

Introducing the Commonwealth Institute of Kentucky's Area Health Resource Files (AHRF) Data

Naiya Patel, BDS, MPH; Kelsey B. White, PhD, BCC; Seyed Karimi, PhD

1. Background

The Area Health Resource Files (AHRF), or Area Resource Files (ARF), data collection started in 1940. The dataset is a result of a collaboration among multiple federal organizations: the Department of Health and Human Services, the Health Resource and Services Administration, the National Center for Health Workforce Analysis, and the Bureau of Health Professions. The historical data files provide national-, state-, and county-level information to policymakers, researchers, and other professionals to make informed decisions on health policy, healthcare planning, and health services research (See: <https://www.cdc.gov/dhdsp/maps/gisx/resources/geo-spatial-data.html>).

2. Documentation

The AHRF datasets are released each year and can be found at <https://data.hrsa.gov/data/download>. They are available at the county, state, and national levels.

The archived files can be found in software-independent formats under the name of Area Resource Files (ARF) at <https://catalog.archives.gov/id/571544>.

The ARF/AHRF data files collect information on six broad areas:

- health care professions
- health facilities
- hospital utilization
- hospital expenditures
- population characteristics
- environment

The health professions category includes information about the supply of individuals within particular disciplines. For example, it reports the number of physicians by specialty, nurses,

and medical technicians.

The health facilities category includes information on the number of hospitals, veterans' hospitals, health maintenance organizations, and nursing homes.

The hospital utilization category includes information on the number of inpatient and outpatient visits to hospitals and other health care-giving facilities.

The hospital expenditure category includes information on Medicare reimbursements and medical facility expenditures such as capital and payroll expenses (costs of medical care).

The population characteristics category includes information on race, sex, and age of the current population, as well as natality and mortality statistics.

The environment category includes information on land use, weather characteristics, and large animal populations.

3. AHRF Summary Statistics

Many of the AHRF data are summarized at the national, state, and county levels and can be found at <https://data.hrsa.gov/data/download> and historical files at <https://catalog.archives.gov/id/571544>. In county files, there are about 3,230 observations and up to 6921 variables. In general, the data file can be divided into the following categories: health professions, health facilities, hospital utilization, hospital expenditures, population characteristics, and environment.

4. Commonwealth Institute of Kentucky's (CIK) AHRF Data

4.1. Years of Data

The CIK has AHRF datasets from 2014 to 2021.

4.2. Data Format

The Database includes data files in multiple formats, namely,

- SAS (2014-2021)
- Stata (2014-2021)

4.3. Components of the Data

For each year, there are about 6,921 variables with multiple years of data incongruent with a specific year data file. For example, the 2014 year data file might have several category variables data belonging to 2010-2013. Following are the AHRF county-level data components:

- **Health Professions (F14675-F14915):**
 1. Number of primary care physicians
 2. Total number of medical doctors (MDs)
 3. Number of MDs by specialty and major professional activity
 4. Number of Physicians by country of graduation
 5. Number of federal MDs
 6. Number of MDs by gender
 7. Number of MDs by specialty and age
 8. Number of doctors of osteopathic medicine (DOs)
 9. Number of dentists
 10. Number of physicians assistants
 11. Number of advanced practice registered nurses
 12. Number of nurse practitioners
 13. Number of clinical nurse specialists
 14. Number of nurse anesthetists
 15. Number of nurse midwives
 16. Number of chiropractors
 17. Number of optometrists
 18. Number of podiatrists
- **Health Facilities (F08868-F15261):**
 1. Total number of facilities
 2. Facilities by accreditation
 3. Number of admissions/readmissions
 4. Number of preventable hospital stays
 5. Number of beds
 6. Services
 7. Number of Inpatient service unit beds
 8. Number of personnel
 9. Distribution by services
 10. Number of providers of services
 11. Number of community health centers
 12. National Health Service Corps (NHSC)
- **Utilization (F09545- F09593):**
 1. Number of Inpatient days/discharges
 2. Utilization rate

3. Number of outpatient visits
 4. Number of emergency department visits
 5. Number of surgical operations
 6. Number of operating rooms
 7. Airborne infection isolation rooms
- **Expenditures (F09593-F15336):**
 1. Total hospital expenditures
 2. Medicare fee-for-service costs
 3. Medicare Advantage rates
 4. Veteran expenditures
 - **Population (F11984-F14619):**
 1. Population by race and gender
 2. Population by gender and age
 3. Population by race, gender, and age
 4. Median age
 5. Population by race detail
 6. Non-English speaking persons
 7. Urban/rural population
 8. Veteran population
 9. Medicare enrollment
 10. Medicaid eligible
 11. Census statistics
 12. Births
 13. Deaths
 14. Income statistics
 15. Poverty level statistics
 16. Divorces
 17. Food stamp/SNAP recipients
 18. Disabled persons
 19. Health insurance
 20. Social security/public assistance
 21. Education levels
 22. Employment
 23. Housing statistics
 - **Environment (F13874-F00811):**
 1. Land area and density
 2. Population per square mile
 3. Air quality
 4. Ground contamination
 5. Elevation

4.4. Documentation

The data comes with detailed technical documentation in which information on the field, column, year of data, variable name, variable characteristics, source of data, and the data collection date are provided. Since the raw data is software independent, these technical documentations for variables allow the conversion of the raw files into any data format. Also, the codebook Excel files are available from 2014 to 2020.

5. Key Data Limitations

- *Delay in Data Reporting:* Some variables for a specific year are not present in that particular year's dataset. For example, in the 2014 AHRF file, the value of some variables was collected in 2010. Therefore, additional data cleaning is needed if a researcher needs all variables to belong to a specific year.
- *5-Year Aggregates:* The demographic data after 2005 are 5-year aggregates, not yearly. This is due to the origin of that measure.
Missing Values: It is common to find variables with a significant number of missing values.
- *Measurement error:* Inability to confirm whether the counts of providers/professionals are connected to their place of practice or place of residence; potential miscalculation if provider/professional works in multiple counties.
- The smallest level of observation is the county, but utilization trends/dynamics are best examined at the Hospital Referral Region or Hospital Service Area.
- Limited definitions of who counts as a particular provider/professional type.

6. Highly Cited Peer-Reviewed Articles Used the AHRF Data

Caldwell JT, Ford CL, Wallace SP, Wang MC, Takahashi LM. Intersection of living in a rural versus urban area and race/ethnicity in explaining access to health care in the United States. *Am J Public Health*. 2016;106(8):1463-1469. doi:10.2105/AJPH.2016.303212

Andrilla CHA, Moore TE, Patterson DG, Larson EH. Geographic Distribution of Providers With a DEA Waiver to Prescribe Buprenorphine for the Treatment of Opioid Use Disorder: A 5-Year Update. *J Rural Heal*. 2019;35(1):108-112. doi:10.1111/jrh.12307

Cheng KJG, Sun Y, Monnat SM. COVID-19 Death Rates Are Higher in Rural Counties With Larger Shares of Blacks and Hispanics. *J Rural Heal*. 2020;36(4):602-608. doi:10.1111/jrh.12511