2022 Healthy Marketplace Index

Technical Appendix

June 2022

Table of Contents

Introduction

Health care spending in the United States has risen dramatically over recent decades and is projected to continue growing into the future. While this trend holds nationally, there is an increasing body of evidence that the sources of both health care spending levels and growth vary dramatically across the country. It is therefore important to understand the factors associated with health care spending in different areas, and how these factors have and may continue to change over time.

The Healthy Marketplace Index (HMI) reports a series of metrics which can be used to assess the economic performance of local commercial health care markets. These metrics are intended to facilitate comparable and consistent assessments of health care market performance both across markets and within markets over time. These metrics are additionally intended to be transparent, both in their availability through public use files, and in their construction through a comprehensive methodological documentation.

This document describes how we use the Health Care Cost Institute (HCCI) commercial claims database to construct our HMI metrics. From HCCI data, we construct a sample containing the health care claims for individuals receiving commercial health insurance through their employer from 2016-2020 residing in one of our 186 sample metro areas across 44 states and the District of Columbia. These data contain more than 4.2 billion claims from 2016-2020 from more than 41 million individuals annually.

Using our analytic sample of claims, we construct indices of metro area health care spending (“Spending Index”), average health care service prices (“Price Index”), volume of health care services used (“Use Index”), and the cost of the mix of services used (“Service Mix Index”). We construct each of these metrics both at a metro area level, as well as by high level service category (Inpatient, Outpatient, Professional Services) in each metro. Separately, we also construct a measure of inpatient hospital market concentration for each metro area (“Concentration Index”).

We gratefully acknowledge financial support for this project from the Robert Wood Johnson Foundation.

1. Using HCCI Claims Data to Construct an Analytic Sample of Claims

Using HCCI claims data, we constructed a sample of health care services provided in geographic areas across the country in each year. The HCCI claims data are primarily organized at the claim line level. That is, for a service performed, the claim file is broken up into multiple claim lines (e.g., one claim line for a particular lab, another for a procedure). To construct a service level sample from the claim line level data, we aggregated data from all claim lines associated with each service to the claim level. This aggregated service will be referred to as a service claim. Our analytic sample consisted of cleaned service claims from enrollees residing in our sample geographic regions (regardless of where services were provided).

1.1 Defining a Sample Population of Members

Using monthly enrollment data, we constructed a sample of member-month observations. For a member month to be included in the sample population, the member, in that given month, must be under the age of 65 and have an identifiable age and gender in the data. We also limited our sample of member-months to individuals with an identifiable five-digit zip code. Further, we excluded member-months in which the individual resided in two or more metro areas in the same month.

Additionally, we restricted our analysis to member-months for individuals with coverage through an employer-sponsored insurance (ESI) plan. Specifically, we limited our sample to individuals with either small or large group commercial insurance coverage with one of the following plan types: Health Maintenance Organization, Preferred Provider Organization, Point of Service Plan, or Exclusive Provider Organization.

1.2 Assigning Member-Months to Core-Based Statistical Areas

Our geographic unit of analysis was the Core-Based Statistical Area (CBSA). Using monthly enrollment data, we mapped the five-digit zip code associated with each member-month to a CBSA. Because CBSA definitions change over time, we used a single five-digit zip-code-to-CBSA crosswalk regardless of the year so that a CBSA in our data refers to the same geographical region across time.

To construct our geographic crosswalk, we used a five-digit zip-code-to-CBSA crosswalk constructed by the United States Department of Housing and Urban Development from our base year (2016).[1] In cases where a zip code is assigned to multiple CBSAs, we assigned zip codes to the CBSA with the greatest "Total Ratio" followed by the greatest "Residential Ratio."

We also mapped five-digit zip codes to states using the National Bureau of Economic Research's "SSA to FIPS State and County Crosswalk" from our base year (2016).[2]

Member months associated with zip codes that did not match either a CBSA or state from the crosswalk were omitted. Member-months whose zip codes matched a state but not a CBSA were assigned to the CBSA "Rural – [State Abbreviation]." Member-months whose zip codes matched to areas outside of the 50 United States or DC were excluded. We assigned each CBSA to a single state based on the state with the largest share of member months observed in our sample within each CBSA.

1.3 Aggregating Claim Lines to Claim Level

Prior to aggregating claim lines, we merged on enrollment information based on the month and year in which a claim line occurred – as defined by the dates associated with each claim line. We excluded all claim lines associated with member-months that were not part of our sample population. We assigned each claim line to the CBSA and state associated with the five-digit zip code attached to the relevant member month.

We defined a service claim as all claim lines for an individual with common dates and service codes. We defined service codes distinctly in each high-level service category (inpatient, outpatient, and professional). For inpatient claims, we defined a service code as Diagnosis Related Group (DRG) codes. For outpatient and professional claims, we defined service codes as Current Procedural Terminology (CPT) codes.

When aggregating claim lines to the service claim level, we summed all allowed amounts (the actual amount paid to for the claim) from each claim line associated with a particular service claim. Allowed amounts comprise the insurer's payment to a provider. We defined the sum of the these allowed amounts as the total spending on a service claim.

---

[1] Specifically, we use the crosswalk titled "ZIP-CBSA" from the 4th quarter of 2013. Available online from the HUD website: https://www.huduser.gov/portal/datasets/usps_crosswalk.html.
[2] Available online from the NBER website: https://data.nber.org/data/ssa-fips-state-county-crosswalk.html

1.4 Cleaning Claims to Construct the Analytic Sample

We applied separate cleaning procedures to inpatient, outpatient, and professional service claims to remove outlier claims.

Inpatient Claims (Admissions)

Our unit of analysis for inpatient service claims was an inpatient admission defined by a combination of year, patient, service code (DRG), and visit dates.

We identified the inpatient facility associated with each claim by encrypted National Plan and Provider Enumeration System Identifiers (NPI). As some inpatient facilities may be assigned multiple NPIs, we mapped all associated NPIs with each facility to a single, consolidated, encrypted NPI (cNPI). Claim lines with missing consolidated cNPIs were assigned the non-missing cNPI within the admission. If a claim contained all missing cNPI values, we assigned it a cNPI of "blank." If a claim was associated with multiple, non-missing cNPIs across claim lines, the cNPI associated with the highest allowed amount was assigned to the entire claim (pcNPI).

Claims were excluded if they contained claim lines with unknown or unidentifiable DRG codes. Claim lines with missing DRG values were assigned the non-missing DRG value within the claim. Claims entirely made up of claim lines with missing DRGs were excluded. Claims with lengths of stay over 180 days were excluded as were admissions with discharge dates preceding first admission dates. Claims were excluded with allowed amounts less than 1 dollar.

Of the remaining claims, within each year and service code combination that had at least 100 claims, the allowed amounts for claims with the highest 1% and lowest 1% of allowed amounts were top coded and bottom coded, where we assigned the 99th percentile allowed amount and 1st percentile allowed amount within that combination, respectively. Additionally, claims in which the ESI plan was not the primary coverage payer were excluded.

In the HCCI data, each provider has an attached provider five-digit zip code. We mapped these provider zip codes to CBSAs and states using the same crosswalk as member zip codes. We omitted claims with an identifiable provider zip code which was not associated with one of the 50 states or D.C. If the provider zip code(s) within a claim matched to multiple CBSAs, the CBSA associated with the highest allowed amount within the claim was assigned as the provider CBSA for that claim.

Using each facility identifier (pcNPI), we merged on hospital characteristics from the American Hospital Association (AHA) in each year. Using characteristics from the AHA survey, we identified inpatient claims associated with general acute care (GAC) inpatient hospitals.[3]

Outpatient Claims (Procedures / Visits)

Our unit of analysis for outpatient claims was the visit or procedure defined by the combination of year, patient, visit dates, and CPT code level.

We identified the provider associated with each claim by encrypted National Plan and Provider Enumeration System Identifiers (NPI). Each claim was assigned a single NPI based on the non-missing NPI value associated with the greatest allowed amount among the visit claim lines. If the greatest allowed amount NPI value for a claim was tied between two or more non-missing NPIs, the NPI that was associated with a greater out-of-pocket payment or claim lines (in that order) was assigned to the claim. If a claim contained all missing NPI values, we assigned it a NPI of “blank.”

The non-negative allowed amounts and out-of-pocket payments associated with claims with missing CPT codes (which could therefore not be assigned a service code) were evenly distributed to claims with non-missing CPT codes for the same patient on the same service dates. If all service codes were missing within a given patient and service date combination, the claims were excluded.

Claims with lengths of stay greater than seven days were excluded (difference between first and last dates). Claims with allowed amounts less than 1 dollar were excluded. Of the remaining claims, within each year and service code combination that had at least 100 claims, the allowed amount for claims with the highest 1% and lowest 1% of allowed amounts were top coded and bottom coded, where we assigned the 99th percentile allowed amount and 1st percentile allowed amount within that combination, respectively. Additionally, claims in which the ESI plan was not the primary coverage payer were excluded.

In the HCCI data, each provider has an attached provider five-digit zip code. We mapped these provider zip codes to CBSAs and states using the same crosswalk as member zip codes. We omitted claims with an identifiable provider zip code which was not associated with one of the 50 states or D.C. If the provider zip code(s) within a claim matched to multiple CBSAs, the

[3] Specifically, we defined GAC hospitals as those with relevant control (excluding federal government facilities) and service codes (general medical and surgical).

CBSA associated with the highest allowed amount within the claim was assigned as the provider CBSA for that claim.

Professional Claims (Procedures / Visits)

Our unit of analysis for professional claims was the visit or procedure defined by the combination of year, patient, visit dates, and CPT code level.

We identified the provider associated with each claim by encrypted National Plan and Provider Enumeration System Identifiers (NPI). Each claim was assigned a single NPI based on the non-missing NPI value associated with the greatest allowed amount among the visit claim lines. If the greatest allowed amount NPI value for a claim was tied between two or more non-missing NPIs, the NPI that was associated with a greater out-of-pocket payment or claim lines (in that order) was assigned to the claim. If a claim contained all missing NPI values, we assigned it a NPI of “blank”.

Claims were excluded if the CPT code was missing and could therefore not be assigned a service code. Claims with lengths of stay greater than seven days were excluded (difference between first and last dates).

Claims with allowed amounts less than 1 dollar were excluded. Of the remaining claims, within each year and service code combination that had at least 100 claims, the allowed amount for claims with the highest 1% and lowest 1% of allowed amounts were top coded and bottom coded, where we assigned the 99th percentile allowed amount and 1st percentile allowed amount within that combination, respectively. Additionally, claims in which the ESI plan was not the primary coverage payer were excluded.

In the HCCI data, each provider has an attached provider five-digit zip code. We mapped these provider zip codes to CBSAs and states using the same crosswalk as member zip codes. We omitted claims with an identifiable provider zip code which was not associated with one of the 50 states or D.C. If the provider zip code(s) within a claim matched to multiple CBSAs, the CBSA associated with the highest allowed amount within the claim was assigned as the provider CBSA for that claim.

Summary Statistics of our Analytic Sample

Our analytic sample include more than 4.2 billion claims from 2016-2020 across more than 170 million member years. This includes an average of more than 1.9 million inpatient claims, 150 million outpatient claims, and 690 million professional claims, annually.

These claims are attributed to an average of more than 41 million individuals annually.

Our analytic sample comprises a consistent subset of all HCCI claims across our study period. As seen in Table 1.1, our sample includes between 78% and 79% of the universe of HCCI claims in each year, and about 85% of total spending (defined as the sum of allowed amounts). Our analytic sample captures a slightly higher percentage of professional claims (81% of claims) than inpatient and outpatient claims (between 66% and 70% of claims).

Table 1.1. Share of HCCI Universe of Claims, Spending Included in Analytic Sample by Year

Service Category	2016	2017	2018	2020	2020
Share of Claims					
Overall	79%	78%	79%	78%	79%
Inpatient	66%	66%	69%	69%	70%
Outpatient	70%	69%	70%	70%	70%
Professional	81%	81%	81%	81%	81%
Share of Spending (Total Allowed Amounts)					
Overall	86%	84%	84%	85%	86%
Inpatient	85%	84%	97%	88%	87%
Outpatient	92%	92%	91%	92%	91%
Professional	81%	79%	78%	79%	80%

2. Constructing Spending, Use, Price, Service Mix Indices

We computed each of the following metrics using the set of claims in our analytic data set (C).

2.1 Measuring Total Spending and Use at the CBSA-Service, CBSA, and National Levels

Defining Spending, Use at the CBSA-Year-Service Level

For a given CBSA-year-service combination, we defined total spending (y_{gts}) as the sum of allowed amounts on all claims *c,* for service *s,* in year *t,* for all residents of CBSA *g*:

$$y_{gts} = AllowedAmt_{gts} = \sum_{c \in C_{gts}} AllowedAmt_{gtsc}$$

For a given CBSA-year-service combination we defined use as the number of claims *c*, for service *s*, in year *t,* for all residents of CBSA *g*:

$$u_{gts} = \sum_{c \in C_{gts}} 1\,;\; u_{gts} \geq 0$$

Defining Total Spending and Use Across Services at the CBSA-Year Level

We defined total spending and use by a CBSA-year combination as the sum of spending on and use of (respectively) each service *s*, in year *t*, for all residents of CBSA *g*:

$$y_{gt} = \sum_{s \in S_{gt}} y_{gts}\,;\; u_{gt} = \sum_{s \in S_{gt}} u_{gts}$$

Here S_{gt} is the subset of services *S* observed in CBSA *g* in year *t:*

$$S_{gt} = \{s | u_{gts} > 0\}$$

Defining Total Spending, Use (Across Services) at the National Level

We defined total spending on and use of our sample set of services *S* nationally in a given year as the sum of spending on and use of (respectively) each service *s,* in year *t,* for all residents of CBSA *g* across all CBSAs in our set of sample CBSAs *G*:

$$y_t = \sum_{g \in G} \sum_{s \in S_{gt}} y_{gts}\;;\; u_t = \sum_{g \in G} \sum_{s \in S_{gt}} u_{gts}$$

2.2 Measuring Price at the CBSA-Service Level

Defining Average Price at the CBSA -Service Level

Given these definitions of spending and use, we can re-write spending on service *s* observed in CBSA *g* in year *t* as the product of spending per claim (average price) and the number of claims (use):

$$y_{gts} = \left(\frac{y_{gts}}{u_{gts}}\right) u_{gts} = \bar{p}_{gts} u_{gts}$$

This allows us to define the average price of a service *s* observed in CBSA *g* in year *t* as total spending on that service divided by its use:

$$\bar{p}_{gts} = \frac{y_{gts}}{u_{gts}}$$

Imputing Average Prices at the National-Service Level for Missing Observations

If there were no observations for service *s* nationally in year *t*, we imputed the price for service *s* by inflating (or deflating) the price from when the service first (or most recently) appeared.

Spending on and use of all services included in our analytic sample were summed within each year. We then divided the resulting spending summation by the use summation to calculate a yearly average national price. The percent change in the yearly price in year *t* compared to yearly price in year *[t-1]* was then used as the national inflator weight in year *t*. The inflator weight was set to 1.0 in the base year 2016. Conversely, the percent change in the yearly price in year *t* was compared to the yearly price in year *[t+1]* was used as the national deflator weight in year *t*. The deflator weight was set to 1.0 in the final year 2020.

If there were no claims for service *s* nationally in year *t*, the average price for that service *s* in year *[t-1]* was multiplied by the inflator weight for year *t* to impute the national average price for service *s* in year *t*. If there were also no claims for service s nationally in year *[t-1]*, the average price in for that service *s* in year *[t+1]* was multiplied by the deflator weight in year *t* to impute the national average price for service *s* in year *t*.

Note, that as the use of any imputed national price will remain 0, no imputed price will affect national spending or national utilization.

Imputing Average Prices at the CBSA-Service Level for Missing Observations

If there were no observations for service *s* in CBSA *g* in year *t*, we imputed the price as the adjusted national average price for that service. In particular, we imputed the price of service $\hat{p_{gts}}$ as the national average price for that service deflated by the ratio of the weighted average of prices in CBSA *g* for the services *s* we do observe and the weighted average of prices nationally for that same set of services:

$$\hat{p}_{gts} = \bar{p}_{ts} * \frac{\sum_{s\prime \in S_{gt}^{f}} \bar{p}_{gts} * w_s^f}{\sum_{s\prime \in S_{gt}^{f}} \bar{p}_{ts} * w_s^f}$$

Using this method of imputing prices for missing CBSA-year-service observations, we defined an adjusted price for each service *s* in CBSA *g* in year *t* as follows:

$$\tilde{p}_{gts} = \begin{cases} \bar{p}_{gts} \; if \; u_{gts} > 0 \\ \hat{p}_{gts} \; if \; u_{gts} = 0 \end{cases}$$

Note that since $\tilde{p}_{gts} \neq \bar{p}_{gts} \Leftrightarrow u_{gts} = 0$, imputing prices for missing CBSA-year-service observations does not change the total spending on any service *s* in any year *t* for residents of any CBSA *g*:

$$\bar{p}_{gts} * u_{gts} = \tilde{p}_{gts} * u_{gts} = \begin{cases} \bar{p}_{gts} * u_{gts} \; if \; u_{gts} > 0 \\ 0 \; if \; u_{gts} = 0 \end{cases}$$

As a result, imputing prices for missing observations does not change our observed total spending at the CSBA-year or national-year level:

$$y_{gt} = \sum_{s \in S} y_{gts} = \sum_{s \in S_{gt}} \bar{p}_{gts} * u_{gts} = \sum_{s \in S} \tilde{p}_{gts} * u_{gts}$$

$$y_t = \sum_{g \in G} \sum_{s \in S} y_{gts} = \sum_{g \in G} \sum_{s \in S_{gt}} \bar{p}_{gts} * u_{gts} = \sum_{g \in G} \sum_{s \in S} \tilde{p}_{gts} * u_{gts}$$

2.3 Weighting Service Use

In the 2022 HMI, we included all services that existed in any year of our analytic sample in our calculations of use, spending, and price. Previously, we had limited our analyses to a limited basket of commonly utilized services that existed in every sample year. Expanding to all services, however, allows us to capture as much of the commercial health care market in each metro area as possible and to paint a more complete and accurate depiction of each geography.

As described above, if a service did not exist nationally in a given year, the price for that service was imputed using a national inflator (or deflator) weight. However, the spending and use on that service remain unchanged at zero.

We included all non-imputed services present in one or more years between 2016 and 2020. In Table 2.1, we show the number of unique services both overall and by category of service (inpatient, outpatient, professional) across all years during the study period, as well as the proportion of the total number of claims in each category of service. The distribution of services remained relatively constant over the five sample years, both overall and by category. Overall, there were more than 30,000 unique services, with nearly 15,000 outpatient codes incorporated as unique services, more than 16,000 professional codes, and nearly 800 inpatient codes. We included claims with the following place of service codes: 11, 20, 21, 22, 23, 24, 81.

Table 2.1. Distribution of Analytic Sample Claims Included in HMI Analysis by Year

Service Category	Number of Services	2016	2017	2018	2020	2020
Overall	32,036	859,437,264	854,932,051	852,559,780	876,238,807	784,151,259
Inpatient	788	0.2%	0.2%	0.2%	0.2%	0.2%
Outpatient	14,950	18.2%	18.1%	18.1%	18.0%	17.3%
Professional	16,298	81.5%	81.6%	81.7%	81.8%	82.5%

We assigned a weight to each service within each service category based on the share of claims they accounted for within each service category nationally across all years, year *T* through year *[T + 4]*. More formally, the weight for service *s* is its share of all services used *nationally* (across CBSAs *g*) among services in category *f* (S^f) across all 5 sample years *T*:

$$w_{T_s}^{f} = \frac{u_{ts}}{u_T} = \frac{\sum_{g \in G} u_{gTs}}{\sum_{s \in S^f} \sum_{g \in G} u_{gTs}}$$

Here, u_{gTs} is defined as in Section 2. 1.

In addition, we assigned each category of service a weight, which we used when computing an "Overall" price index. When we constructed our price index, we used these weights to compute a weighted-average price across services within each service category (i.e., across different inpatient services). This enabled us to compare prices for the same market basket of services across geographies and over time.

2.4 Constructing CBSA Level Indices by Category

We constructed Per Capita Spending, Per Capita Use, Price, and Case Mix Indices at the CBSA-Service Category level in each year of our data. Where relevant, we designated our base year *T* as 2016.

Defining Member Years at the CBSA-Year Level

For a given CBSA-year combination, we defined member years as the sum of member months *m* attributed to CBSA *g* in year *t* divided by 12:

$$Mem.Years_{gt} = \frac{1}{12} * \sum_{m \in M_{gt}} 1$$

Here, M_t represents the set of all individuals *i* in our sample population attributed to CBSA *g* in year *t*.

Spending Index

Our spending index measures the average spending among individuals in each CBSA in each year. To standardize area-level population differences, we computed our spending index as a per-capita metric. Within each category of services *f*, we defined per-capita spending as the sum of total spending across services *s* divided by the number of member years in CBSA *g* in year *t:*

$$Y_{gt}^{f} = \sum_{s \in S^f} \frac{y_{gts}}{Mem.Years_{gt}}$$

Use Index

Our use index measures the average volume of health care services used by individuals in each CBSA in each year. To standardize area-level population differences, we computed our use index as a per-capita metric. Within each category of services *f*, in each CBSA *g* in year *t*, we define per-capita service use as a simple count of claims per capita across services *s:*

$$U_{gt}^{f} = \sum_{s \in S^f} \frac{u_{gts}}{Mem.Years_{gt}}$$

Price Index

Our price index measures the “price” (average spending per service) paid by members residing in each CBSA in each year. To provide a standardized comparison across areas and over time, we measured the average price if each CBSA used services at in national proportions across the five-year sample. For each category of services *f*, in each CBSA *g* in year *t* we computed our price index using a weighted average of the average prices across services:

$$P_{gt}^{f} = \sum_{s \in S^f} \tilde{p}_{gts} * w_{Ts}^{f}$$

The weights utilized in the price index, applied to all 186 reported CBSAs for all three categories of service, created a single weight for each service that existed in any sample year across all sample years. Thus, the weight for service *s* is equal to the total use of service *s* across all five years divided by all use of all services in all five years.

Service Mix Index

Our service mix index measures whether CBSAs use, on average, a more or less expensive mix services than our national basket. For each CBSA *g* in year *t*, we defined our service mix index as the difference between our spending index and the product of our price and use indices.

$$M_{gt}^{f} = Y_{gt}^{f} - \left(P_{gt}^{f} * U_{gt}^{f}\right)$$

Within each category of services, our use index measures the volume of services used by each CBSA in each year. Our price index measures the average cost per service each CBSA would pay in each year if they used services in national proportions across all five sample years. Therefore, the product of our price and use indices captures expected total spending in each CBSA in each year, assuming that services were used in the same proportion as nationally across all 5 years given the average price paid for each service and volume of services used:

$$iY_{gt}^{f} = P_{gt}^{f} * U_{gt}^{f}$$

The difference between the actual spending observed and expected spending represents the additional cost of or savings from the mix of services each CBSA used in each year relative to national use:

$$M_{gt}^{f} = Y_{gt}^{f} - iY_{gt}^{f} = Y_{gt}^{f} - \left(P_{gt}^{f} * U_{gt}^{f}\right)$$

Re-writing our expression for our spending index, we can derive an expression for our mix index at the service category level. Our mix index measures the cost of using a different mix of services than the nation by the average price per service multiplied by the difference between the share of all claims on service s in CBSA *g* in year *t* and the share of all claims on service *s* nationally in our base year *T:*

$$M_{gt}^{f} = \sum_{s \in S^f} \frac{u_{gts}}{Mem.Years_{gt}} * \sum_{s \in S^f} \tilde{p}_{gts} * \left(w_{gts}^{f} - w_{Ts}^{f}\right)$$

Here w_{Ts}^{f} is defined as above and w_{gts}^{f} represents an analogous CBSA-year specific weight for service *s*, the share of claims on service *s* among all sample services in category *f* in CBSA *g* in year *t*:

$$w_{gts}^{f} = \frac{u_{gts}}{\sum_{s \in S^f} u_{gts}}$$

2.5 Constructing CBSA-Level Indices Across Service Categories

Using our service category indices, we constructed overall indices across service categories as follows:

Per Capita Spending Index

Our overall spending index captures total spending per capita on medical services across service categories. We defined total per capita spending as the sum of per capita spending across service categories *f* in each year *t:*

$$Y_{gt} = \sum_{f \in F} Y_{gt}^{f}$$

Per Capita Use Index

Our overall use index captures total medical service use across service categories. We defined total services used per capita as the sum of services per capita used across service categories *f* in each year *t:*

$$U_{gt} = \sum_{f \in F} U_{gt}^{f}$$

Price Index

Our overall price index captures the average spending per service across service categories. We calculated an overall price index value as a weighted average of CBSA price index values across service categories in each year *t:*

$$P_{gt} = \sum_{f \in F} P_{gts}^{f} * w_{T}^{f}$$

Here, the weights assigned correspond to the share of total services across all years accounted for by each service category *f* across all years nationally:

$$w_{T}^{f} = \frac{U_{T}^{f}}{\sum_{f \in F} U_{T}^{f}} = \frac{\sum_{s \in S^f} \sum_{g \in G} u_{gTs}}{\sum_{f \in F} \sum_{s \in S^f} \sum_{g \in G} u_{gTs}};\ U_{T}^{f} = \sum_{g \in G} U_{gt}^{f}$$

Note that this measure is equivalent to defining our overall set of services as the union of each of our category-level sets of services. We computed our price index as the weighted average prices across service categories:

$$P_{gt} = \sum_{f \in F} P_{gts}^{f} * w_{T}^{f} = \sum_{f \in F} \left(\sum_{s \in S^f} \tilde{p}_{gts} * w_{Ts}^{f} \right) * w_{T}^{f} = \sum_{s \in \tilde{S}} \tilde{p}_{gts} * w_{s}$$

Where for each service *s* within each service category *f*:

$$w_{S} = w_{Ts}^{f} * w_{T}^{f} = \left(\frac{\sum_{g \in G} u_{gTs}}{\sum_{s \in S^f} \sum_{g \in G} u_{gTs}} \right) \left(\frac{\sum_{s \in S^f} \sum_{g \in G} u_{gTs}}{\sum_{f \in F} \sum_{s \in S^f} \sum_{g \in G} u_{gTs}} \right) = \frac{\sum_{g \in G} u_{gTs}}{\sum_{s \in S} \sum_{g \in G} u_{gTs}}$$

Service Mix Index

Our Service Mix index captures the degree to which the difference in spending between a particular CBSA is driven by *which* services a CBSA uses. In other words, the degree to which spending in a CBSA is higher (or lower) than the national median due to the use of more (or less) expensive services, on average. More specifically, for each CBSA *g* in year *t*, we defined our mix index as the difference between our spending index and the product of our price and use indices:

$$M_{gt} = Y_{gt} - \left(P_{gt} * U_{gt}\right)$$

Re-writing our expression for our spending index we can derive an expression for our overall mix index:

$$M_{gt} = \underbrace{\sum_{f\in F} U_{gt}^{f} * \sum_{f\in F} P_{gts}^{f} * \left(w_{gt}^{f} - w_{T}^{f}\right)}_{\text{Cost of Mix Across Categories}} + \underbrace{\sum_{f\in F} M_{gt}^{f}}_{\text{Cost of Mix Within Categories}}$$

Where:

$$w_{gt}^{f} = \frac{U_{gt}^{f}}{\sum_{f\in F} U_{gt}^{f}}$$

Here, our mix index incorporates the cost of both the mix of services within service categories as well as across service categories. The cost of mix within service categories is captured as the sum of within category mix indices across categories. The cost of mix across service categories is analogously defined, i.e., the average price of each service category multiplied by the deviation in the share of services attributed to each service category in a given CBSA-year observation relative to our national set of services, scaled by the volume of services used.

2.6 Constructing CBSA-Level Indices Across Service Categories

Reporting Index Values: Level of Spending / Price / Use / Mix

For each index – both within and across service categories – we reported indices as deviations from the national median. Note that we refer to the national median as the median among our sample of reportable CBSAs and states for the relevant geographic metric (e.g., CBSA index values, state index values, respectively).

For our spending, price, and use indices, we reported index values as percent deviations from the national median. For example, for service category *f*, in CBSA *g*, in year *t*:

Spending Index:

$$\frac{Y_{gt}^{f} - \bar{Y}_{t}^{f}}{\bar{Y}_{t}^{f}}$$

Use Index:

$$\frac{U_{gt}^{f} - \bar{U}_{t}^{f}}{\bar{U}_{t}^{f}}$$

Price Index:

$$\frac{P_{gt}^{f} - \bar{P}_{t}^{f}}{\bar{P}_{t}^{f}}$$

Service Mix Index:

For our service mix index, we reported a slightly different metric because the median service mix is close to zero. This property is intuitive and by construction. The mix index essentially measures the cost (positive or negative) associated with deviating in the mix of services used from the mix of services used by the entire country. On average, CBSAs tend to use services in the similar proportion as the nation. Deviations from this basket – both in using a more and less expensive mix of services – should therefore be balanced around zero.

As a result, we alternatively reported our service mix index as the percentage deviation in CBSA spending from the national median attributable to using a different mix of services.

$$\frac{M_{gt}^{f} - \bar{M}_{t}^{f}}{\bar{Y}_{t}^{f}}$$

This metric captures the degree to which spending in a particular metro is different from the national median solely due to mix of services that CBSA uses, that is, the proportion of all services used accounted for by each individual service relative to the nation as a whole. A value of 10%, for example, indicates that in that CBSA individuals used a higher proportion of more expensive services than the nation as a whole. As a result, this mix of services used increased spending by 10% relative to the national median. To see this, note that the difference in spending between CBSA *g* and the national median in year *t* is as follows:

$$Y_{gt}^{f} - \bar{Y}_{t}^{f} = \left(P_{gt}^{f} * U_{gt}^{f} + M_{gt}^{f}\right) - \left(\bar{P}_{t}^{f} * \bar{U}_{t}^{f} + \bar{M}_{t}^{f}\right) = \left(P_{gt}^{f} * U_{gt}^{f} - \bar{P}_{t}^{f} * \bar{U}_{t}^{f}\right) + \left(M_{gt}^{f} - \bar{M}_{t}^{f}\right)$$

Here, we can re-write the difference in spending between CBSA *g* and the median CBSA as the sum of the difference between implied spending – assuming a CBSA used the national composition of services given prices paid and volume of services used (the product of price and use indices) – in CBSA *g* and the median CBSA, and the difference in the cost of the actual service mix between CBSA *g* and the median CBSA.

Hence, if we assumed the price and use levels in CBSA *g* were equivalent to the national median, the remaining difference in spending would be equal to the difference between the service mix index for CBSA *g* and the national median:

$$P_{gt}^{f} = \bar{P}_{t}^{f}, U_{gt}^{f} = \bar{U}_{t}^{f} \Rightarrow Y_{gt}^{f} - \bar{Y}_{t}^{f} = M_{gt}^{f} - \bar{M}_{t}^{f}$$

We can therefore express the difference between per capita spending in CBSA *g* from the national median attributable to the mix of services used in CBSA *g* as follows:

$$\frac{M_{gt}^{f} - \bar{M}_{t}^{f}}{\bar{Y}_{t}^{f}}$$

Note that because we calculated median spending, price, and use measures separately as sample medians, the CBSA with the median price, use, and spending values were not necessarily the same, both across and within service categories. Consequently, the metric we reported does not perfectly match the assumptions laid out. For example, in the CBSA with median inpatient spending, price and use were not necessarily equal to the national median (as assumed above). To understand how this affects our analysis, we constructed a composite median CBSA spending using the median price, use and mix index value for each service category. We subsequently compared how the percent of spending deviation from the national median used to report our mix index differed whether we used actual median spending or composite median spending. Across service categories, the two set of metrics were perfectly (positively) correlated and had nearly identical distributions. Consequently, we chose the simpler method of reporting our mix index (reporting our mix index using sample medians for spending, price, and use).

For each of these indices, we reported overall index values (across service categories) analogously.

Reporting Index Values: Change in Spending / Price / Use / Mix over Time

For our spending, price, and use indices, we reported changes in index values as percent changes from our base year. For example, for service category *f*, in CBSA *g*, in year *t*:

Spending Index:

$$\frac{Y_{gt}^{f} - Y_{T}^{f}}{Y_{T}^{f}}$$

Use Index:

$$\frac{U_{gt}^{f} - U_{T}^{f}}{U_{T}^{f}}$$

Price Index:

$$\frac{P_{gt}^{f} - P_{T}^{f}}{P_{T}^{f}}$$

Service Mix Index:

For our service mix index, we reported the percent change in per capita spending accounted for by service mix from our base year to the current year. For example, for service category *f*, in CBSA *g*, in year *t*:

$$\frac{M_{gt}^{f} - M_{gT}^{f}}{Y_{gT}^{f}}$$

To see this, note that the percent change in per capita spending in CBSA g from year *T* to year *t* can be written as follows:

$$\frac{Y_{gt}^{f} - Y_{gT}^{f}}{Y_{gT}^{f}} = \frac{\left(P_{gt}^{f} * U_{gt}^{f} + M_{gt}^{f}\right) - \left(P_{gT}^{f} * U_{gT}^{f} + M_{gT}^{f}\right)}{Y_{gT}^{f}} = \left[\frac{P_{gt}^{f} * U_{gt}^{f} - P_{gT}^{f} * U_{gT}^{f}}{Y_{gT}^{f}}\right] + \left[\frac{M_{gt}^{f} - \bar{M}_{t}^{f}}{Y_{gT}^{f}}\right]$$

The latter term, therefore, represents the degree to which spending in CBSA *g* changed solely due to changes in the mix of services CBSA *g* used over time.

For each of these indices, we reported overall index values (across service categories) analogously.

Geographic Levels of Reporting: CBSA Sample Inclusion Criteria

We calculated and reported all of our metrics using data from the 186 CBSA observed in our analytic sample, rather than the universe of CBSAs. Similarly, when computing national medians, we only did so among our subset of sample CBSAs.

The CBSAs included in the study had to meet certain population, coverage, and utilization criteria. First, the sample CBSAs had to have a minimum average HCCI coverage of 10% over the 5-year period (2016-2020). Yearly HCCI coverage estimates were calculated by dividing HCCI's member years (total member months divided by 12) within a CBSA by the American Community Survey (ACS) 5-year average employer sponsored insurance (ESI) population in that same CBSA. Each sample CBSA had to have an average of at least 25,000 member years in the HCCI data over the 2016-2020 period. Using data from the American Hospital Association (AHA), included CBSAs had to have a minimum of 5 distinct, non-governmental General Medical and Surgical Hospitals. This resulted in a final geographic sample of 186 CBSAs across 44 states, as well as the District of Columbia.

Geographic Levels of Reporting: State Level Metrics

In addition to CBSA metrics, we also reported state level metrics. These are computed analogously to our CBSA level measures. However, for our state level metrics, we treated the geographic unit of analysis as the state rather than the CBSA. The states reported meet the same reporting criteria as CBSAs.

3. Constructing the Inpatient Hospital Market Concentration Index

Limiting our full sample only to those claims that occurred at an inpatient facility, we constructed a Herfindahl-Hirschman Index (HHI) measure at the CBSA level. These index values are intended to provide descriptive, relative comparisons of the inpatient facility market concentration between the CBSAs within our sample. A high HHI level indicates high market concentration in the area, which typically signifies a lack of local market competition.

3.1 Defining a CBSA Hospital Market

Our concentration measure is best thought of as a "patient-flow" HHI where we treat the market as the set of hospital systems at which patients from a particular CBSA received care. More formally, for CBSA *g* in year *t*, we consider the market to consist of all hospitals to which individuals who reside in CBSA *g* in year *t* are admitted.

Previous work has argued that "patient-flow" concentration measures are more robust than alternative geographic market definitions than a "geographic-based" concentration measures – where a market is defined as all providers located within a geographic area.[4]

It is important to note that our HHI measures for each CBSA were calculated based on market definitions that were not chosen to represent product markets suitable for regulatory or antitrust enforcement purposes. Rather, our geographic measures were chosen to weigh both the relevance of our HHI measure to a broad spectrum of research and policy evaluations and our ability to publicly report an HHI measure at a local level. As such, our measures should not be used or interpreted to inform regulatory or antitrust conclusions.

3.2 Calculating the hospital system level Herfindahl-Hirschman Index

Given our market definition, we computed an HHI measure as the sum of squared hospital system shares of inpatient admissions for individuals from CBSA *g* in year *t*. To do so, we first counted admissions for each hospital *h* in system *s* for members from CBSA *g* in year *t* for the set of hospitals to which they are admitted:

$$Admit_{hsgt} = \sum_{a \in A_{hsgt}} 1_{ahsgt}$$

[4] John Graves, "Defining Markets for Health Care Services," health-care-markets, accessed August 20, 2020, https://graveja0.github.io/health-care-markets/.

Here, A_{hsgt} represents the set of admissions at hospital *h* in system *s* for individuals for members from CBSA *g* in year *t*.

Next, we sum the count of admissions for each hospital *h* in system *s* for members from CBSA *g* in year *t* to the system level:

$$Admit_{sgt} = \sum_{h \in H_{sgt}} 1_{hsgt}$$

Here, H_s represents the set of all hospitals *h* admitting members from CBSA *g* belonging to system *s* in year *t*.[5] If a hospital does not belong to a system, we treat the hospital as its own system ($Admits_{sgt}$ = $Admits_{hsgt}$).

Finally, we can count all admissions for members from CBSA *g* in year *t*:

$$Admit_{gt} = \sum_{a \in A_{gt}} 1_{agt}$$

Note that the set of admissions for members from CBSA *g* in year *t* (A_{gt}) is equivalent to the union of sets of admissions for each hospital *h* in system s for individuals from which members of CBSA *g* are admitted in year *t*:

$$A_{gt} = \bigcup_{h \in H_{gt}} A_{hsgt}\,; H_{gt} = \bigcup_{s \in S_{gt}} H_{sgt}$$

Here, S_{gt} represents the set of hospital systems containing at least one hospital which admits a member from CBSA *g* in year *t*.

Given these system level admission counts, we can compute our HHI measure as the sum of squared system shares of admissions for members from CBSA *g* in year *t:*

$$HHI_{gt} = \sum_{s \in S_{gt}} \left[\frac{Admits_{sgt}}{Admits_{gt}}\right]^2$$

As before, S_{gt} represents the set of hospital systems containing at least one hospital which admits a member from CBSA *g* in year *t*.

[5] This is equivalent to expanding the set of hospitals *h* in system *s* to include all hospitals in system *s* inyear *t,* regardless of whether they admit individuals from CBSA *g*.

4. Constructing Community and Social Factors

4.1 Defining the Minority Health Social Vulnerability Index

The Minority Health Social Vulnerability Index (SVI) was developed in 2021 through a collaboration between Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry Social Vulnerability Index (CDC/ATSDR SVI). Derived from publicly available data from the 2018 U.S. Census Bureau American Community Survey (5-year estimates), CDC, Department of Homeland Security, and Institute for Health Metrics and Evaluation. The index, which is expanded from the CDC's SVI, includes 15 social factors divided into 6 themes (Socioeconomic Status, Household Composition and Disability, Minority Status and Language, Housing Type and Transportation, Health Care Infrastructure and Access, Medical Vulnerability), and an overall score. Values range from 0 (indicating a less vulnerable community) to 1 (indicating a more vulnerable community). Detailed documentation of the methodology can be found here).

For the 2022 HMI, we downloaded the Minority Health Social Vulnerability rankings at the county level and cross walked each county to a CBSA in the HMI dataset. All counties in the SVI that were not located within an HMI CBSA were dropped. All CBSAs in Puerto Rico were also dropped from the dataset, due to missing information in the SVI. Because most CBSAs contain more than one county, we created a population weight for each SVI, to account for the different county sizes within the CBSA. The population weight is the share of a CBSA's population (*pop*) in a given county (*x*). These values are multiplied by the county's SVI value, and summed across all counties in a CBSA to generate the weighted CBSA-level SVI value ($\mathrm{SVI}_{\mathrm{CBSA}}^{\mathrm{w}}$):

$$\mathrm{SVI}_{\mathrm{CBSA}}^{\mathrm{w}} = \sum_{\mathrm{x} \in \mathrm{CBSA}} \frac{pop_{\mathrm{x}}}{pop_{\mathrm{CBSA}}} \cdot \mathrm{SVI}_{\mathrm{x}}$$

The result is a CBSA-level file with an SVI value for each of the six themes and an overall composite score. We include additional variables relevant to understanding the SVI, including the total population within a CBSA, the county count, and the share of the population that had employer-sponsored insurance, to understand how representative HMI data are of the social factors within a community.

4.2 Defining the Share of a CBSA's Population with Employer-Sponsored Insurance

HCCI data may vary in their representation of a CBSA's population, since individuals with employer-sponsored insurance (ESI) tend to be underrepresented in areas with higher social vulnerability. In particular, this metric provides broader context for the HMI health care use and spending indices, since ESI data may be differently representative across areas. A higher share of the population with ESI suggests that HCCI data are more representative of the population.

We used the U.S. Census American Community Survey to estimate the number of individuals within each CBSA that were represented in the HMI data in 2020 (i.e. those that had ESI only) (Table B27010). Specifically, we divided the sum of individuals with ESI across each age band (variables B27010_004, B27010_020, B27010_036, and B27010_053) by the sum of the estimated population in each age band (variables B27010_002, B27010_018, B27010_034, and B27010_051). The files were downloaded at the 5-digit zip code level, and then cross walked to HCCI's CBSAs. The estimates were aggregated to the CBSA level. All rows that did not crosswalk to an HMI CBSA were dropped. The result is a variable at the CBSA level showing the percent of population with ESI insurance by age group.