiple reasons. As a result, we have created a consistent panel of hospital identifiers that more accurately and comprehensively track hospital mergers over time.

First, the AHA reports data for a reference period preceding the year of the survey. As a result, system information in the AHA typically refers to the year following the reference year. In general, we deal with this issue by utilizing the lagged system information.

Second, the AHA sometimes deals with mergers and acquisitions in a way that would complicate analysis if used directly. In a case where one hospital “merges” with another, the AHA often contains a single observation for the merged entity.<sup>1</sup> However, when a system “acquires” a hospital, it sometimes retains its unique AHA ID and experiences a change in its system ID. Dealing with this discrepancy is particularly important for understanding the geographic distribution of hospital ownership because hospital demand is strongly determined by travel distance. In order to obtain a complete picture of hospital geography and ownership, we generate imputed observations for those hospitals, which are deleted from the data as a result of a merger, while noting the change in ownership structure through the system information.

We also incorporate several additional data sources to verify the existence and timing of ownership in our AHA-based series: Irving-Levin Associates, Factset, and SDC Platinum. Each database contains detailed information (e.g. parties involved, announcement and closing dates) on both completed and failed mergers and acquisitions. We use these data to corroborate the existence and timing of mergers, following the corrections noted above. In the event that there were conflicts between the our data and the supplemental databases, we used the Becker’s Hospital Review and local newspaper articles to determine the correct

---

<sup>1</sup>We use the two hospital case as an illustrative example of why the AHA might be making a distinction between the way that it treats ownership changes. However, in some cases, this type of treatment in the data may involve more than two hospitals.

tracking of hospital ownership.<sup>2</sup>

### 3 Imputing Observables for Hospitals Impacted by Consolidation

This section describes the method by which we impute observable information for hospitals that were deleted from the AHA survey or agglomerated because of mergers. In order to comply with our data use agreement, we do not include these variables in the accompanying data. However, users who already have access to the AHA data can obtain them by contacting us at [stucraig@upenn.edu](mailto:stucraig@upenn.edu).

In Section 2, we describe the process by which we “undo” the consolidation of merging firms so that we have observations even for hospitals that are “absorbed” by an acquiring firm. However, in many cases, we also wish to obtain an estimate of each firm’s observable information as reported in the AHA. For example, once we “undo” this consolidation of IDs, we do not know the correct bed count (and other observables) at the hospital site-level after consolidation.

We address this by imputing the information at the consolidated level to the site level for all continuous variables in the following manner. Consider the following example of imputing bed counts. Two separate hospitals have distinct IDs  $A$  and  $B$  and merge at time  $\tau$ , at which point they become hospital  $C$ .<sup>3</sup> Hospitals  $A$  and  $B$  cease to exist in the data from time  $\tau$  onward. Let  $b_{ht}$  denote the number of beds at hospital  $h$  at time  $t$ , where  $h \in \{A, B, C\}$  and  $t \in \{2001, \dots, 2014\}$ . Let

$$w_h = \frac{b_{h,\tau-1}}{\sum_{h \in \{A,B\}} b_{h,\tau-1}},$$

which is the share of hospital  $h$ ’s beds between hospital  $A$  and  $B$  at  $\tau - 1$ . If hospital  $C$  (in time  $\tau$ ) has a bed count that is within 20 percent of the sum of  $A$  and  $B$  at  $\tau - 1$ :

$$\frac{\left| b_{C,\tau} - \sum_{h \in \{A,B\}} b_{h,\tau-1} \right|}{\left( b_{C,\tau} + \sum_{h \in \{A,B\}} b_{h,\tau-1} \right) / 2} \leq 0.2$$

then we assume hospital  $h$ ’s bed count is  $w_h b_{C,t}$  for all  $t$  in which hospital  $C$  exists in the AHA data. Otherwise, we assume hospital  $h$ ’s bed count is  $b_{h,\tau-1}$  in all periods after the merger. We have provided our imputed bed count to facilitate the construction of hospital HHIs. However one can use this same methodology to impute any continuous variable.

---

<sup>2</sup>If the closing date was not available from any of the databases and we could not find a news article or report that documented the closing of the deal, we used the announcement date as an estimate of when the deal was completed.

<sup>3</sup>Hospital  $C$  may have already exist at  $\tau - 1$  or may be a new hospital created from the merger of  $A$  and  $B$  at  $\tau$ .

## 4 Identifying Mergers in the Data

In Section 2, we detail our method for tracking hospital ownership using the system ID variable. We identify merger events by tracking year-over-year changes to that variable over time. We provide Stata code below to verify their construction:

---

```
use HC_ext_mergerdata_public.dta, clear
assert sysid!=" "

// Target hospitals are those with changes to system ID
qbys id (year): gen lagsys = sysid[_n-1]
qui gen target2 = (sysid != lagsys) ///
                  & (lagsys != "") ///
                  & (year - 1 == year[_n-1])
tab target target2

// Acquirer hospitals are part of the destination system but
// without changes to the system ID
qui egen acquirer2 = max(target), by(sysid year)
qui replace acquirer2 = 0 if (target==1)
tab acquirer acquirer2
```

---

## 5 Codebook

<b>id</b>	Unique Hospital ID
<b>year</b>	Calendar Year
<b>sysid</b>	Corrected System Ownership ID
<b>lat</b>	Latitude
<b>lon</b>	Longitude
<b>target</b>	Indicator for Acquired Hospital
<b>acquirer</b>	Indicator for Membership in Acquiring System
<b>id_defunct</b>	Original AHA ID
<b>id_parent</b>	AHA ID of Absorbing Hospital

The data are structured as a panel of hospital IDs, which have a unique observation for each year. **sysid** contains the corrected system ownership information for each hospital as described in Section 2. The latitude and longitude variables (**lat** and **lon**) were generated using ArcGIS to look up location information for the addresses available in the AHA.<sup>4</sup> **target**

---

<sup>4</sup>The AHA contains latitude and longitude data, however this data is often incorrect in early years of the data. For a small number of cases, ArcGIS was unable to find coordinates for the address. In these cases, we used Google Maps to complete the data.

and **acquirer** contain the binary indicators for merger participation as described in Section 4. **id\_defunct** and **id\_parent** contain the original AHA ID or consolidated AHA ID respectively. We include these to facilitate merging with the AHA data.