Evaluating the Quality and Cost of Palliative and Hospice Care in the U.S.

Background

The U.S. Census Bureau[1][1] (USCB) has flagged the year 2030 as "an important demographic turning point in U.S. history" as the year when the USCB expects approximately 20% of U.S. residents to be older than age 65, as compared to 13% in 2010. This rapid demographic change is expected to transform the U.S. economy and social fabric. While this demographic change has many social and economic implications, one of the most critical to understand is the stress it will place on our already fragile healthcare system, particularly as it relates to end-of-life care (or EoLC). EoLC generally involves two critical health care services: (1) palliative and (2) hospice care. Palliative care encompasses specialized medical services for people living with a chronic, terminal or serious illness. Hospice care focuses on supporting the patient's quality of life once treatment options are exhausted and the illness is deemed incurable. Typically, hospice care is initiated within the last 6 months of the patient's life and while it may incorporate palliative components, hospice care does not involve a full-time medical team.

The U.S. government administers two separate healthcare programs (i.e., Medicare[2] and Medicaid[3]) that influence the delivery and cost coverage of palliative and hospice care. The purpose of our project is to conduct an exploratory data analysis (EDA) of key healthcare variables tracked for palliative and hospice services in the U.S. in order to understand the utilization, quality, and cost of the current services provided.

Key Research Questions

- How extensive are palliative and hospice care services in the United States? How are these services distributed across states and counties?
- What are the (1) quality, (2) utilization and (3) cost characteristics of palliative and hospice care services delivered in the U.S.? Is there significant geographic variation?
- How do patient satisfaction levels differ between wealthy and poor regions?

Hypotheses

Given that palliative services are only partially covered by health care programs in the U.S., we expect that (1) income levels will directly impact the quality and cost metrics, and (2) the demand, quality and cost structure will vary at the state/county level.

Methods

To paint a more informed picture of palliative and hospice care in the U.S., we relied on publicly available data from the Center for Medicaid and Medicare Services (CMS) and the USCB, which provided metrics on demographics, quality of care, utilization and cost. Specifically, open source CMS data called "Hospice Compare" provided quality of care metrics, such as patient satisfaction. These data and their availability was federally mandated by The Hospice Quality Reporting Program (HQRP), established under section 1814(i)(5) of the Social Security Act.[4] The Act also requires the Secretary to publicly report, on a CMS website, quality measures that relate to the care provided by hospice programs across the country.[5] The stated goal of Hospice Compare is "to help consumers compare hospice providers on their performance and assist consumers in making decisions that are right for them. Providers can start a conversation with their patients and family members about how the new Hospice Compare website impacts them by:

- Explaining that the compare website provides a snapshot of the quality of care a hospice offers;
- Encouraging patients and their family members to review quality ratings;
- Helping to strengthen patients and family members' ability to make the best decisions for their care."[6]

---

[1] https://www.census.gov/newsroom/press-releases/2018/cb18-41-population-projections.html

[2] Medicare is an insurance program administered by the federal government that provides health coverage for patients over the age of 65 (or under 65 for those with a disability) regardless of the patient's income.

[3] Medicaid is a state and federal assistance program that serves low-income individuals regardless of age.

[4] For the act see: https://www.ssa.gov/OP_Home/ssact/title18/1814.htm

[5] For a summary of the background to the act and the the data see: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospice-Quality-Reporting/Hospice-Quality-Public-Reporting.html

[6] Ibid

In 2013, CMS also released a series of publicly-available synthetic claims files (DE-SynPUF) that mimicked demographic, utilization and cost patterns of Medicare beneficiaries for years 2008 to 2010. We appreciate that these synthetic data come with limitations in their ability to derive causal inference[7], but they provide a solid training ground into performing an EDA on complex observational datasets. Furthermore, they allow refining code and analytical approaches without the need to apply for access to data containing personal health information, which is both costly and administratively cumbersome. Additionally, using these synthetic data provides the opportunity to assess their usefulness in deriving insights, compared to metrics derived from real Medicare data and publicly available in the peer-reviewed literature.

The DE-SynPUF datasets contain information that has similar content and structure to healthcare claims data. As such, it contains transactional information about interactions of patients with the US Medicare Healthcare system, which effectively constitute the bills that healthcare providers send to Medicare for reimbursement. Information is divided by the type of care and venue:

- Inpatient data represent bills that come from the hospital.
- Outpatient data represent bills that come from office visits.
- Carrier data represent bills that come from other types of peripheral healthcare services[7] (i.e.: ambulances, social worker interactions, etc.).
- Finally, beneficiary summary files contain information on Medicare coverage.

Finally, we also acquired publicly available census records, that provided contextual information on the communities where hospice care facilities are provided. Metrics such as income and race at the county level provided contextual measures of socioeconomic status.

Data Preparation

The three data sources listed above were manipulated to create county-level indicators on socioeconomic status of the general population, as well as quality of care, cost and utilization of palliative and hospice services. The section below describes how each data source was manipulated to generate these county-level metrics.

Hospice-compare

This section of the analysis relied on four publicly available datasets and/or ranking measures to perform the analyses:

1. Hospice Compare Data: Provider CAHPS Hospice Survey data from cms.gov. Hospice Compare data is served in CSV flat files, last updated in May of 2019. These data have 114K rows and 17 columns, and contain a list of hospice agencies with data on their scores on the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Hospice Survey measures.There is an API, but we elected to read the files as CSVs. We parsed these data by “State” and “Score” columns, and purged the files of “Not Available” and “Not Applicable” entries.

2. “State-by-State Income Data”: Publicly available data for state-by-state and city income levels necessary for identifying potential disparities in care between states based on US Census data on median household income.[8] We parsed the national level by the top five and bottom five states, preparing dataframes for each of these analyses.

3. “Most Economically Disadvantaged City Data”: These include publicly available data to identify the “poorest” cities in the country. From this ranking we identified 10 cities where hospice care service was provided. This was to make sure state-level aggregation of care quality data was not obscuring a more focused investigation of the “costs” of disparity. We parsed these data by “City” and “Score” columns to create a separate data frame for analyses.

4. “Palliative Care ‘Report Card’ Data”: to investigate the possible correlation between palliative care rankings by state, and hospice care rankings by state. Based on data and rankings on the “poorest” cities in America, we checked this list of 50 against the hospice data and located 9 cities that also had hospice care. We imported scores provided by palliative care analyses and prepared the data to create a national map to display these scores.[9]

---

[7] https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/SynPUF_DUG.pdf

[8] “U.S. Census Bureau’s 2017 American Community Survey”

[9] https://reportcard.capc.org/

Data download

We downloaded inpatient, outpatient, carrier and beneficiary DE-SynPUF files from CMS.gov[10]. There were a total of 20 samples, each representing a 0.25% sample of the entire Medicare population between 2008 and 2010. We downloaded inpatient, outpatient, carrier and beneficiary files. In total, this represented 1400 DE-SynPUF datasets. The format available from CMS.gov was csv.

Data ingestion

We created some utility functions to ingest these 140 files in a meaningful way. For each type of DE-SynPUF dataset (i.e.: inpatient, outpatient, etc.), we first worked on appending the numerous csv's into a single dataframe. Each individual csv was appended as a dataframe into a single list. We then used a concatenation function from pandas to append the dataframe elements of the list into a master dataset. This resulted in four main dataframes (i.e.: inpatient, outpatient, carrier and beneficiary dataframes).

Palliative cohort definition

Identifying palliative care events proved to be a challenge, which we further detail in Appendix 2.

We proceeded to identify other diagnosis and procedure codes called ICD9 codes that contained palliative care services[11]. Records matching these codes were found in our datasets. Furthermore, we created a function called "palliative_cohort" that successfully identified all palliative events in inpatient, outpatient and carrier files. Using aggregating functions such as groupby, we identified that roughly 35 thousand individuals in our 10 samples has ever received a ICD9 diagnosis or ICD9 procedure code related palliative care services.

We note that that this number represents Medicare patients who had full Medicare coverage between 2008 and 2010. We excluded patients who did not have full Medicare eligibility during our time period of interest by filtering on coverage variables available in the beneficiary files.

In order to draw meaningful comparisons of Medicare populations, our final cohort contains all patients whom we defined as "eligible" for palliative care. That is, we included all individuals with full Medicare eligibility, who had at least one chronic condition, such as end-stage chronic kidney disease, cancer, alzheimers, ischemic heart disease, chronic heart failure or/and chronic obstructive pulmonary disease, based on data available in the beneficiary files. We made this selection based on data available in the beneficiary files and based on the ICD9 variables above, flagged the patients who, in fact, received palliative care services.

Using additional aggregating functions, we included the number of palliative care events and total cost of care for each of the members of our cohort.

Validation of cohort

We referred to the peer-reviewed literature to identify whether the number of individuals we linked to palliative or hospice care related in any way to the true number found in the Medicare population between 2008 and 2010[12]. Those numbers were specifically available for deceased patients in Medicare. It was reported that Medicare-certified hospice facilities serve roughly 40% of the deceased Medicare population. Unfortunately, we hypothesize that because of the lack of provider data in our synthetic datasets, we only captured that 2.5% of the deceased population received palliative/hospice services. This is a gross underestimate, and made us question the validity of synthetic data to study palliative/hospice care. Nevertheless, we used the data that were available to us and proceeded with our analyses.

U.S Census Data Source Data

In order to enrich the Medicare datasets described in the previous section, we needed to define a process to aggregate income and population data at the level of granularity of the data elements in the Medicare datasets. The Medicare datasets, however, aggregate healthcare related data elements based on standardized social security administration (SSA) state and county codes, whereas government reported U.S. Census data[13] is reported publicly at the Micropolitan Statistical Area (MSA) and state levels.

---

[10] https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF.html

[11] https://icd.codes/icd9cm/V667

[12] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3214714/

[13] https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

We could not find a direct mapping for MSA level data to SSA state and county code in order to tie the Census and Medicare datasets. Further research revealed other datasets we could use in a multi-step mapping and aggregation process to achieve our objectives.

The datasets listed below were used as source data for our multi-step data enrichment process.

- Zip Code Characteristics: Mean and median household income[14]
- FIPS / ZIP Crosswalk: Mapping of FIPS state and county codes to ZIP codes[15]
- Medicare Beneficiary Summary File: same file used in the previous section
- SSA State Code / State Abbreviation / State Name Mapping: created manually from Medicare data dictionary PDF
- SSA / FIPS State County Crosswalk: 2011 County Crosswalk[16]

Data Ingestion, Validation and Aggregation[17]

Given that multiple files from different sources were used in order to create a cleansed mapping of ZIP codes to FIPS/SSA state and county codes, we could not standardize an intake approach but rather had to do extensive data cleansing and wrangling to aggregate the final datasets that was bespoke to each source file.

At each step of the data cleansing process, we needed to run a repeated set of validation steps, so we created a utility function (*df_properties*) to facilitate the data cleansing process. The *df_properties* function takes a dataframe as an argument and returns the dataframe's shape, header, data elements, data element types, and the number of missing values by data element. This function was called before data cleansing adjustments were made to each dataset and and after to ensure that changes were reflected correctly.

The process outlined below was followed to create the final cleansed mapping. Import the data sets: Zip Code Characteristics; FIPA/ZIP Crosswalk; Medicare Beneficiary Summary; SSA State Code/Abbreviation/State Name Mapping; SSA/FIPS State County Crosswalk. A preliminary cleansing of each of these datasets was performed and certain data elements were renamed to facilitate later aggregation. The *df_properties* function was used to check each of the bespoke data preparation steps described below.

1) Use the FIPS / ZIP Crosswalk to create separate dataframes to understand and reconcile the ZIP to FIPS relationships. Specifically, create dataframes for the following:
    a) Unique FIPS state and county pairs
    b) Unique ZIP codes
    c) Unique FIPS State, FIPS County and ZIP code triplets
2) Use the Medicare Beneficiary Summary dataset to create a dataframe of unique SSA State and SSA County pairs.

The next step of our data preparation process entailed joining and aggregating cleansed datasets.

Data Analysis

Leveraging these three sources of data, our exploratory data analysis focuses on identifying any variation in palliative care services at the national level. Census data in particular provides an anchor point to visualize the fact that the U.S population is diverse both in terms of income and in terms of racial make-up. Based on census, we narrowed down our national analysis to top wealthiest and poorest geographic areas, which represent good proxies for income and potential racial diversity. Our EDA aims to assess whether these variations have any impact on the quality and the utilization of palliative care services.

Based on these national-level findings, we go on to perform subnational analyses. Because our data sources are unfortunately not aligned in time - the hospice-compare data is from 2017, while the DE-SynPUF data is from 2008/2010 - we use different years of census data to compare national income and racial trends with our hospice-compare and DE-SynPUF datasets. We nevertheless take advantage of this limitation to identify whether geographical disparities in utilization and quality of care displayed any time trends. We explore, for instance, if poor utilization metrics in 2008 are tied to the nascent awareness that palliative services is a helpful component of healthcare[18], but that by 2017, quality metrics show a robust adoption of palliative care.

---

[14] https://www.psc.isr.umich.edu/dis/census/Features/tract2zip/

[15] https://www.kaggle.com/danofer/zipcodes-county-fips-crosswalk

[16] https://www.nber.org/data/ssa-fips-state-county-crosswalk.html

[17] Supporting Python notebook: Project2_Census_Data_Enrichment_vFinal_190809.ipynb

[18] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3214714/

Data Visualization

Through our EDA, we start by using accessible visualization tools, like histograms or bar charts, to understand the distribution of our variables of interest. For example, with Hospice-compare data, we assess the distribution of quality of care responses at national and state levels. This is a simple but compelling way to see how patterns differ - or not - between different subgroups of data.

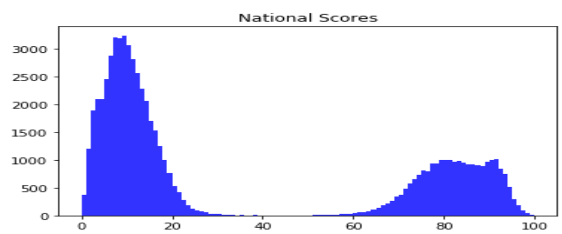
Because of the geographical nature of our data, we include maps into our visualization methods to easily visualize any national or subnational level trends. As we dive deeper into our EDA, we are compelled to look at variation by sets of two measures at a time (i.e: by state and race, or state and income, etc.). We find that using heatmaps to highlight variations by sets of two measures in our data is a very compelling visualization method.

We leveraged the visualization tools available in the matplotlib library, as well as cartopy to create maps. We also used heatmap visualization tools from the seaborn library.

Results

National-level analysis: Hospice Care Data: Exploratory Visualization and Appropriate Measures of Central Tendency

The Hospice Data analysis begins, as most exploratory data analyses begin, with a consideration of the data structure and summary statistics. This is a critical step as an appropriate measure of central tendency in this case will be dictated, in part, by the distribution of the scores in the data. For example, if the data is bimodal, mean could be a misleading measure.[19] An exploratory visualization of the data reveal that survey scores are, in fact, distributed bimodally.



This is not surprising. This distribution is consistent with the literature on survey data. Extreme response style (ERS), the proclivity to gravitate toward the two extremes of a response scale, rather than choosing an intermediate or moderate response is a commonly noted measurement bias among research in survey rating scales. [20] When responding to hospice surveys, respondents gravitated toward the two extremes with more respondents centered around a negative response. We elected, based on this distribution, to take two approaches: (1) To use mode as the measure of central tendency and (2) to treat these results as two distinct distributions, an "upper" and a "lower" response distribution. Based on this approach our findings at the national level are as follows:

Upper mean = 81.79656630144078 **Lower mean** = 9.736688498219005 **Mode** = 14

Comparing Individual State Data

Among our stated hypotheses is that quality care satisfaction ratings would be higher in states with higher median household incomes. Our reasoning was informed, in part, by the fact that only a portion of hospice care is covered by insurance. People with more money might be able to supplement their care and, therefore, be more satisfied with the care in general. On the other hand, if state care is all that is available, perhaps those that cannot supplement their care with private providers would be more satisfied. There is some evidence in the literature that acquiescence bias is exhibited most by individuals most harmed by a given social system. In other words, "yea-saying" and an unconscious bolstering of the

---

[19] S. Manikandan, "Measures of central tendency: Median and Mode," Journal of Pharmacology and Pharmacotherapeutics," July 2011, 2(3): 214-5

[20] See: D. L Paulhus, "Measurement and control of response bias", in J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes, Vol. 1. Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA, US: Academic Press; Guy Moors et al., "The Effect of Labeling and Numbering of Response Scales on the Likelihood of Response Bias," *Sociology Methodology*, 2014, 44(1) 369-399. For an exploration of the cultural and gender impact of ERS see:

Robert A. Peterson et. al., "The Cross-National Comparison of Extreme Response Style Measures," *International Journal of Market Research,* 2014, 56(1), 89-110.

status quo due to an internalization of inferiority and the belief that better isn't deserved might make states with lower income to give better survey results.[21]

To investigate this question we elected to compare the states with the top five highest median household incomes with the states with the lowest five median household incomes. Our selection of these states was based on median household income figures from the U.S. Census Bureau's 2017 American Community Survey. [22] Thus, our analyses included the following:

<u>States with lowest median household income:</u>

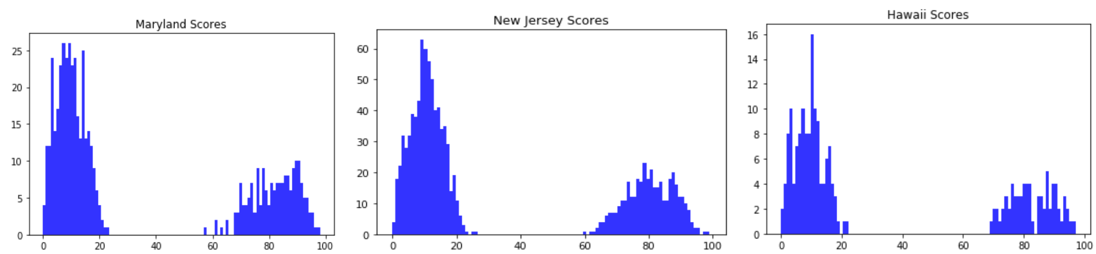
1. West Virginia: $43,469
2. Mississippi: $43,529
3. Arkansas: $45,869
4. Louisiana: $46,145
5. New Mexico: $46,744

<u>States with highest median household income:</u>

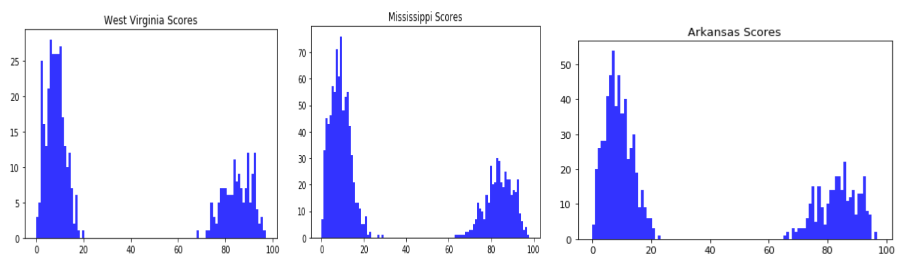
1. Maryland = $80,776
2. New Jersey = $80,800
3. Hawaii = $77,765
4. Massachusetts = $77,385
5. Connecticut= $74,16

A visualization of the scores from the 10 states exhibits a very similar bimodal distribution, and very similar high and low distribution means.

Top Three Highest Median Income States:

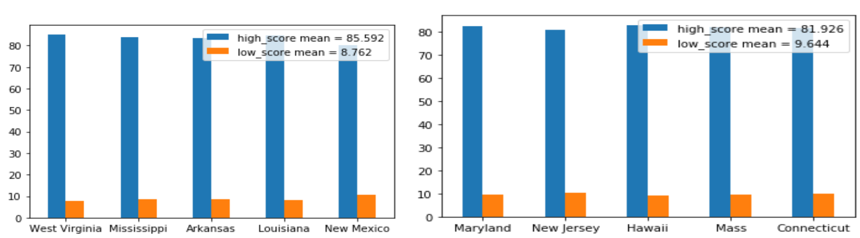


Bottom Three Median Income States:



"Positive" and "Negative" Response Comparison

Even though the visualization suggests little difference between the states, we analyzed the mean of each state. Creating a positive and negative mean score for each. As the figures show, there are only slight differences between the wealthiest and poorest states in the United States.



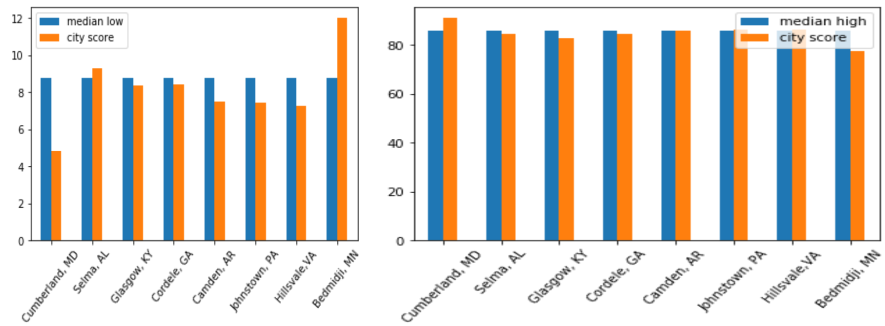
As good data scientists, we were naturally skeptical of these findings. It was surprising that the survey data returned such similar values and findings. We hypothesized that findings at the "state level" might obscure disparities. For example, Camden, New Jersey is one of the poorest cities in the country, despite the fact that New Jersey is one of the wealthiest states in the country. In order to assess this hypothesis, we decided to consider the survey scores with greater granularity. We turned next to seven of the poorest cities fifty cities in the U.S.[23] We selected these cities based on two factors: their rankings, that is, they had to be in the bottom fifty based on median household income, and they had to have hospice care offered in the cities. Our criteria resulted in the following eight cities:

[21] For an excellent comprehensive overview see: John T. Jost et al, "A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo," Political Psychology, 25 (6), 2004.

[22] https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2017/

[23] Samuel Stebbins, "50 million Americans live in poverty — here are the poorest Poorest Towns in America: Methodology Source: towns in every US state", *Business Insider*, 24/7 Wall St, Oct. 1, 2018

Cumberland,MD Selma,AL Glasgow,KY Cordele,GA Camden, AR Johnstown, PA Hillsvale, VA Bemidji, MN

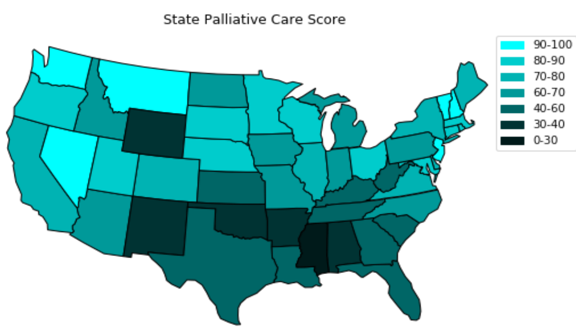


There are some subtle, though potentially interesting findings at the city level. Cumberland, MD has a much lower satisfaction mean than comparable cities, and the state-wide mean. 4 of the 8 states considered have a lower negative score mean. Bedmidji, Minnesota, however, has a much higher low score satisfaction rating than the national average. It also has a lower upper score than the national average. That said, none of these data suggest a significant difference in experience from state to state, from city to city.

Context for our findings: Palliative Care Report Card

Given these findings, we looked to evaluate the lack of significant difference with a comparable, though distinct data score measure. We turned to national palliative care scores to evaluate our findings. If Hospice Quality Care scores are accurately capturing hospice patient experience, we would expect to see a similar lack of distinct experiences nationwide.

To assess palliative care scores, we turned to the National Palliative Care Research Center and the Center to Advance Palliative Care who published a state-by-state “report card” and based on multiple factors assigned scores to each state. Their findings significantly challenge the Hospice Care Quality Report due to the diversity of experiences from state to state, as figure below illustrates:



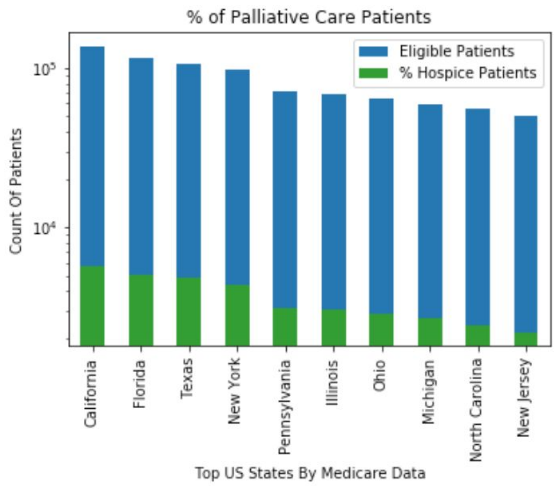
National-level analysis: Synthetic Medicare Data

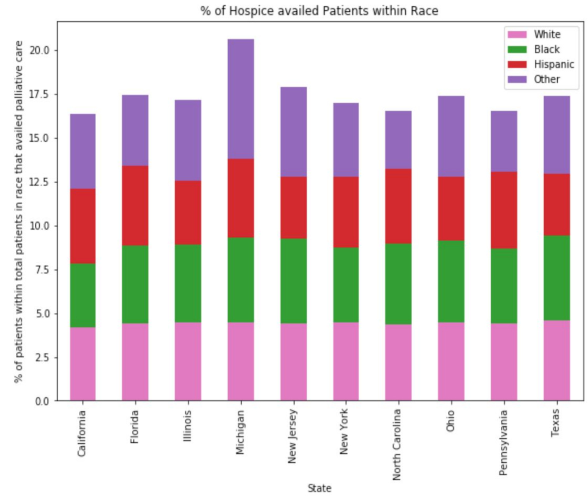
Assessment of the palliative care population

As we mention in our methods section, we identified that there were 35,280 patients who received at least one palliative care service between 2008 and 2010. We were able to derive that only 2.5% of deceased patients in our cohort received some palliative care, compared to patients eligible for some palliative care but who received none. Overall, about 4.3% of all patients eligible for palliative care received it. We appreciate that this is likely an underestimate due to having proper variables available in our dataset.

Top 10 states with most palliative care eligible patients - rate of palliative care administration

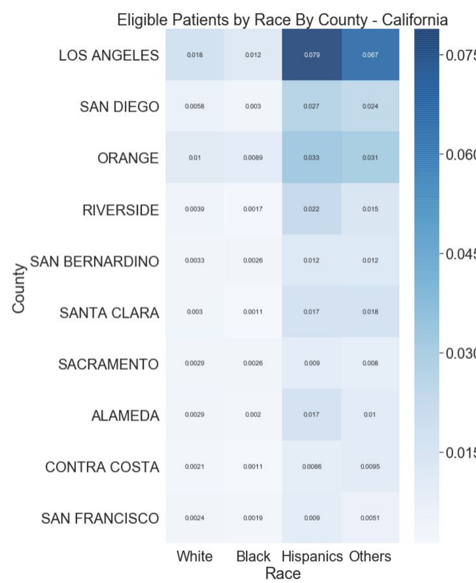
In our proposal, we hypothesized that geographical variation would exist in the way palliative care is delivered. Given our small number of palliative care events, we narrowed down our analysis to the top 10 states with most palliative care-eligible patients. We used the logarithmic count of patients, so as to visualize differences between states.





We observe from this bar chart that reception of actual palliative care services is actually proportional to the total number of patients eligible for palliative care. No unexpected trends surge from these findings. This chart allows assessing whether the proportion of patients that received palliative care services in the top 10 states with most eligible patients, varies by race. We notice for instance that in Michigan, the proportion of "Other" is greater than in other states. This could be due to underlying population differences, so we set out to understand the baseline racial differences in those top 10 states and control for those differences.

Racial distribution of patients eligible for palliative care services in the top 10 states and counties.



We then further explored whether variations existed in the number of eligible patients in the top 10 states and by race using a normalized measure of counts. The darker areas indicate that more eligible patients are available for a specific state and race subgroups. We notice that while the rate of Medicare patients enrolled by race, as seen in the histograms above, is fairly well distributed by race, the underlying population of Medicare eligible patients does show some variations.

National-Level Analysis: Hospice Beneficiaries and Relative Per Capita Income[24]

Exploratory Data Analysis and VisualizationsDensity and Histogram for Medicare Hospice Patients and All Medicare Patients as a Percentage of State Population

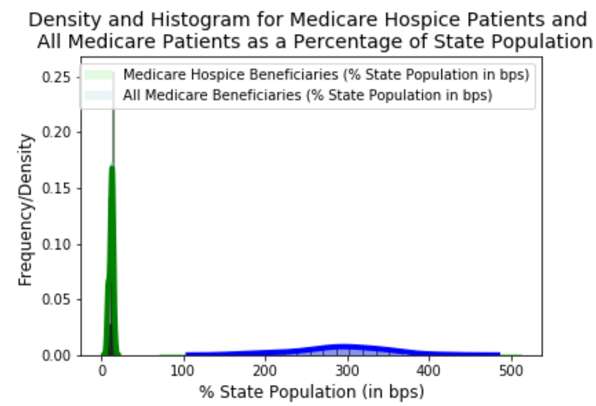
The respective density distributions and histograms show two very different distributions for medicare hospice beneficiaries and overall medicare beneficiaries. The rate of medicare hospice beneficiaries as a percentage of the population is very low with a very small standard deviation about the mean.

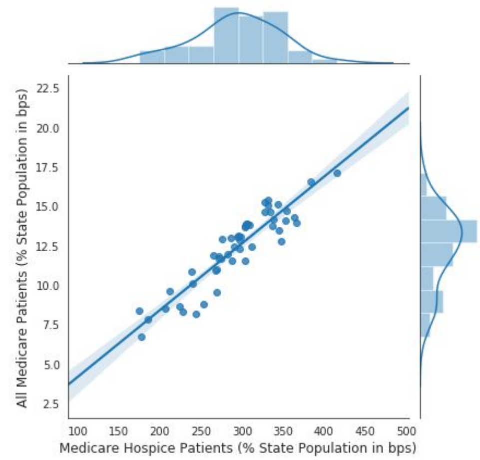
As expected, overall medicare beneficiaries as a percentage of the population have a much higher mean. But surprisingly, the distribution is very flat.

---

[24] Supporting Python notebook: Project2_Census_Analysis_vFinal_190809

Also, there is a strong linear relationship (with positive slope) between the rates of medicare hospice care and overall medicare services, given that the former is a subset of the latter.





Summary Statistics for All Medicare Patients as a Percentage of State Population (bps_ben_pop)

count	51
mean	292.6014
std	52.5207
min	174.6422
25%	268.1296
50%	297.4042
75%	331.7323
max	415.7726

Summary Statistics for Medicare Hospice Patients a Percentage of State Population (bps_ben_hosp_pop)

count	51
mean	12.3088
std	2.4055
min	6.7290
25%	10.9244
50%	12.8444
75%	13.8958
max	17.1182

Comparing Hospice Rates of Top and Bottom Quartile States by Log of Per Capita Income

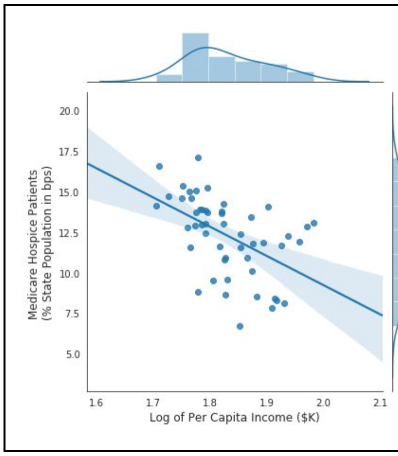
Now that we have a sense of the underlying census and medicare beneficiary rate distributions, we wanted to understand the relationship between medicare hospice beneficiary rates as a percentage of the population relative to per capita income.

Joint Distribution Between Medicare Hospice Patients as a Percentage of State Population and Log of Per Capita Income

We initially plotted the joint distribution between Medicare hospice patient rate (as a percentage of population) and per capita income. The resulting plot showed both clustering and heteroskedasticity, so we chose to log-transform the per capita income explanatory variable.

The resulting joint distribution of the log-transformed explanatory variable and the Medicare hospice patient rate (as a percentage of population) results less clustering and a strong linear relationship between the log-transformed explanatory variable and population as reflected in the high R-squared. However, there is still significant heteroskedasticity making it a poor regression.

On an aggregate national level, it appears that higher rates of Medicare hospice services per capita occur at lower log-transformed per capita income levels. This is consistent with our initial intuition since we expect that wealthier individuals to be more likely to have private insurance as well as savings to address end of life needs.



OLS Regression Results

Dep. Variable:	bps_ben_hosp_pop	R-squared (uncentered):	0.956
Model:	OLS	Adj. R-squared (uncentered):	0.955
Method:	Least Squares	F-statistic:	1078
Date:	Fri, 09 Aug 2019	Prob (F-statistic):	1.69E-35
Time:	19:24:05	Log-Likelihood:	-121.87
No. Observations:	51	AIC:	245.7
Df Residuals:	50	BIC:	247.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.0
log_incpp_state_k	6.6945	0.204	32.831	0	6.2

Omnibus:	1.594	Durbin-Watson:	1.142
Prob(Omnibus):	0.451	Jarque-Bera (JB):	1.503
Skew:	-0.312	Prob(JB):	0.472
Kurtosis:	2.436	Cond. No.	1

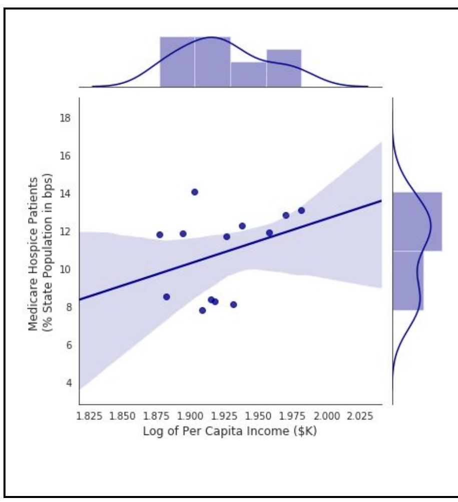
The clustering and greater heteroskedasticity on either end of the plot suggests there may be other relationships that are being masked in this aggregate view. Consequently, we decided to split the data set into quartiles by per capita income and compare the top quartile to the bottom quartile.

Top Quartile States by Log of Per Capita Income

Joint Distribution Between Medicare Hospice Patients as a Percentage of State Population and Log of Per Capita Income, and Associated Marginal Distributions

The regression of the Medicare hospice patient rate (as a percentage of population) on the log-transformed explanatory variable (log of per capita income in thousands) for the top quartile data set results less clustering and a strong positive linear relationship between income and population as reflected in the high R-squared and statistical significance of the t-test. However, there is a lot of heteroskedasticity, and given the small number of observations, this would not be considered a good regression.

The regression of the top quartile data set suggests that states with higher per capita income have a higher per capita rate of usage of Medicare hospice services—a different conclusion than the one drawn from the aggregate regression, and counter our initial intuition that wealthier individuals are more likely to have other forms of insurance and savings to address end of life needs.



OLS Regression Results

Dep. Variable:	bps_ben_hosp_pop	R-squared (uncentered):	0.964
Model:	OLS	Adj. R-squared (uncentered):	0.961
Method:	Least Squares	F-statistic:	321.8
Date:	Fri, 09 Aug 2019	Prob (F-statistic):	4.94E-10
Time:	19:24:06	Log-Likelihood:	-28.042
No. Observations:	13	AIC:	58.08
Df Residuals:	12	BIC:	58.65
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
log_incpp_state_k	5.6333	0.314	17.94	0	4.949	6.317

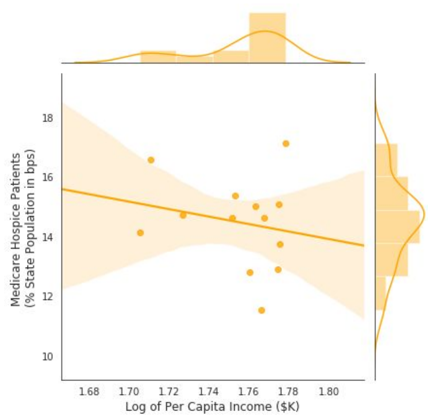
Omnibus:	3.817	Durbin-Watson:	1.903
Prob(Omnibus):	0.148	Jarque-Bera (JB):	1.255
Skew:	-0.211	Prob(JB):	0.534
Kurtosis:	1.538	Cond. No.	1

Bottom Quartile States by Log of Per Capita Income

Joint Distribution Between Medicare Hospice Patients as a Percentage of State Population and Log of Per Capita Income, and Associated Marginal Distributions

The regression of the Medicare hospice patient rate (as a percentage of population) on the log-transformed explanatory variable (log of per capita income in thousands) for the bottom quartile data set results in a little more clustering in the range of 1.74 and 1.78 of the log-transformed explanatory variable. There is a strong negative linear relationship between per capita income and population as reflected in the high R-squared and statistical significance of the t-test. However, as with the regression for the top quartile data set, there is a lot of heteroskedasticity and given the small number of observations, this would not be considered a good regression.

The regression of the bottom quartile data set suggests that states with lower per capita income have a decreasing rate of per capita rate of usage of Medicare hospice services, more in line with the conclusions implied by the aggregate regression. Given the weaknesses of the regresion, however, the latter is not a reliable conclusion.



OLS Regression Results

Dep. Variable:	bps_ben_hosp_pop	R-squared (uncentered):	0.989
Model:	OLS	Adj. R-squared (uncentered):	0.988
Method:	Least Squares	F-statistic:	1102
Date:	Fri, 09 Aug 2019	Prob (F-statistic):	3.53E-13
Time:	19:24:07	Log-Likelihood:	-23.8
No. Observations:	13	AIC:	49.6
Df Residuals:	12	BIC:	50.17
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
log_incpp_state_k	8.2464	0.248	33.203	0	7.705	8.788

Omnibus:	0.225	Durbin-Watson:	2.192
Prob(Omnibus):	0.893	Jarque-Bera (JB):	0.158
Skew:	-0.194	Prob(JB):	0.924
Kurtosis:	2.624	Cond. No.	1

Discussion

Despite the federal mandate, Hospice Quality Care scores are inadequate at measuring the hospice patient experience and, therefore, fall far short of their stated goal to inform patients and their families of the quality of options available to them. While ERS bias may be present in all survey results, to more actively satisfy the federal mandate, Hospice Quality Care metrics must rely on more than survey data. We recommend multiple metrics and methods using the Palliative Care Report Card as a guide for Hospice Quality Care metrics.

Consistent with findings identified in the Hospice-Compare data, we found no striking differences in the way palliative care services are administered for geographical and racial subgroups. While we did find that more patients were eligible for palliative services in certain geographical and racial subgroups, the way palliative care is administered across these subgroups seems quite homogenous. It is unclear whether this homogeneity stems from the fact that Medicare is a federal program that aims to provide care in a similar manner to its beneficiaries, irrespective of income and race. We also suspect that this homogeneity is attributed to working with synthetic data, which - by design - smoothes granular differences that could make data identifiable. Furthermore, because we struggled with identifying palliative care events, our numbers may not reflect the true rate of palliative care patients.

When joining our datasets with Census data, we identified that lower-income patients were more likely to receive palliative care than wealthier patients. However, the relationship between income and palliative care services was somewhat inconsistent across various analyses, and requires further investigation.

Conclusion

While our original hypotheses relied on identifying variation by various subgroups, we were surprised to see little to no variation of palliative care among these subgroups. Analyses not shown in this report looked at other variables, such as cost and utilization, but demonstrated similar trends. While we appreciate the federal mandate from Medicare to provide fair and equal healthcare delivery to all, existing measures, such as the palliative care scorecard, seemed to indicate that regional patterns do exist. Further investigation is required using additional, non-synthetic Medicare data to see if our lack of trends is an artifact from data available to us, or if these trends are sustained across various data sources. If anything, these analyses identified that, while open data exist on Medicare, they do not prove to be reliable sources for deriving meaningful insights.

Appendix 1

Creating the FIPS to ZIP relationship:

This three-step process revealed that there is a many-to-one relationship between ZIP codes and FIPS State/County pairs; however, there isn't a unique many-to-one relationship between FIPS State/County pairs and ZIP codes (i.e., the same ZIP code can be mapped to different FIPS State/County pairs).

This creates a problem because we do not have sufficient information to be able to split income and population of a single ZIP code across FIPS State/County pairs.

Consequently, we chose to retain only unique FIPS State, FIPS County and ZIP code triplets in order to ensure that the mapping was bi-directional. As such, aggregated income and population metrics for FIPS State and FIPS County are just proxies.

The mapping for ZIP codes to FIPS State level codes, however, is bidirectional, so state level aggregations of income and population will be accurate.

Cleansing of master census dataset

1) Creation of Master Census Data Frame
    a) Merge the cleansed income by ZIP code data frame (dat_IncByZIP) to the cleansed bidirectional ZIP/FIPS mapping (dat_ZIPtoFIPS) using an inner join on the ZIP code data element.
    b) Check for duplicate (i) FIPS State/County pairs in the merged file, as well as potential FIPS/SSA duplicates (ii) SSA / FIPS State County Crosswalk
        i) This step revealed that all of teh SSA/FIPS mappings were bidirectional one-to-one mappings with the exception of one FIPS county code that was mapped two distinct SSA county codes.
        ii) Given it was only one FIPS county code in question, we created a one-to-one mapping for this instance.
    c) The resulting SSA/FIPS state and county codes relationship is one-to-one in the cleansed master census data frame. It was then merged with the cleansed SSA / FIPS State County Crosswalk data frame (dat_xwalk_SSAtoFIPS).
    d) We then used the dataframe from the previous step to aggregate population and income based on SSA / FIPS state and county codes. As noted above, these values are proxies for income and population at the county code level, but are accurate reflections of income and population at the state level.
    e) Augment data frame created in previous step to include full state names by merging it with the cleansed SSA State Code / State Abbreviation / State Name Mapping (dat_ssa_state).
2) Output cleansed census data set to a file (master_census_by_SSA_FIPS.csv) for use in statistical analysis.

Appendix 2: Identification of palliative care events.

We leveraged the DE-SynPUF documentation and data dictionary to identify whether "palliative" or "hospice" was mentioned in any of the existing variables. We identified that a variable called "Provider Number" contained combinations of digits (3rd and 4th digits of the provider number ID) that allowed deriving whether the provider was a hospice facility. Using slicing of strings, we were able to extract these 2 digits, but realized that in the DE-SynPUF datasets, no provider number was available in the range that pertained to hospice providers, despite being defined in the data dictionary. Identifying palliative care events proved to be a challenge, which we further detail in Appendix 1.

We then looked for alternatives. Based on the data dictionary, we also identified that inpatient, outpatient and carrier files contained variables related to diagnoses and procedures recorded during a patient's encounter. We found that specific procedure codes called HCPCS codes and available in our data identified palliative care services[25]. We scanned the files to identify these events. Unfortunately, we identified that procedure code data was scarcely populated. We used the specific panda function "isna(var).all()" to confirm that entire HCPCS variables were completely empty. We also used conditional statements to see if any of the scarcely available HCPCS values related to palliative care. Unfortunately, none of the procedure-related variables contained codes related to palliative care procedures.

---

[25] https://www.chcf.org/wp-content/uploads/2019/05/DocumentationCodingHandbookPalliativeCare.pdf