HCUP Methods Series

Methods for Calculating Patient Travel Distance to Hospital in HCUP Data

Report #2021-02

Agency for Healthcare Research and Quality

U.S. Department of Health and Human Services
Agency for Healthcare Research and Quality

Contact Information:
Healthcare Cost and Utilization Project (HCUP)
Agency for Healthcare Research and Quality
5600 Fishers Lane
Room 07W17B
Mail Stop 7W25B
Rockville, MD 20857
www.hcup-us.ahrq.gov

For Technical Assistance with HCUP Products:

Email: hcup@ahrq.gov

or

Phone: 1-866-290-HCUP

TABLE OF CONTENTS

INDEX OF TABLES AND FIGURES

LIST OF REPORT SUPPLEMENTS

Supplement 1. Distribution of difference (in miles) between geocode methods, by patient region, division, and urban/rural location, 2018

Supplement 2. Distribution of travel distance (in miles) by scenario, patient region, division, and urban/rural location, 2018

Supplement 3. Distribution of difference in travel distance (in miles) between scenarios, by patient region, division, and urban/rural location, 2018

EXECUTIVE SUMMARY

This report examines patient travel distance to the hospital using inpatient data from the Healthcare Cost and Utilization Project (HCUP). Different methods are compared for identifying the latitude and longitude coordinates (geocodes) of the patient and hospital locations. Patient location is determined from the patient ZIP Code in HCUP data using two different methods of geocoding: geographic centroid versus population-weighted centroid of the ZIP Code. Hospital location is determined from the hospital address provided by the American Hospital Association (AHA) using two different geocoding methods: AHA-provided geocodes versus Google Maps geocodes. Finally, the distance between the patient and hospital location was determined using two different methods for calculating travel distance: straight line versus driving distance.

There was a relatively small difference between the two patient geocoding methods (median of 0.6 miles) and a very small difference between the two hospital geocoding methods (median of 0.02 miles).

Overall, the median straight-line patient-hospital travel distance was 6.6 miles, with 75 percent of distances less than 15 miles and 90 percent of distances less than 30 miles. The median driving distance was 8.7 miles. Most of the difference in travel distance was due to the distance metric (driving vs. straight-line) with little difference due to the patient or hospital geocoding method. Driving distances were approximately 30 percent longer on average than straight-line distances, and this relationship was relatively constant with only small variations for geographic area.

Geographically, the longest distances were in the South region, East South Central division, and noncore (rural) areas. The shortest distances were in the Northeast region, New England division, and large central metropolitan areas. Distances were generally longer for patients residing in ZIP Codes with greater area (covering more square miles).

These results may be useful to researchers studying how patient travel distance to the hospital relates to topics such as healthcare access, decisions about where to obtain hospital care, and outcomes of medical and surgical treatment.

INTRODUCTION

Background

Patient travel distance to the hospital is an important factor related to access to care and where patients obtain inpatient care. The distance between a patient's residence and the nearest hospital has increased in recent years.[1,2] From 2005 to 2015, the number of people who lived more than 60 minutes from any hospital increased by more than 80 percent.[3] Research has

---

[1] Centers for Disease Control and Prevention. "Distance to Nearest Hospital" Files: NAMCS and NHAMCS (1999 to 2009). www.cdc.gov/nchs/data/ahcd/distance_to_nearest_hospital_file.pdf. Accessed November 6, 2020.
[2] Diaz A, Schoenbrunner A, Pawlik TM. Trends in the geospatial distribution of inpatient adult surgical services across the United States. Annals of Surgery. 2021;273(1):121–7.
[3] Ibid.

focused on how patients' distance to the nearest hospital has been affected by hospital closures and mergers.[4]

In addition to studies focused on healthcare access, travel distance also is an important metric for research that explores how patient choice, type of illness and surgical needs, and hospital quality impact where patients receive inpatient care. For example, patients will travel farther for an elective admission than they will for an emergent admission.[5] Travel distance also may be related to treatment outcomes. In one study, patients who traveled farther to high-volume centers for treatment of pancreatic cancer had better postoperative outcomes than patients who were treated locally at low-volume centers.[6] Other studies exploring factors related to travel distance have found longer travel distances for patients who are younger, have higher levels of education, are of White race, are from higher income areas, reside in rural areas, are in better health, and are privately insured.[7,8,9,10] Understanding the distance patients travel to obtain hospital care also has important policy implications, because disparities in travel distance may affect access to care, costs associated with travel, and inequities in care.[11,12]

Methods for defining travel distance require determining the patient and hospital locations and specifying a metric for measuring the distance between them. For patient location, patient ZIP Code centroids are frequently used because exact patient residence address information is frequently unavailable in research datasets due to confidentiality.[13] For hospital location, hospital addresses are known and precise locations can typically be determined. To measure the distance between the patient location and the point of care (hospital or emergency department), the shortest or "straight-line" distance (i.e., the geodetic or great circle distance) is commonly used because it can be readily calculated (e.g., through statistical software programs such as SAS®).[14,15] An alternative distance metric that has been used is the driving distance or

[4] Wishner J, Solleveld P, Rudowitz R, Paradise J, Antonisse L. A look at rural hospital closures and implications for access to care: three case studies. Kaiser Family Foundation Issue Brief. July 2016. www.kff.org/report-section/a-look-at-rural-hospital-closures-and-implications-for-access-to-care-three-case-studies-issue-brief/. Accessed August 10, 2021.

[5] Johnson JE. An analysis of distance traveled for healthcare services utilizing a GIS. Unpublished manuscript. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.392.2139. Accessed November 22, 2021.

[6] Lidsky ME, Sun Z, Nussbaum DP, Adam MA, Speicher PJ, Blazer DG. Going the extra mile: improved survival for pancreatic cancer patients traveling to high-volume centers. Annals of Surgery. 2017;266(2):333–8.

[7] Tai WC, Porell FW, Adams EA. Hospital choice of rural Medicare beneficiaries: patient, hospital attributes, and the patient-physician relationship. Health Services Research. 2004;39(6):1903–22.

[8] Buhn S, Holstiege J, Pieper D. Are patients willing to accept longer travel times to decrease their risk associated with surgical procedures? A systematic review. BMC Public Health. 2020;20:253.

[9] Jia P, Wang F, Xierali IM. Differential effects of distance decay on hospital inpatient visits among subpopulations in Florida, USA. Environmental Monitoring and Assessment. 2019:19(Suppl 2):381.

[10] Gutacker N, Siciliani L, Moscelli G, Gravelle H. Choice of hospital: which type of quality matters? Journal of Health Economics. 2016;50:230–46.

[11] Diaz A, Schoenbrunner A, Pawlik TM. Trends in the geospatial distribution of inpatient adult surgical services across the United States. Annals of Surgery. 2021;273(1):121–7.

[12] Tai WC, Porell FW, Adams EA. Hospital choice of rural Medicare beneficiaries: patient, hospital attributes, and the patient-physician relationship. Health Services Research. 2004;39(6):1903–22.

[13] Fitzgerald JD, SooHoo NF, Losina E, Katz JN. The potential impact on patient-hospital travel distance and access to care under a policy of preferential referral to high-volume knee replacement hospitals. Arthritis Care & Research (Hoboken). 2012;64(6):890–7.

[14] Brown AM, Decker SL, Selck FW. Emergency department visits and proximity to patients' residence, 2009–2010. NCHS Data Brief #192. March 2015. Hyattsville, MD: National Center for Health Statistics.

[15] Centers for Disease Control and Prevention. "Distance to Nearest Hospital" Files. www.cdc.gov/nchs/data/ahcd/distance_to_nearest_hospital_file.pdf. Accessed July 22, 2021.

driving times that can be obtained from various mapping software such as Google Maps,[16,17,18] MapQuest,[19] OpenStreetMaps,[20] and ArcGIS Network Analyst.[21]

HCUP data are the most comprehensive source of hospital inpatient stays in the United States and may be a valuable resource for research involving the distance that patients travel to the hospital. This report focuses on describing methods for calculating patient-hospital travel distance using HCUP data.

Objective

The objective of this report is to compare patient-hospital travel distances based on HCUP inpatient data using two different methods of geocoding patient and hospital locations and two different methods of calculating the distance between the patient and hospital. Geocoding entails obtaining the latitude and longitude coordinates of the patient or hospital location. Table 1 provides a summary of the three scenarios that are compared in this report.

Table 1. Scenarios to compare patient location, hospital location, and travel distance measures

Scenario	Patient geocoding method	Hospital geocoding method	Distance measure
Baseline: standard, most available approach	A. Geographic centroid of patient's ZIP Code (calculated in SAS)	A. AHA-provided geocode of AHA-defined hospital	A. Straight-line distance (calculated in SAS)
Scenario 1: more precise location methods	B. Population-weighted centroid of patient's ZIP Code (from Esri)	B. Google Maps geocode of AHA-defined hospital address	A. Straight-line distance (calculated in SAS)
Scenario 2: more precise location methods and distance metric	B. Population-weighted centroid of patient's ZIP Code (from Esri)	B. Google Maps geocode of AHA-defined hospital address	B. Driving distance (from Google Maps)

Abbreviation: AHA, American Hospital Association

The Baseline scenario represents a relatively simple and accessible approach to obtain travel distance. Patient location is defined as the geographic centroid of the patient's ZIP Code, the coordinates of which can be readily obtained using SAS software. Hospital location is defined as the geocode provided by the American Hospital Association (AHA). Finally, the distance between the patient and hospital is determined by the straight-line distance calculated using SAS software.

---

[16] Lam O, Broderick B, Toor S. How Far Americans Live From the Closest Hospital Differs by Community Type. Pew Research Center. December 12, 2018. www.pewresearch.org/fact-tank/2018/12/12/how-far-americans-live-from-the-closest-hospital-differs-by-community-type/. Accessed July 22, 2021.

[17] Tsuang WM, Arrigain S, Lopez R, Snair M, Budev M, Schold JD. Patient travel distance and post lung transplant survival in the United States: a cohort study. Transplantation. 2020;104(11):2365–72.

[18] Rocque GB, Williams CP, Miller HD, Azuero A, Wheller S, Pisu M, et al. Impact of travel time on health care costs and resource use by phase of care for older patients with cancer. Journal of Clinical Oncology. 2019;37(22):1935–45.

[19] Bliss RL, Katz JN, Wright EA, Losina E. Estimating proximity to care: are straight line and zipcode centroid distances acceptable measures? Medical Care. 2012;50(1):99–106.

[20] Diaz A, Burns S, D'Souza D, Kneuertz P, Merrit R, Perry K, et al. Accessing surgical care for esophageal cancer: patient travel patterns to reach higher volume center. Diseases of the Esophagus. 2020;33:1–10.

[21] Jia P, Wang F, Xierali IM. Differential effects of distance decay on hospital inpatient visits among subpopulations in Florida, USA. Environmental Monitoring and Assessment. 2019:19(Suppl 2):381.

Scenario 1 employs the same straight-line distance method (from SAS) as the Baseline scenario but uses alternative patient and hospital geocodes that may be somewhat more precise but also may be more difficult to obtain. Patient location is defined as the population-weighted centroid of the patient's ZIP Code obtained from ZIP Code files available from Esri. Hospital location is the geocode of the AHA-defined hospital address produced by Google Maps.

Scenario 2 employs the same patient and hospital geocoding methods (population-weighted ZIP Code centroids and Google Maps, respectively) as Scenario 1 but uses an alternative distance metric. Specifically, the distance between the patient and hospital locations is determined by the shortest driving distance obtained from Google Maps.

STUDY DATA

HCUP Data

State-level data on U.S. hospitalizations (inpatient care) are available through the HCUP State Inpatient Databases (SID), which include the patient residence ZIP Code (full patient address is not available in HCUP data) and an identifier for the hospital. The actual hospital address can be obtained by linking hospital identifiers in the SID to the AHA data (see the "Non-HCUP Data" section below). HCUP SID are available in 2018 for 47 States and the District of Columbia; 2018 SID are not available for Alabama, Idaho, and New Hampshire. The data used for this report are based on all available 2018 SID. Data exclusions are described below and summarized in Table 2 under the "Analysis File" section.

1. The data for this report consist of the 48 HCUP 2018 SID. Data for patients residing in the three States without 2018 HCUP SID but treated in hospitals in States with HCUP data are included. Data included are from community hospitals, which are defined as short-term, non-Federal, general, and other hospitals. Excluded are data from long-term care facilities such as rehabilitation, psychiatric, and alcoholism and chemical dependency hospitals. Hospitals in HCUP data that could not be linked to hospitals in the 2018 AHA Annual Survey were not included.[22]

2. We limited discharges to those where both the hospital location and patient residence location were within the continental United States. Specifically, we excluded discharges for hospitals in Alaska and Hawaii and for patients residing in Alaska, Hawaii, or the U.S. territories (American Samoa, Guam, Northern Mariana Islands, Puerto Rico, and the U.S. Virgin Islands). We implemented these exclusions in order to eliminate atypical travel patterns (e.g., air, boat) that many patients from these areas would need to undertake to obtain hospital care. Additionally, patients with ZIP Codes identifying them as homeless were assigned to the ZIP Code of the hospital destination.

3. We excluded discharges with patient origin ZIP Codes that were not found in the Esri 2018 USA ZIP Code Areas reference data file (see the "Non-HCUP Data" section below). The Esri file was used to identify valid ZIP Codes and to map ZIP Codes such

---

[22] Of 5,678 hospitals in HCUP data identified by the HCUP Partner-supplied identifiers, 147 could not be mapped to an AHA identifier (2.6%).

as post office (PO) boxes that do not represent a geographic area to the surrounding ZIP Code areas.

4. We excluded discharges representing patient transfers in from another hospital and included discharges for patients transferred out to another hospital. Discharges for patients transferred out represent the initial hospital destination for the patient—the first hospital to which the patient traveled or was transported for care.

5. We excluded discharges with extreme patient-hospital travel distances calculated under the Baseline scenario. Extreme distances were defined as those that were more than either the 99th percentile distance for that State (based on the patient origin) or 250 miles, whichever was smaller.[23]

6. We excluded discharges that were not part of the sampled pairs used to obtain driving distances under Scenario 2. For analysis, the sampled discharge data were weighted to represent the entire universe of discharges.

7. We excluded discharges with no patient-hospital travel distance available under Scenario 2.

Non-HCUP Data

The following non-HCUP data were used to obtain hospital and patient geocodes (latitude and longitude coordinates) and to associate point ZIP Codes such as PO boxes with area-level ZIP Codes.

2018 American Hospital Association (AHA) Annual Survey of Hospitals

The AHA database contains self-reported hospital information on a wide range of hospital characteristics, including organizational structure, service lines, and staffing. AHA data are used as part of the standard development of the HCUP databases (e.g., SID) to identify the types of hospitals (e.g., community hospitals) and to provide supplemental data elements, such as bed size, teaching status, and control/ownership of the hospital. We used the AHA data for this report to obtain the geocode of the hospital for the Baseline scenario and the hospital address to use for obtaining the Google Maps geocode.

We identified a few cases where the AHA hospital geocodes appeared to be incorrect (i.e., the ZIP Code of the hospital address provided in the AHA data and the AHA-provided geocode did not coincide). In some cases, we were able to determine that the ZIP Code was incorrect, and in other cases, the geocode was incorrect. In two cases, these appeared to be extreme geocode errors whereby the geocode put the hospital in a different State from the ZIP Code. These errors were infrequent. We used the AHA hospital address and geocode data as provided in the AHA database. Our Scenario 1 used a more precise measure of hospital location based on Google Maps that would eliminate these AHA data errors.

SAS SASHELP.ZIPCODE

This SAS dataset contains the geographic centroid and county assignment for each ZIP Code. SAS obtains the ZIP Code geocodes included in this dataset from ZipCodeDownload.com. We

---

[23] Note: Because the Baseline scenario was used to implement the extreme distance exclusion, some travel distances longer than 250 miles were observed in the results for Scenarios 1 and 2.

used this SAS dataset to obtain the geocodes of the patients' ZIP Code geographic centroids to define patient location under the Baseline scenario.

Visual examination comparing the SAS version of the ZIP Code geographic centroids with geographic centroids computed directly from the Esri 2018 USA Zip Code Areas file (described below) using GIS software (QGIS version 3.14.16, open source) revealed instances where the SAS-provided geographic centroids appeared to be incorrect (e.g., in a far corner of a ZIP Code or in an entirely different ZIP Code). We used the SAS dataset geographic ZIP Codes as provided for the Baseline scenario. Our Scenarios 1 and 2 used an alternative measure of patient location defined as the population-weighted ZIP Code centroid provided by Esri.

Esri 2018 USA ZIP Code Points[24]

This file contains "five-digit U.S. ZIP Code areas as points, plus all ZIP Codes that have no associated area such as post office box ZIP Codes and single site ZIP Codes (government, building, or organization)."[25] Points are the latitude and longitude coordinates of the population-weighted centroid of the ZIP Code. Fields in the data file include ZIP Code, post office name, and type for the ZIP Code locations in the United States. There are two types of ZIP Codes: those that cover a defined geographic area and those that do not cover any geographic area (e.g., post offices or large-volume customers). We accessed this file to obtain the population-weighted centroids of the patient's ZIP Code for Scenarios 1 and 2 and to identify ZIP Codes that do not cover any geographic area in order to map them to their enclosing ZIP Code area.

Esri 2018 USA Zip Code Areas[26]

This file contains "five-digit ZIP Code areas used by the U.S. Postal Service to deliver mail more effectively."[27] Areas are defined by a sequence of latitude/longitude pairs that create a polygon defining each ZIP Code. Fields in the data file include ZIP Code, post office name, population, square mile area, and latitude/longitude coordinates for the ZIP Code area. This file was used to associate ZIP Codes that do not cover any geographic area (i.e., ZIP Codes with a post office box or other single delivery site) with the ZIP Code area enclosing it using GIS software (QGIS version 3.14.16, open source)[28] and to obtain the area (square miles) of the ZIP Code.

Analysis File

Table 2 provides a summary of the HCUP 2018 inpatient data included in the analysis file used for this report. Data exclusions, as described in the "HCUP Data" section above, are presented in the table along with the number of impacted records.

---

[24] ArcGIS. USA ZIP Code Points. Updated May 13, 2021. www.arcgis.com/home/item.html?id=1eeaf4bb41314febb990e2e96f7178df. Accessed May 26, 2021.
[25] Ibid.
[26] ArcGIS. USA ZIP Code Areas. Updated May 13, 2021. www.arcgis.com/home/item.html?id=8d2012a2016e484dafaac0451f9aea24. Accessed May 26, 2021.
[27] Ibid.
[28] QGIS. Download QGIS for Your Platform. www.qgis.org/en/site/forusers/download.html. Accessed May 26, 2021.

Table 2. Construction of analysis file: inclusions and exclusions from HCUP SID, 2018

Distance analysis file build steps	Included records			Excluded records (from prior step)		
	Discharges	Patient ZIP Codes	Unique hospital/ patient ZIP Code pairs	Discharges	Patient ZIP Codes	Unique hospital/ patient ZIP Code pairs
1. Initial file: All 2018 HCUP SID data for community, nonrehabilitation, non-LTAC hospitals that map to AHA data.	34,254,432	41,635	1,141,891	—	—	—
2. Exclude data for hospitals and patients not in the continental United States (i.e., Alaska, Hawaii, U.S. territories).	34,072,002	32,848	1,130,291	182,430	8,787	11,600
3. Exclude data for patient ZIP Codes with no Esri reference data and use reference data to map unique ZIP Codes (e.g., PO boxes) to surrounding ZIP Code area.	34,028,846	30,511	1,074,774	43,156	2,337	55,517
4. Exclude data for patients transferred in from another hospital.	30,687,994	30,494	985,802	3,340,852	17	88,972
5. Exclude extreme travel distances under the Baseline scenario. ANALYTIC UNIVERSE	30,278,849	30,419	690,451	409,145	75	295,351
6. Exclude pairs not sampled for Google Maps processing (Scenario 2).	11,296,867	29,499	261,704	18,981,982	29,840	428,747
7. Exclude sampled pairs with no distances available under Scenario 2. ANALYTIC SAMPLE	11,296,457	29,476	261,639	410	23	65

Abbreviations: AHA, American Hospital Association; HCUP, Healthcare Cost and Utilization Project; LTAC, long-term acute care hospital; PO, post office; SID, State Inpatient Databases

The initial HCUP data were 2018 SID from 47 States and the District of Columbia (not included are Alabama, Idaho, and New Hampshire) with a total of 34,254,432 discharges, 41,635 unique

patient ZIP Codes, and 1,141,891 unique hospital/patient ZIP Code pairs. Following exclusions, the final analytic universe for this report as produced in step 5 of Table 2 was 30,278,849 discharges, 30,419 patient ZIP Codes, and 690,451 unique hospital/patient ZIP Code pairs. Compared with the initial data, the final analytic universe had:

- An 11.6 percent reduction in discharges, primarily due to discharges that were excluded for patients transferred into the hospital (step 4). Discharges were retained for patients transferred out to another hospital so that the initial hospital destination for transferred patients was retained and the same episode of care was not counted twice.
- A 26.9 percent reduction in patient ZIP Codes, primarily due to the exclusion of data for patients residing in Alaska, Hawaii, and the U.S. territories (step 2).
- A 39.5 percent reduction in hospital/patient ZIP Code pairs, primarily due to the exclusion of extreme travel distances (step 5).

Steps 6 and 7 represent an approximately one-third sample of hospital/patient ZIP Code pairs used to obtain driving distances from Google Maps (Scenario 2). Discharges in the analytic sample were weighted to represent the universe of all discharges (step 5) in the analysis, as subsequently described.

METHODS OF GEOCODING AND CALCULATING TRAVEL DISTANCE

Patient Geocoding

HCUP data only provide patient residence ZIP Code; more specific patient residence location information (e.g., patient street address) is not part of the HCUP data in order to protect patient identity.[29] To approximate the patient location for calculating travel distances, we used two different patient geocoding methods:

- Method (A), used for the Baseline scenario, obtains the latitude and longitude coordinates of the geographic centroid of the patient's ZIP Code using the SAS dataset SASHELP.ZIPCODE.[30]
- Method (B), used for Scenarios 1 and 2, obtains the latitude and longitude coordinates of the population-weighted centroid of the patient's ZIP Code using the Esri 2018 USA Zip Code Points data file.

Hospital Geocoding

We used two different hospital geocoding methods to determine the hospital location for calculating travel distances:

- Method (A), used for the Baseline scenario, obtains the hospital geocode (latitude and longitude coordinates) directly from the AHA 2018 Annual Survey dataset.
- Method (B), used for Scenarios 1 and 2, obtains the hospital geocode using the hospital's address from the AHA Annual Survey and Google Maps.

---

[29] Note that not all HCUP Partners release patient ZIP Code on the publicly available SID.

[30] Hadden LS, Zdeb MS. Paper 219-2010: ZIP Code 411: Decoding SASHELP.ZIPCODE and Other SAS® Maps Online Mysteries. SAS Global Forum 2010. www.support.sas.com/resources/papers/proceedings10/219-2010.pdf. Accessed May 26, 2021.

The hospital geocodes under Method (B) were originally obtained for all hospitals in the 2018 HCUP data (including hospitals that were not part of this report) as part of the development of the HCUP Hospital Market Structure (HMS) files. This geocoding process is summarized below:

- Hospital geocodes were obtained using the Google Maps Application Programming Interface (API) web service. The *ggmap* function in R was used to submit URL requests with hospital address information to the Google Maps Platform Geocoding API, which returned the hospital coordinates.[31,32] Example request:

  https://maps.googleapis.com/maps/api/geocode/json?address=123+Main+Street,+Anytown,+IL &key=YOUR_API_KEY

- Up to four versions of hospital address were submitted, as necessary, to attempt to obtain a hospital geocode from Google Maps:
  - Hospital address
  - Hospital name, State, and ZIP Code
  - Hospital name and State
  - Hospital name and ZIP Code
- For hospitals for which a Google Maps geocode was obtained from the API, the distance between the Google Maps geocode and the AHA geocode was calculated. If the difference in distance was < 1 kilometer, the Google Maps geocode was retained. If the difference in distance was ≥ 1 kilometer, or if no Google Maps geocode was obtained, hospitals were manually reviewed using publicly available tools (e.g., Bing, Google Maps, TomTom, and the American Hospital Directory [ahd.com]) to determine the most accurate final hospital location. In total, less than 5 percent of all hospital records required manual review.

Patient-Hospital Travel Distance

Two different patient-hospital distance measures were used to calculate the travel distance between the patient and hospital.

- Method (A), used for the Baseline scenario and Scenario 1, obtains the geodetic (shortest or "straight-line") distance between the patient and hospital geocodes using the SAS GEODIST function.[33]
- Method (B), used for Scenario 2, obtains the driving distance between the patient and hospital geocodes using the Google Maps API web service.

The process for obtaining travel distances under Method (B) is summarized below:

[31] Kahle D, Wickham H. ggmap: Spatial Visualization with ggplot2. The R Journal. 2013;5(1):144–61.
[32] Google Maps Platform. Documentation, Web Services: Geocoding API. Updated May 26, 2021. www.developers.google.com/maps/documentation/geocoding/overview?_ga=2.44829095.2029319474.1622070491-425700908.1610059053. Accessed May 28, 2021.
[33] SAS Help Center. GEODIST Function. Updated May 17, 2021. www.documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/lefunctionsref/n1korpfg2e18lon1nwpow9qijdxe.htm. Accessed May 26, 2021.

- The *gmapsdistance* function in R was used to submit URL requests with patient and hospital geocodes to the Google Maps Platform Distance Matrix API, which returned the driving distance between the coordinates.[34,35] Example request:

  https://maps.googleapis.com/maps/api/distancematrix/json?units=imperial&origins=40.330161, -89.855780&destinations=40.330161,-90.078253&key=YOUR_API_KEY

- Although API requests were submitted in batches, there was substantial time involved to obtain the Google Maps driving distances. As a result, we decided to obtain driving distances for a sample of our data rather than the full dataset. A total of 15 batches of hospital/patient ZIP Code geocode pairs were submitted in the order shown below. The first three batches were test batches used to determine the optimal size and configuration of data to submit using the API. Pairs sampled in a prior batch were excluded before the next batch was selected. For example, before selecting the random pairs in Batch 2, Rhode Island pairs (Batch 1) were excluded.

  1. Batch 1: all pairs with a hospital location in Rhode Island (1,704 patient-hospital pairs)
  2. Batch 2: 9,836 randomly sampled pairs
  3. Batch 3: 10,164 pairs for 63 selected hospitals
  4. Batches 4–15: 20,000 randomly sampled pairs per batch

- Discharge data from sampled pairs were weighted for Scenario 2 to represent the universe of all discharges. Pairs in Batch 1 (Rhode Island) and Batch 3 (select hospitals) were weighted 1 because they were sampled with certainty. The discharge counts and weights are summarized in Table 3.

Table 3. Weights applied to discharges sampled for Scenario 2

Pairs sampled	Sample, discharges	Universe, discharges	Weight
Certainty pairs	489,370	489,370	1.000000
Probability pairs	10,807,497	29,789,479	2.756372
Total	11,296,867	30,278,849	

- Driving distances in Google Maps are based on the road nearest to the provided geocodes. For instance, if the geocode for the geographic centroid of the ZIP Code (patient location) is in the middle of a lake, Google Maps will start the distance calculation at the nearest road to the provided coordinates.
- A very small number of submitted pairs (65) were returned with no driving distance— "route not found." We examined several of these cases and found that there was no drivable route between the patient and hospital (e.g., the patient location was on an island with no road possible to the hospital).
- Of the universe of 690,451 unique hospital/patient ZIP Code pairs, 261,704 were sampled to obtain driving distances (37.9 percent). Valid driving distances were

---

[34] Azuero Melo R, Rodriguez DT, Zarruk D. Package 'gmapsdistance'. August 28, 2018. www.cran.r-project.org/web/packages/gmapsdistance/gmapsdistance.pdf. Accessed June 1, 2021.
[35] Google Maps Platform. Documentation, Web Services: Distance Matrix API. Updated May 26, 2021. www.developers.google.com/maps/documentation/distance-matrix/overview. Accessed May 26, 2021.

obtained for 261,639 of the 261,704 sampled hospital/patient ZIP Code pairs submitted to Google Maps (99.98 percent).

COMPARING METHODS OF GEOCODING AND CALCULATING TRAVEL DISTANCE

Comparing Methods of Geocoding Patients and Hospitals

Table 4 compares the distance between the two patient geocoding methods (geographic centroid vs. population-weighted centroid of the patient's ZIP Code) and the distance between the two hospital geocoding methods (Google Maps vs. AHA). For reference, the median area of the patient's ZIP Code in square miles is also provided. Analysis is based on the universe of 30,278,849 discharges from 690,451 unique hospital/patient ZIP Code pairs.

The median and interquartile range (IQR) are provided overall and for three patient geographic areas: census region, census division, and urban/rural location. Census region and division are defined by the U.S. Census Bureau geography as illustrated in Figure 1.

Figure 1. Nine U.S. census divisions across four U.S. census regions
Northeast (orange), Midwest (grey), South (blue), and West (green)
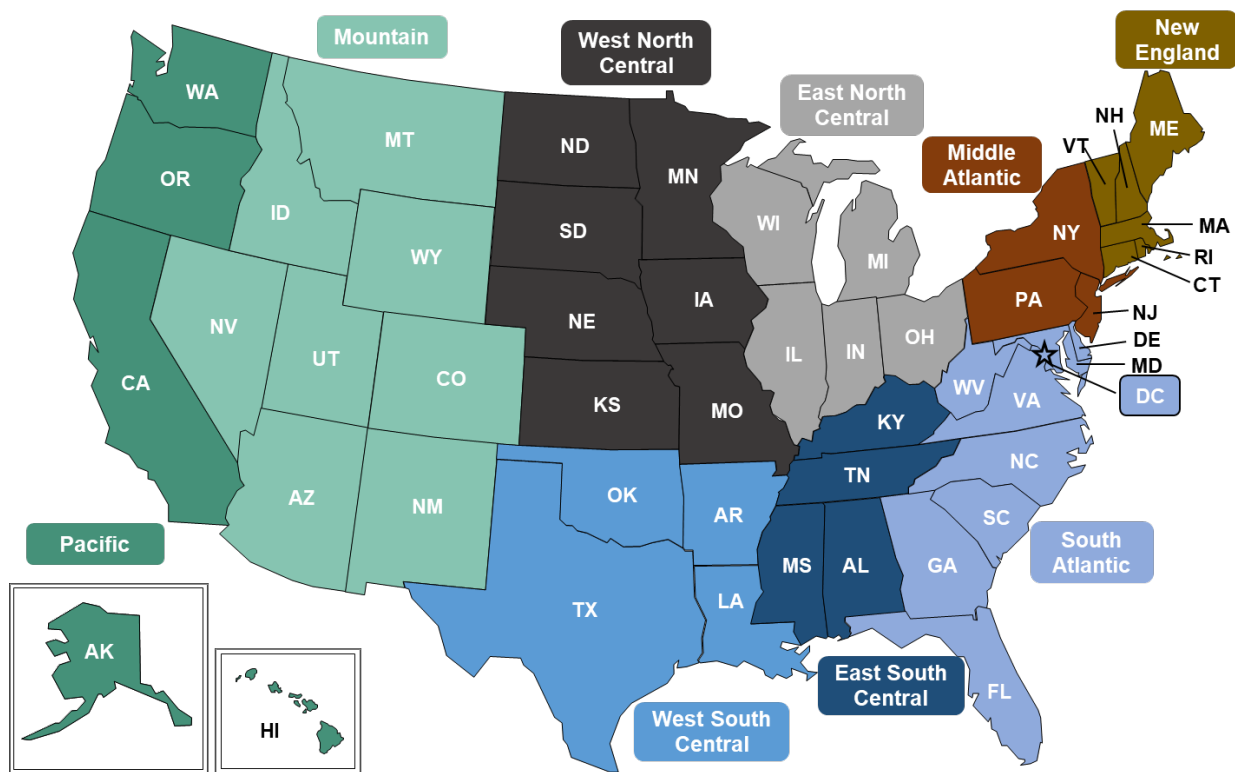
Urban/rural location uses the classification scheme for U.S. counties developed by the National Center for Health Statistics (NCHS). This classification scheme is based on the Office of Management and Budget (OMB) definition of a metropolitan service area as including a city and a population of at least 50,000 residents.[36]

- Large Central Metropolitan: Counties in a metropolitan area with 1 million or more residents that satisfy at least one of the following criteria: (1) containing the entire population of the largest principal city of the metropolitan statistical area (MSA), (2) having their entire population contained within the largest principal city of the MSA, or (3) containing at least 250,000 residents of any principal city in the MSA
- Large Fringe Metropolitan: Counties in a metropolitan area with 1 million or more residents that do not qualify as large central metropolitan counties
- Medium Metropolitan: Counties in a metropolitan area of 250,000–999,999 residents
- Small Metropolitan: Counties in a metropolitan area of 50,000–249,999 residents
- Micropolitan: Counties in a nonmetropolitan area of 10,000–49,999 residents
- Noncore: Counties in a nonmetropolitan and nonmicropolitan area

Supplement 1 provides additional geocoding distance difference statistics (minimum and maximum values, and the 1st, 5th, 10th, 90th, 95th, and 99th percentiles) as well as statistics for census division by urban/rural location.

---

[36] ZIP Codes were assigned to counties using data from Claritas, a vendor that produces population estimates and projects based on data from the U.S. Census Bureau. (Claritas. Claritas Demographic Profile by ZIP Code. www.claritas360.claritas.com/mybestsegments/. Accessed January 22, 2021.) Approximately 10 percent of the ZIP Codes could not be found in the Claritas data file. For these ZIP Codes, we obtained the county assignment from the SAS SASHELP.ZIPCODE data file.

Table 4. Difference (in miles) between geocode methods, by patient region, division, and urban/rural location, 2018

Patient geographic area	Number of discharges	Distance between patient geocodes (geographic vs. population centroids)		Distance between hospital geocodes (Google Maps vs. AHA)		Median area of patient's ZIP Code, square miles
		Median	IQR	Median	IQR	
Overall	**30,278,849**	**0.57**	**0.95**	**0.02**	**0.04**	**21.21**
Census region						
Northeast	5,642,604	0.36	0.62	0.02	0.05	9.38
Midwest	6,813,637	0.54	0.79	0.02	0.05	24.84
South	11,840,725	0.74	1.24	0.02	0.05	33.82
West	5,981,883	0.53	0.86	0.02	0.03	13.47
Census division						
New England	1,316,203	0.40	0.57	0.02	0.04	13.77
Middle Atlantic	4,326,401	0.34	0.63	0.03	0.05	6.90
East North Central	4,756,279	0.53	0.77	0.03	0.05	20.49
West North Central	2,057,358	0.55	0.84	0.01	0.04	41.66
South Atlantic	6,593,006	0.72	1.17	0.02	0.05	28.99
East South Central	1,485,736	0.95	1.47	0.02	0.07	71.81
West South Central	3,761,983	0.71	1.28	0.02	0.04	34.64
Mountain	1,869,929	0.76	1.32	0.02	0.03	19.71
Pacific	4,111,954	0.46	0.71	0.02	0.03	11.87
Urban/rural location						
Large central metropolitan	9,476,906	0.34	0.46	0.03	0.04	7.36
Large fringe metropolitan	7,495,376	0.56	0.89	0.02	0.04	19.26
Medium metropolitan	6,278,831	0.69	1.04	0.02	0.05	29.30
Small metropolitan	2,789,001	0.91	1.30	0.02	0.04	66.30
Micropolitan	2,487,059	1.06	1.54	0.02	0.05	127.89
Noncore	1,751,676	1.16	1.80	0.02	0.04	141.11

Abbreviation: AHA, American Hospital Association; IQR, interquartile range

Overall, there was a small difference in the patient location based on the two geocoding methods and an even smaller difference in the hospital location based on the two geocoding methods.

For patient location, the median difference in the patient geocodes (geographic centroid vs. population-weighted centroid of the patient's ZIP Code) was approximately 1/2 mile (0.57 miles) with an IQR of 0.95 miles. By census region, the largest difference was in the South and the smallest difference was in the Northeast (median difference of 0.74 vs. 0.36 miles). By census division, the largest difference was in the East South Central division and the smallest difference was in the Middle Atlantic division (median difference of 0.95 vs. 0.34 miles). By urban/rural location, the difference increased with rurality, from a 0.34 mile median difference in metropolitan areas to a 1.16 mile median difference in noncore areas. These results reflect differences in the size of the patient's ZIP Code. For example, the largest median ZIP Code

area is in the South and the smallest median ZIP Code area is in the Northeast (33.82 vs. 9.38 square miles). Similarly, ZIP Code area increases with increasing rurality.

For hospital location, the median difference in the hospital geocodes (Google Maps vs. AHA) was only 0.02 miles overall with an IQR of 0.04. The IQR did not exceed 1/10 of 1 mile for any of the geographic areas.

Comparing Methods of Calculating Travel Distance

Travel Distances

Figure 2 presents the distribution of travel distances observed under the Baseline scenario (patient location based on geographic centroid of ZIP Code, hospital location based on AHA coordinates, straight-line distance measure) including the median, 75th, 90th, 95th, and 99th percentiles. Because extreme travel distances were excluded from the Baseline scenario, the maximum observed travel distance was 250 miles.

Figure 2. Frequency of travel distances under the Baseline scenario, 2018
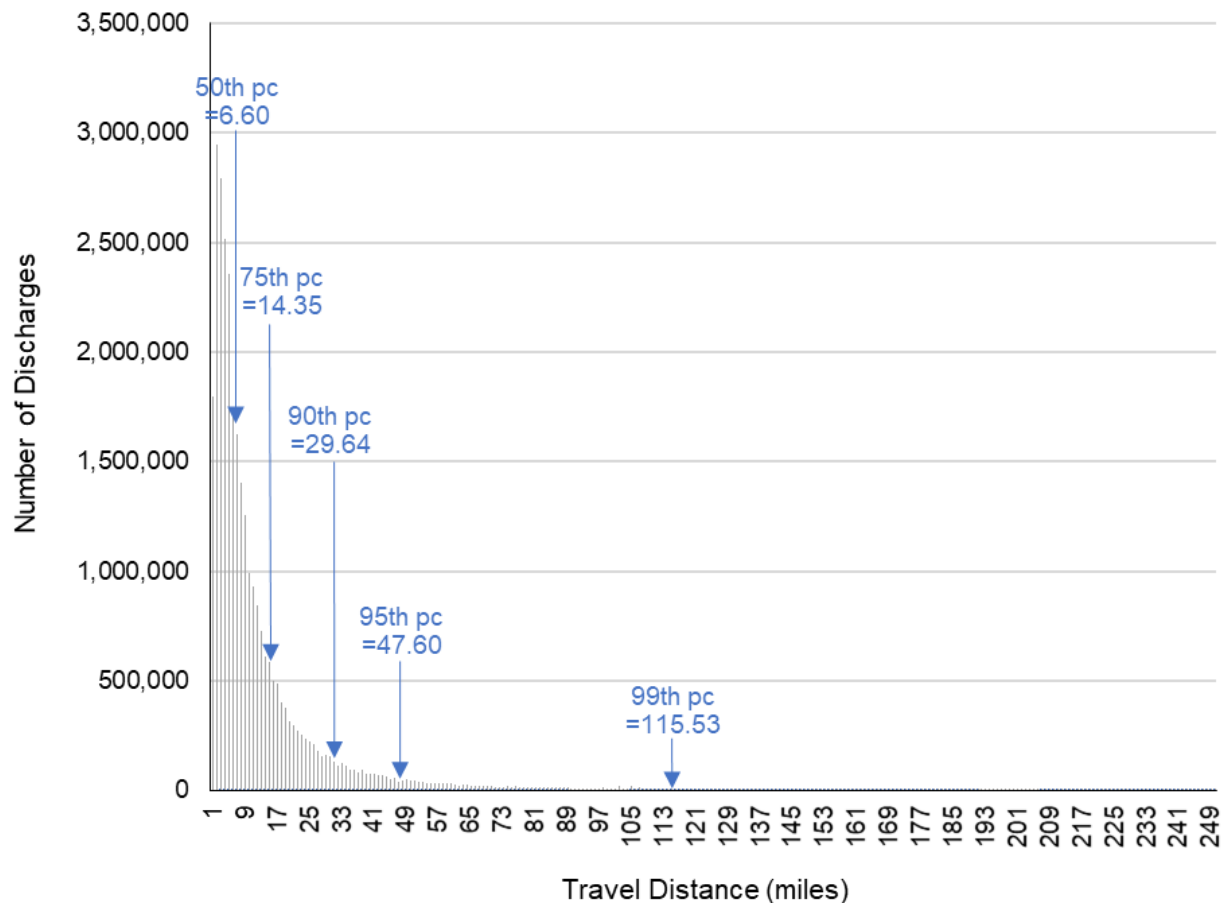
Abbreviation: pc, percentile

The median travel distance under the Baseline scenario was 6.60 miles. Half of all travel distances fell between 3.02 and 14.35 miles (IQR=11.33 miles). A total of 80 percent of distances fell between 1.39 and 29.64 miles (between the 10th and 90th percentiles). Fully 99 percent of travel distances were 115.53 miles or less.

Table 5 provides travel distances for each of the three scenarios:

- <u>Baseline</u>: patient location defined as the geocode of the geographic centroid of the patient's ZIP Code, hospital location defined as the AHA geocode, and straight-line distance measure
- <u>Scenario 1</u>: patient location defined as the geocode of the population-weighted centroid of the patient's ZIP Code, hospital location defined as the Google Maps geocode, and straight-line distance measure
- <u>Scenario 2</u>: patient location defined as the geocode of the population-weighted centroid of the patient's ZIP Code, hospital location defined as the Google Maps geocode, and Google Maps driving distance measure

Analysis is based on 30,277,975 discharges for which a distance measure was available for each of the three scenarios (i.e., the universe of 30,278,849 discharges less 874 discharges for which no valid driving distance was available under Scenario 2). The median travel distance and IQR are provided overall and for three patient geographic areas: census region, census division, and urban/rural location. Supplement 2 provides additional travel distance statistics (minimum and maximum values, and the 1st, 5th, 10th, 90th, 95th, and 99th percentiles) as well as statistics for census division by urban/rural location.

Table 5. Travel distance (in miles) by scenario, patient region, division, and urban/rural location, 2018

Patient geographic area	Number of discharges	Baseline		Scenario 1		Scenario 2*	
		Median	IQR	Median	IQR	Median	IQR
Overall	**30,277,975**	**6.60**	**11.33**	**6.38**	**11.22**	**8.70**	**14.52**
Census region							
Northeast	5,668,069	5.37	8.78	5.23	8.78	7.19	11.80
Midwest	6,796,353	6.44	12.03	6.31	12.13	8.51	15.41
South	11,821,334	7.83	13.50	7.74	13.35	10.37	17.07
West	5,992,220	5.83	9.36	5.49	9.19	7.62	12.25
Census division							
New England	1,279,539	5.35	10.17	5.41	10.45	7.03	13.44
Middle Atlantic	4,388,530	5.40	8.29	5.20	8.32	7.22	11.29
East North Central	4,729,934	6.25	10.62	6.12	10.70	8.18	13.95
West North Central	2,066,419	7.41	16.41	7.18	16.27	9.63	19.81
South Atlantic	6,554,272	7.07	11.92	6.98	11.55	9.40	15.16
East South Central	1,490,042	10.11	19.51	10.37	18.79	13.51	22.69
West South Central	3,777,020	8.61	14.29	8.40	14.59	11.11	18.65
Mountain	1,805,465	6.53	10.58	5.84	10.11	7.91	13.27
Pacific	4,186,755	5.57	8.89	5.37	8.87	7.44	11.69
Urban/rural location							
Large central metropolitan	9,456,092	5.09	6.34	4.99	6.32	6.79	8.85
Large fringe metropolitan	7,522,861	7.42	10.88	7.32	10.90	9.77	14.62
Medium metropolitan	6,270,798	6.27	10.11	6.18	9.87	8.27	12.34
Small metropolitan	2,817,827	5.98	12.95	5.41	12.97	7.38	16.13
Micropolitan	2,398,600	14.20	31.30	13.78	31.25	16.89	38.57
Noncore	1,811,797	26.21	31.10	26.05	31.73	32.13	38.77

Abbreviation: IQR, interquartile range

Notes: The analysis for all three scenarios excluded 874 discharges associated with 65 unique patient-hospital pairs that had no driving distance (returned from Google Maps with "route not available") under Scenario 2. For Scenario 2, analysis is based on weighted discharges.

The travel distance was quite similar between the Baseline scenario and Scenario 1, both of which used the straight-line distance measure. The distance was longer for Scenario 2, which used the driving distance measure.

The median straight-line distance was approximately 6 1/2 miles (6.60 miles under Baseline and 6.38 miles under Scenario 1) with an IQR of about 11 miles (11.33 miles under Baseline and 11.22 miles under Scenario 1). Under the Baseline scenario (with similar results for Scenario 1):

- By census region, the longest travel distance was in the South and the shortest distance was in the Northeast (7.83 vs. 5.38 miles). The greatest variability in distances also was in the South, and the least variability was in the Northeast (IQR=13.50 vs. 8.78).

- By census division, the travel distance was nearly twice as long in the East South Central division as in the Middle Atlantic division (10.11 vs. 5.40 miles) with the greatest and least variability in distances in these divisions as well (IQR=19.51 vs. 8.29).
- By urban/rural location, the longest travel distance was in noncore areas and the shortest distance was in large central metropolitan areas (26.20 vs. 5.09 miles). The greatest variability in distance was in the most rural areas (IQRs, micropolitan=31.30 miles; noncore=31.10 miles), and the smallest variability in distance was in the most urbanized areas (IQR large central metropolitan=6.34 miles).

The median driving distance was approximately 30 percent longer than the median straight-line distance, overall and by patient geographic area. The pattern of driving distances across geographic areas was similar to the pattern for straight-line distances.

- The median driving distance was 8.70 miles with an IQR of 14.52 miles.
- The longest and most variability in driving distance was in the South, and the shortest and smallest variability in driving distance was in the Northeast (median=10.37 vs. 7.19 miles; IQR=17.07 vs. 11.80 miles).
- The median driving distance ranged from 6.79 miles in large central metropolitan areas to 32.10 miles in noncore areas. The greatest variability in driving distance was in rural areas (IQRs, micropolitan=38.57 miles; noncore, 31.10 miles), and the smallest variability was in large central metropolitan areas (IQR=8.85 miles).

Difference in Travel Distances

Table 6 provides the difference in travel distances across the three scenarios examined in this report:

- Total difference: difference in travel distance between Scenario 2 (driving distance with population-weighted ZIP Code centroid geocode of patient location and Google Maps geocode of hospital location) and Baseline (straight-line distance with geographic ZIP Code centroid geocode of patient location and AHA geocode of hospital location)
- Geocoding difference: difference in travel distance between Scenario 1 (straight-line distance with population-weighted ZIP Code centroid geocode of patient location and Google Maps geocode of hospital location) and Baseline (straight-line distance with geographic ZIP Code centroid geocode of patient location and AHA geocode of hospital location)
- Distance metric difference: difference in travel distance between Scenario 2 (driving distance with population-weighted ZIP Code centroid geocode of patient location and Google Maps geocode of hospital location) and Scenario 1 (straight-line distance with population-weighted ZIP Code centroid geocode of patient location and Google Maps geocode of hospital location)

Analysis is based on 30,277,975 discharges for which a distance measure was available for each of the three scenarios (i.e., the universe of 30,278,849 discharges less 874 discharges for which no valid driving distance was available under Scenario 2). The median difference in travel distances and IQR are provided for three patient geographic areas: census region, census division, and urban/rural location. Supplement 3 provides additional travel distance

difference statistics (minimum and maximum values, and the 1st, 5th, 10th, 90th, 95th, and 99th percentiles) as well as statistics for census division by urban/rural location.

Table 6. Difference in travel distance (in miles) between scenarios, by patient region, division, and urban/rural location, 2018

Patient geographic area	Number of discharges	Scenario 2 vs. Baseline (total difference)		Scenario 1 vs. Baseline (geocoding difference)		Scenario 2 vs. Scenario 1 (distance metric difference)	
		Median	IQR	Median	IQR	Median	IQR
Overall	**30,277,975**	**1.97**	**3.55**	**(0.04)**	**0.65**	**2.00**	**3.33**
Census region							
Northeast	5,668,069	1.58	3.21	(0.04)	0.41	1.71	2.98
Midwest	6,796,353	1.88	3.41	(0.03)	0.60	1.90	3.28
South	11,821,334	2.31	3.96	(0.03)	0.86	2.33	3.66
West	5,992,220	1.80	3.10	(0.08)	0.67	1.89	2.97
Census division							
New England	1,279,539	1.58	3.28	(0.06)	0.46	1.76	3.03
Middle Atlantic	4,388,530	1.58	3.14	(0.04)	0.40	1.66	2.97
East North Central	4,729,934	1.77	3.20	(0.05)	0.60	1.77	3.05
West North Central	2,066,419	2.14	3.97	(0.00)	0.58	2.21	3.86
South Atlantic	6,554,272	2.13	3.63	(0.02)	0.82	2.19	3.35
East South Central	1,490,042	2.87	4.70	0.02	1.13	2.67	4.13
West South Central	3,777,020	2.45	4.12	(0.04)	0.81	2.47	3.93
Mountain	1,805,465	1.78	3.47	(0.10)	0.93	1.89	3.32
Pacific	4,186,755	1.80	3.00	(0.07)	0.60	1.89	2.87
Urban/rural location							
Large central metropolitan	9,456,092	1.66	2.63	(0.04)	0.41	1.71	2.52
Large fringe metropolitan	7,522,861	2.22	3.62	(0.05)	0.63	2.27	3.44
Medium metropolitan	6,270,798	1.79	3.12	(0.03)	0.77	1.81	2.79
Small metropolitan	2,817,827	1.67	3.70	(0.08)	1.07	1.68	3.21
Micropolitan	2,398,600	2.87	6.67	0.01	1.19	2.71	6.45
Noncore	1,811,797	5.17	9.02	(0.03)	1.24	5.25	8.73

Abbreviation: IQR, interquartile range

Notes: The analysis for all three scenarios excluded 874 discharges associated with 65 unique patient-hospital pairs that had no driving distance (returned from Google Maps with "route not available") under Scenario 2. For Scenario 2, analysis is based on weighted discharges.

The greatest difference in travel distance was due to the definition of the distance metric (straight-line vs. driving distance) and not due to improvements in patient or hospital geocoding.

The total difference in travel distance, accounting for both the distance metric and geocoding methods used, which is represented by the comparison of Scenario 2 versus Baseline, was approximately 2 miles (median=1.97 miles). Geographically, the largest differences were in the South (2.31 miles), the East South Central division (2.87 miles), and noncore areas (5.18 miles). The smallest differences were in the Northeast (1.58 miles), the New England and Middle

Atlantic divisions (1.58 miles), and large central and small metropolitan areas (1.66 and 1.67 miles, respectively).

The geocoding difference, reflecting the different methods used for patient and location geocoding, which is represented by the comparison of Scenario 1 versus Baseline, was nearly zero (median = –0.04 miles). This was true across geographic areas, with the largest difference in the Mountain division (median = –0.10 miles).

The distance metric difference, which is represented by the comparison of Scenario 2 versus Scenario 1, was 2 miles longer for the driving than straight-line distance. Geographically, the difference varied from 1.66 miles longer driving than straight-line distance in the Middle Atlantic to 2.67 miles longer in the East South Central division, and from less than 2 miles longer in large central metropolitan, medium metropolitan, and small metropolitan areas to 5.24 miles longer in noncore areas.

Table 7 compares the distribution of the travel distances across the three scenarios using the root mean square error (RMSE). The RMSE is a measure of the central tendency of the difference between the travel distance values at each percentile of the cumulative distribution function in one scenario compared with another scenario. The RMSE provides a measure of the average separation in miles between the distributions of travel distances under the two scenarios. Higher values indicate greater difference between the distribution of distance values in the two scenarios. For each patient geographic area, the RMSE was computed by calculating each percentile (1–99) across the distribution of travel distances for each scenario, then computing the difference in miles between the two scenarios at each percentile, squaring those differences, and averaging the squared differences across the distribution.

Table 7. RMSE of travel distance between scenarios, by patient region, division, and urban/rural location, 2018

Patient geographic area	Number of discharges	Scenario 2 vs. Baseline (total difference)	Scenario 1 vs. Baseline (geocoding difference)	Scenario 2 vs. Scenario 1 (distance metric difference)
		RMSE	RMSE	RMSE
Overall	**30,277,975**	**4.70**	**0.39**	**5.03**
Census region				
Northeast	5,668,069	4.05	0.14	4.11
Midwest	6,796,353	4.88	0.13	4.97
South	11,821,334	4.61	0.98	5.28
West	5,992,220	5.61	0.35	5.75
Census division				
New England	1,279,539	4.08	0.18	4.13
Middle Atlantic	4,388,530	3.99	0.16	4.09
East North Central	4,729,934	4.35	0.13	4.42
West North Central	2,066,419	6.22	0.23	6.35
South Atlantic	6,554,272	4.19	1.65	5.12
East South Central	1,490,042	5.86	0.26	5.95
West South Central	3,777,020	5.19	0.17	5.27
Mountain	1,805,465	7.30	0.59	7.58
Pacific	4,186,755	4.67	0.25	4.78
Urban/rural location				
Large central metropolitan	9,456,092	3.09	0.12	3.20
Large fringe metropolitan	7,522,861	4.11	0.16	4.23
Medium metropolitan	6,270,798	3.61	2.16	4.49
Small metropolitan	2,817,827	5.71	0.40	5.94
Micropolitan	2,398,600	8.13	0.45	8.22
Noncore	1,811,797	10.56	0.40	10.64

Abbreviation: RMSE, root mean square error

Notes: The analysis for all three scenarios excluded 874 discharges associated with 65 unique patient-hospital pairs that had no driving distance (returned from Google Maps with "route not available") under Scenario 2. For Scenario 2, analysis is based on weighted discharges.

The results when examining the RMSE of the travel distances between scenarios were consistent with the median results provided in Tables 5 and 6. In particular, the primary difference in travel distance was due to the distance metric (driving distance vs. straight-line distance) rather than the patient and hospital geocoding methods used.

The RMSE of the travel distance between Scenario 1 and Baseline, reflecting geocoding differences, was 0.39 miles. Geographically, the greatest differences were in the South region, the South Atlantic division, and medium metropolitan areas (RMSE=0.98, 1.65, and 2.16 miles, respectively).

The RMSE of the travel distance between Scenario 2 and Baseline, reflecting the distance metric difference, was 5.03 miles. Geographically, the greatest differences were in the West region, Mountain division, and noncore areas (RMSE=5.75, 7.58, and 10.64 miles, respectively).

A MODEL FOR CONVERTING STRAIGHT-LINE TO DRIVING DISTANCE

Straight-line travel distances are relatively easy to obtain (e.g., from SAS), especially for large-volume datasets such as HCUP that are often used for research. However, driving distances, which may be more difficult to obtain from mapping software, particularly in large volumes, may be of interest in some research applications. Accordingly, a simple approach to estimate driving distances from straight-line distances may be useful. Using a discharge-weighted sample (from Scenario 2), we ran a simple linear model with the driving distance as the response variable and the Baseline scenario travel distance, patient census division, and urban/rural location as the predictor variables. The linear model specification was as follows:

$$D_i = \alpha B i + \boldsymbol{C_i\beta} + \boldsymbol{L_i\gamma} + \varepsilon_i$$

Where:

- i indexes patients
- D_i : driving distance
- B_i : baseline distance
- $\boldsymbol{C_i}$: vector of census division dummy variables
- $\boldsymbol{L_i}$: vector of urban/rural location dummy variables
- α : coefficient for baseline distance
- $\boldsymbol{\beta}$: vector of coefficients for census division dummy variables
- $\boldsymbol{\gamma}$: vector of coefficients for urban/rural location dummy variables
- ε_i : mean-zero error term

The model fit was very high (R^2=0.96). Table 8 provides the model coefficients that can be used to estimate driving distance from straight-line distance.

Table 8. Coefficients to estimate driving distances from straight-line distances

Covariate	Coefficient
Baseline distance	**1.1797**
Census division	
New England	1.9454
Middle Atlantic	1.9970
East North Central	1.7714
West North Central	1.7067
South Atlantic	1.5099
East South Central	1.8704
West South Central	1.7828
Mountain	2.0252
Pacific	2.0232
Urban/rural location	
Large central metropolitan	–0.7346
Large fringe metropolitan	–0.5275
Medium metropolitan	–1.4546
Small metropolitan	–1.0043
Micropolitan	–0.8427
Noncore (reference category)	0.0000

Overall, the coefficient to adjust the baseline distance is 1.1797; thus, a 5-mile straight-line distance would convert to a 5.90 mile driving distance (5*1.1797) before incorporating geographic characteristics. The coefficients allow for further refinement to the driving distance estimate based on the patient residence census division and urban/rural location. For example, a 5-mile straight-line distance in a medium metropolitan area in the South Atlantic division translates to a 5.95-mile driving distance: (5*1.1797) + 1.5099 – 1.4546 = 5.95. A 5-mile straight-line distance in a noncore area in the Mountain division translates to a 7.92-mile driving distance: (5*1.1797) + 2.0252 – 0.0000 = 7.92.

The conversion of straight-line distance to driving distance varies by geographic area. The relative impact of geographic area on driving distance is larger for shorter distances. For example, for a patient residing in the Northeast in a large metropolitan area, the driving distance estimate is 2.4 miles for a corresponding 1-mile straight-line distance (more than double the distance), whereas the driving distance estimate is 36.6 miles for a corresponding 30-mile straight-line distance (less than one-fourth longer).

CONCLUSION

Patient-hospital travel distance is an important metric for a wide range of research involving patient access to healthcare and patient choice of care. HCUP State inpatient data include patient residence ZIP Code and AHA-identifiable hospitals. In this report, we compared two methods of geocoding patient location, two methods of geocoding hospital location, and two travel distance measures. Most of the difference in travel distance between methods is

associated with the distance measure (driving vs. straight-line) with little difference due to geocoding methods.

The SAS ZIP Code and AHA hospital source data used for geocoding patient and hospital locations did have some errors. We did not attempt to correct these data errors in order to identify the overall impact of the data on geocoding and travel distance differences. Even with some source data errors, there was little change in travel distance using more refined approaches to geocoding patient and hospital location. Depending on the number of hospital-patient ZIP Code pairs, and the level of precision desired, HCUP data users may wish to closely examine and potentially resolve any source data errors (e.g., AHA hospital location, SAS ZIP Code geographic centroids). The results presented in this report suggest that this step may be unnecessary in most cases given the minimal impact on the results, but data correction may be more important when studying a small geographic area with limited sample size.

HCUP data only provide patient residence ZIP Code and not complete address, so some imprecision exists in the patient location. The results presented in this report indicate that there is little difference in travel distance overall using patient geocodes defined as the geographic versus population-weighted centroids of the patient's ZIP Code. However, the impact of imprecise patient location may be greater in ZIP Codes that cover large areas.

The information presented in this report suggests approaches that researchers can use to estimate patient-hospital travel distances. The Baseline method (patient ZIP Code geographic centroid geocodes, AHA hospital geocodes, straight-line distance) provides a reasonable estimate of travel distances, especially for comparative purposes. Driving distance may be of more interest for studies where the distance itself is inherently important. In such cases, estimating driving distances from straight-line distances via the equation provided in this report may be a more tenable solution than obtaining driving distances directly from mapping software, especially in large volume.

APPENDIX. HCUP PARTNERS

Alaska Department of Health and Social Services

Alaska State Hospital and Nursing Home Association

Arizona Department of Health Services

Arkansas Department of Health

California Office of Statewide Health Planning and Development

Colorado Hospital Association

Connecticut Hospital Association

Delaware Division of Public Health

District of Columbia Hospital Association

Florida Agency for Health Care Administration

Georgia Hospital Association

Hawaii Laulima Data Alliance

Hawaii University of Hawai'i at Hilo

Illinois Department of Public Health

Indiana Hospital Association

Iowa Hospital Association

Kansas Hospital Association

Kentucky Cabinet for Health and Family Services

Louisiana Department of Health

Maine Health Data Organization

Maryland Health Services Cost Review Commission

Massachusetts Center for Health Information and Analysis

Michigan Health & Hospital Association

Minnesota Hospital Association

Mississippi State Department of Health

Missouri Hospital Industry Data Institute

Montana Hospital Association

Nebraska Hospital Association

Nevada Department of Health and Human Services

New Hampshire Department of Health & Human Services

New Jersey Department of Health

New Mexico Department of Health

New York State Department of Health

North Carolina Department of Health and Human Services

North Dakota (data provided by the Minnesota Hospital Association)

Ohio Hospital Association

Oklahoma State Department of Health

Oregon Association of Hospitals and Health Systems

Oregon Office of Health Analytics

Pennsylvania Health Care Cost Containment Council

Rhode Island Department of Health

South Carolina Revenue and Fiscal Affairs Office

South Dakota Association of Healthcare Organizations

Tennessee Hospital Association

Texas Department of State Health Services

Utah Department of Health

Vermont Association of Hospitals and Health Systems

Virginia Health Information

Washington State Department of Health

West Virginia Department of Health and Human Resources, West Virginia Health Care Authority

Wisconsin Department of Health Services

Wyoming Hospital Association