

***Geography, Geographic Information Systems and
Cancer Comparative Effectiveness Research***

Karl Eschbach, Ph.D.

Internal Medicine-Geriatrics and

Sealy Center on Aging

December 3, 2010



Working together to work wonders.™

Why geo-reference?

- Environmental exposures—
 - Communicable disease risks
- Social-economic exposures
 - As proxy for characteristics not measured on administrative records
 - For monitoring trends in relation to SES
 - As context for health risks and access barriers

Why geo-reference: Geography of health care resources:

- Distance as a barrier to access
- To define locally available resources (Capacity)
- Contexts for practice patterns
- Medically underserved areas

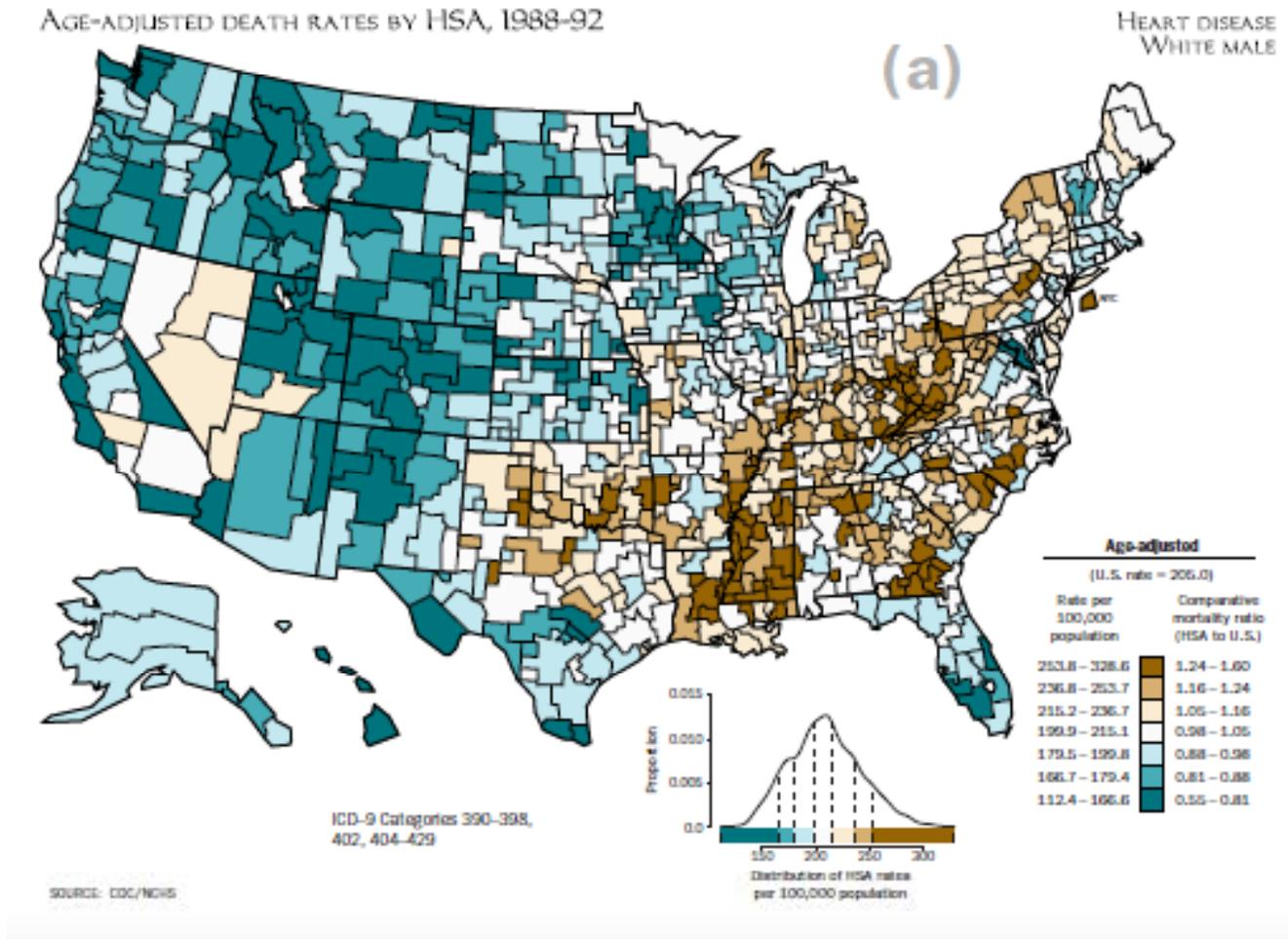
Modifiable Area Unit Problem

- Containing areas may be described at different scales
- For a given scale, boundaries are arbitrary
- Boundaries are often crossed
- Statistical relationships may vary with different aggregations
 - Ecological fallacy

Common issues for geographic analysis

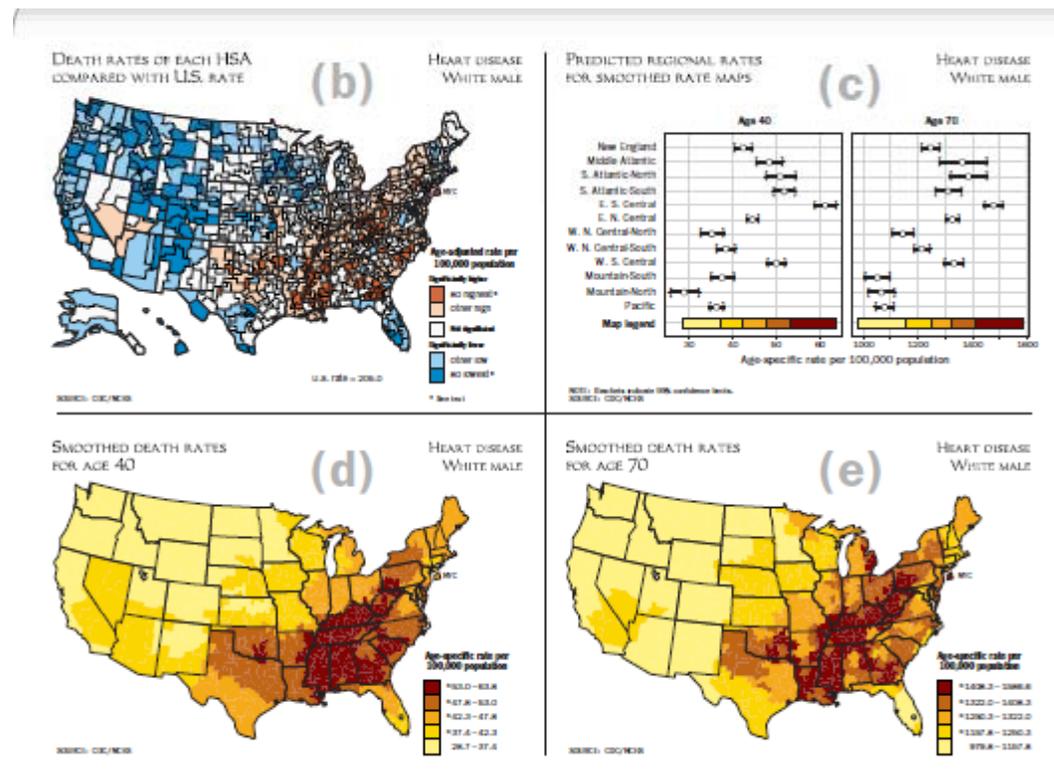
- Heteroskedasticity of rates in area units because of differences in size
- Spatial auto correlation
 - Contagion
 - Excluded variables and correlated errors
- Visual misrepresentation of relationships:
 - Size/population may be inversely related

Health Service Area—Mortality Rates for Heart Disease, White Males



Source: Centers for Disease Control Atlas of United States Mortality (1996)

Smoothed Rate Maps



Source: Centers for Disease Control Atlas of United States Mortality (1996)

Administrative units vs. analytic units

- Political boundaries: Cities, Counties, States
- Census geographies
- Zipcodes

VS

- Metropolitan areas
- Market areas for health care
- Neighborhoods

Administrative Units

- Subjectively meaningful
- County is a common unit for reporting population (Census and Census Estimates) and vital statistics
- Administrative entities sometimes control relevant resources, e.g. county hospital districts
- Analytically arbitrary
- Different in size and organization

Zip codes

- Not geographic areas
- Defined by USPS Service Route Manuals, which are constantly updated.
- Census ZIP code tabulation area (3/5 digit) is a time frozen approximation.

Analytical Units

- May also have different sizes
- Boundaries are not rigidly defined

Match Concept to Geography

- Why are you using geography in a particular context?
 - Access barriers
 - Resources or practice patterns
 - “Neighborhood” effects
 - Risk mapping
 - Denominators
 - Communicate with the public

Context vs. Composition

- Neighborhood effects: Contextual unit affects outcomes and behavior, net of individual characteristics
- Composition implies differences in socio-economic risk levels
- Confounding problems:
 - Selective sorting
 - Unmeasured variation in composition may drive apparent “neighborhood” effects

Where do denominators come from?

- Incidence: Census
- Survival: Registry or claims database
- Mortality: Registry, Claims or Census

Geographic Information Systems

- **Geographic information systems (GIS)** or **geospatial information systems** are sets of tools that capture, store, analyze, manage, and present data that are linked to location(s). In the simplest terms, GIS is the merging of cartography, statistical analysis, and database technology.

-Wikipedia

ESRI.com: ArcGIS®; ArcInfo®, ArcMap®, ArcView®

Geocoding: Putting Points on a Map and in Areas

- Compare to address ranges or parcels
- Zipcode and Zip+4
- Associate point with area by administrative look-up or point-in-polygon

- Based on string matches
- More difficult in areas without city style addresses
- Extremely labor intensive for the last n%
- Errors arising at the point of entry
- Errors in databases and in boundary files

MAF/TIGER® DATABASE

- MAF/TIGER® is an acronym for the Master Address File/Topologically Integrated Geographic Encoding and Referencing system or database. It is a digital (computer-readable) geographic database that automates the mapping and related geographic activities required to support the Census Bureau's census and survey programs.
- The Census Bureau developed the TIGER® System to automate the geographic support processes needed to meet the major geographic needs of the 1990 census: producing the cartographic products to support data collection and map presentations, providing the geographic structure for tabulation and dissemination of the collected statistical data, assigning residential and employer addresses to the correct geographic location and relating those locations to the geographic entities used for data tabulation, and so forth. During the 1990s, the Census Bureau developed an independent Master Address File (MAF) to support field operations and allocation of housing units for tabulations. After Census 2000, both the address-based MAF and geographic TIGER® databases merged to form MAF/TIGER®. The content of the MAF/TIGER® database is undergoing continuous updates and is made available to the public through a variety of TIGER®/Line shapefiles.
- <http://www.census.gov/geo/www/tiger/index.html>

Census Data

- Decennial Census (Through Census 2000)
 - Short Form
 - Race, Hispanic Origin, Age, Sex, Household Relationships, Housing Tenure
 - Long Form (1 in 6 sample)
 - Social economic domains including education, income and poverty, language, work and occupation, migration, nativity and citizenship, disability, commuting and vehicle ownership, house values and expenses, limited housing facilities

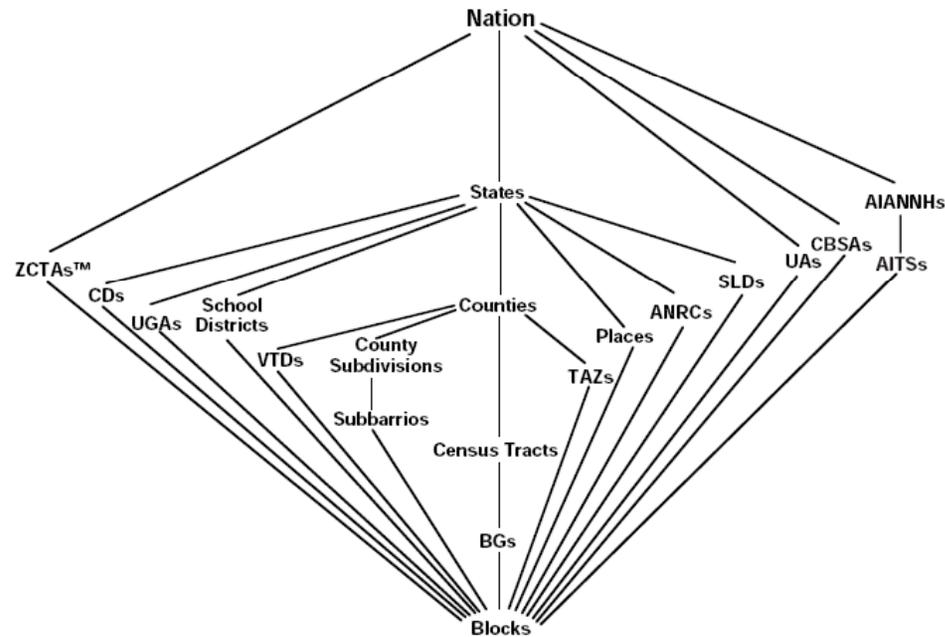
American Community Survey

- Enhanced long-form content
 - Health insurance
 - Fertility
- Continuous measurement
 - Data Release for
 - 1-year (65,000 person geographies)
 - 3-year (20,000 person geographies)
 - 5-year aggregations (all geographies)

ACS—2009 Sample Size and Fraction

Area	Population Estimate	Sample Count: Households + GQ	Percent	Occupied Housing Unit Estimate	Completed Household Interviews	Percent
United States	307,006,556	4,483,439	1.46%	113,616,229	1,917,748	1.69%
Texas	24,782,302	316,098	1.28%	8,527,938	128,120	1.50%
Harris County	4,070,989	44,486	1.09%	1,414,814	17,630	1.25%
Galveston County	286,814	3,618	1.26%	112,068	1,637	1.46%

Census Reporting Geographies (Summary Levels)



AIANNH: American Indian, Alaska Native, and Native Hawaiian area
 AITS: American Indian Tribal Subdivision
 ANRC: Alaska Native Regional Corporation
 BG: Block Group
 CD: Congressional District
 CBSA: Core Based Statistical Area (Metropolitan and Micropolitan Statistical Areas)
 SLD: State Legislative District
 TAZ: Traffic Analysis Zone
 UA: Urban Area
 UGA: Urban Growth Area
 VTD: Voting District
 ZCTA™: ZIP Code Tabulation Area

Census Blocks

- **Census blocks are statistical areas bounded by visible features, such as streets, roads, streams, and railroad tracks, and by nonvisible boundaries, such as city, town, township, and county limits, and short line-of-sight extensions of streets and roads. Generally, census blocks are small in area; for example, a block in a city bounded on all sides by streets. Census blocks in suburban and rural areas may be large, irregular, and bounded by a variety of features, such as roads, streams, and/or transmission line rights-of-way. In remote areas, census blocks may encompass hundreds of square miles. All territory in the United States, Puerto Rico, and the Island Areas has census block numbers. Census blocks nest within all other tabulated census geographic entities.**
- **(Approx 100 people). Limited data is available at this level of geography.**

Census Tracts

Census Tracts are small, relatively permanent statistical subdivisions of a county or equivalent entity and are updated by local participants prior to each decennial census as part of the Census Bureau's Participant Statistical Areas Program. The Census Bureau delineates census tracts in situations where no local participant existed or where local or tribal governments declined to participate. The primary purpose of census tracts is to provide a stable set of geographic units for the presentation of statistical data.

- Census tracts generally have a population size between 1,200 and 8,000 people with an optimum size of 4,000 people. The spatial size of census tracts varies widely depending on the density of settlement.

Signposts on the Royal Road

- My Census Block: 48167724402016
- 48= State of Texas
- 167=Galveston County
- 724400=Census tract number 724400
- 2=Block Group 2
- 016=Block 016

- Internal Point (Latitude)
- +29.298005
- Internal Point (Longitude)
- -094.793021

Where to Get Aggregated Census Data

- American Factfinder:
http://factfinder.census.gov/home/saff/main.html?_lang=en
- Underlying files: <http://www2.census.gov>
- And many people repackage and resell or give away

- Summary Files 1 and 2 (Short Form Subjects)
- Summary Files 3 and 4 (Long Form Subjects)
- (2 and 4 include data for detailed race/ethnic groups with some data suppression)
- American Community Survey: Summary Files

Census Public-Use Microdata (PUMS) Files

- Decennial Census (1980,1990,2000) 1 and 5% files
- American Community Survey: Annual 1% file and composite 3 and 5 year files
- Base Geography: Public-use microdata area: Areas with 100,000 persons: partly county-based.

Between Censuses

- Census Bureau produces county-level estimates by race/Hispanic Origin, gender and age
- Used and distributed by National Center for Health Statistics (National Vital Statistics System) and SEER
- Commercial providers make small area estimates:
Caveat Emptor
- ACS Files 5-year files will contain population estimates, DO NOT USE THEM FOR DENOMINATORS

Urban Concepts

- Incorporated place (City or town)
- Census Designated Place (e.g. The Woodlands)
- Urban: Highly complex definition based on core (1,000 persons per sq. mile) and peripheral density (500 ppsm)
 - Urbanized area (50,000+)
 - Urban cluster (10,000 < 50,000)
- Rural: Not urban
- http://www.census.gov/geo/www/ua/ua_2k.html
- Analytical concepts and data of rurality:
USDA Economic Research Service (Measuring Rurality):
<http://www.ers.usda.gov/Briefing/Rurality/>

Core-Based Statistical Areas

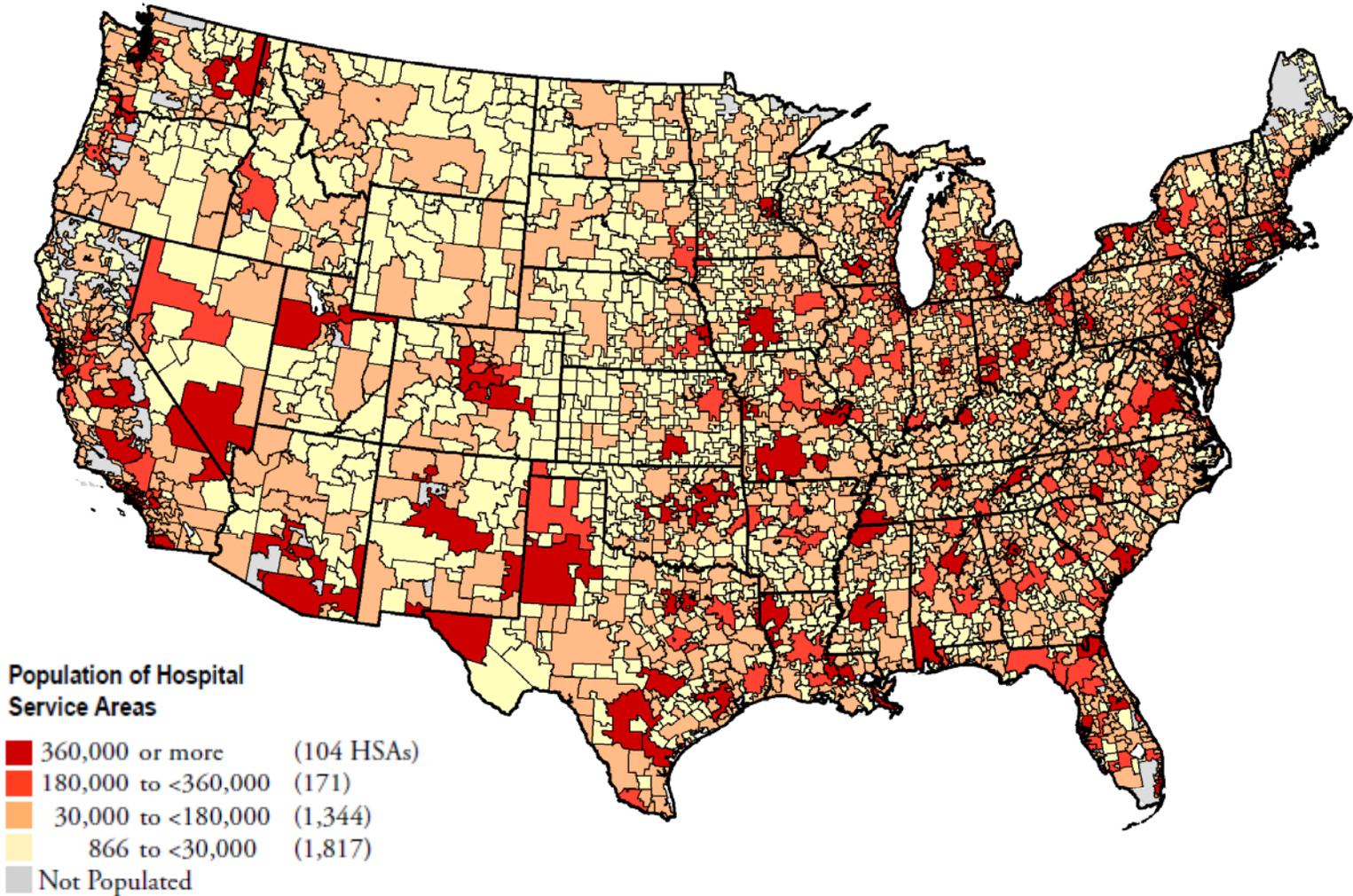
- Rules set by Office of Management and Budget
- Metropolitan Areas (50,000 Core)
- Micropolitan Areas: (10,000 Core < 50,000 Core)
- Non-Metropolitan: Not Metropolitan
 - Metro/non-metro does not equal urban/rural
- County-based

<http://www.census.gov/population/www/metroareas/metroarea.html>

Dartmouth Atlas Geographies

- **Hospital service areas (HSAs)** are local health care markets for hospital care. An HSA is a collection of ZIP codes whose residents receive most of their hospitalizations from the hospitals in that area. HSAs were defined by assigning ZIP codes to the hospital area where the greatest proportion of their Medicare residents were hospitalized. Minor adjustments were made to ensure geographic contiguity. This process resulted in 3,436 HSAs. When these regions were created in the early 1990s, most hospital service areas contained only one hospital. In the intervening years, hospital closures have left some HSAs with no hospital; these HSAs have been maintained as distinct areas in order to preserve the continuity of the database.
- Data served at on physician resources, hospital resources, Medicare reimbursements, procedure rates, mortality rates:
- <http://www.dartmouthatlas.org/tools/downloads.aspx>

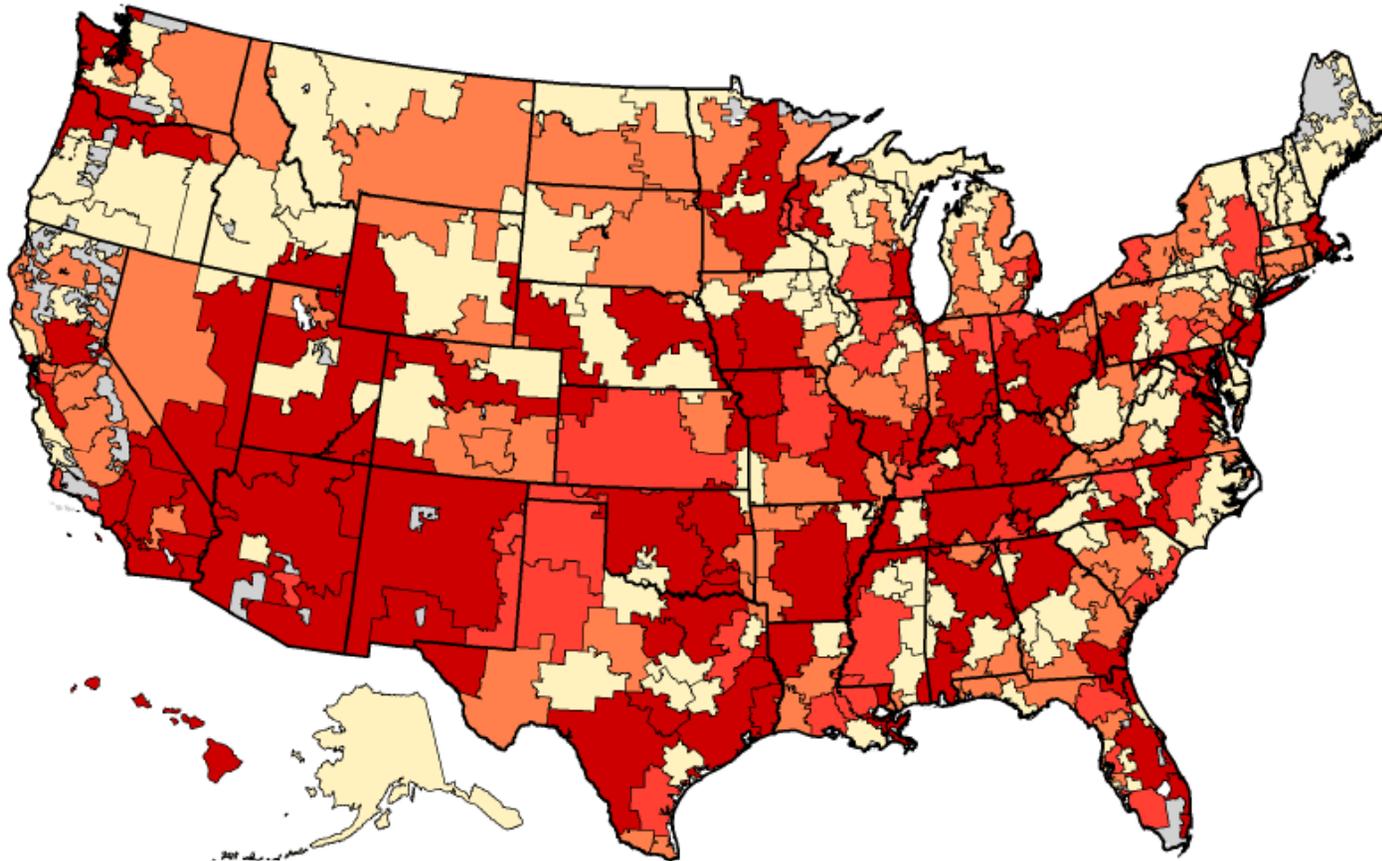
Hospital Service Areas (3,436)



Dartmouth Atlas Geographies

- **Hospital referral regions (HRRs)** represent regional health care markets for tertiary medical care that generally requires the services of a major referral center. The regions were defined by determining where patients were referred for major cardiovascular surgical procedures and for neurosurgery. Each hospital service area (HSA) was examined to determine where most of its residents went for these services. The result was the aggregation of the 3,436 hospital service areas into **306** HRRs. Each HRR has at least one city where both major cardiovascular surgical procedures and neurosurgery are performed.

Hospital Referral Regions (306)



Dartmouth Atlas Geographies

- **Primary care service areas (PCSAs)** reflect Medicare patient travel to primary care providers. Each of the 6,542 PCSAs includes a ZIP code area with one or more primary care providers and any contiguous ZIP code areas whose Medicare populations seek the plurality of their primary care from those providers.

DHA Definitions of Service Areas

- Zipcode-based analysis of market areas for primary, secondary, and tertiary care based on Medicare billing records.
- Hospital Service Areas are nested in Hospital Referral Regions
- Defined in early 1990s; Maintained for continuity purposes even as hospitals close

Limitations

- Though defined as market areas, there is considerable cross-commuting with respect to Hospital Service Areas.
- How do we allocate resources to service areas given travel?
- HSAs vary greatly in population.
- Some HSA areas may have lost their hospital.
- Most people live in large areas.

Other Resource Concepts: County Based

- NCHS Health Service Areas (802 defined in 1988 for continental U.S.)
 - County Based
- SEER Data at <http://seer.cancer.gov/seerstat/variables/countyattribs/hsa.html>
- Area Resource File: County-Based Resources and Populations

HRSA Underserved/Shortage Areas

- Health Professional Shortage Areas (HPSAs) are designated by HRSA as having shortages of primary medical care, dental or mental health providers.
- Medically Underserved Areas/Populations Designated by HRSA as having too few primary care providers, high infant mortality, high poverty, and/or high elderly population.
- County, tract or population based.
- Out of date; HRSA has been trying unsuccessfully to update concept for over a decade
 - http://datawarehouse.hrsa.gov/HPSA_MUA_Logic.aspx

Contact

Karl Eschbach, Ph.D.

Professor and Director of Population Research

Internal Medicine-Geriatrics

Sealy Center on Aging 0177

University of Texas Medical Branch

Galveston, TX 77555-0177

(409) 266-9680 (Office)

(409) 772-8931 (Fax)

(409) 392-1778 (Mobile)