Can You Do Health Disparities Research with Publicly Available Datasets?

Danny R. Hughes, PhD

Rationale and Objectives: Given the growing importance of identifying and reducing health disparities, it is important for radiologist researchers to engage in this space to promote evidence-based imaging disparities policy. However, researchers are often hindered by access to appropriate data to perform quality research.

Materials and Methods: This paper reviews existing publicly available data sets that may be useful for performing imaging disparities research.

Results: Multiple data sources are publicly available and have been used by previous researchers to examine imaging disparities.

Conclusions: This paper provides an overview of publicly available data sources that radiologists can use for imaging disparities research. Appropriate use of these data sources will require researchers to carefully consider the overall research question and level of analysis.

Key Words: Disparities; health services research; radiology; health policy.

© 2018 Published by Elsevier Inc. on behalf of The Association of University Radiologists.

INTRODUCTION

Over the past three decades, the examination of health disparities has grown from a niche research area to an important component of the National Institutes of Health's research portfolio through the National Institute on Minority Health and Health Disparities (1,2). Although there is no standard definition for "health disparities," it can be broadly construed to mean variation in the prevalence and incidence of disease, health status, mortality, and other patient outcomes—as well as differences in health-care access, utilization, and quality—related to specific populations such as ethnicity, gender, geography, or socioeconomic status (SES) (3,4). Given the persistence of disparities across a wide range of health and access measures (5), some consideration of health disparities is now included in almost all clinical and health services studies through mechanisms such as the stratification or adjustment of results by appropriate population and SES characteristics—in addition to being the primary focus of the studies themselves (3,6).

Academic radiologists are involved in this effort, with calls for radiologists to study health disparities and translate results to practice in the hopes of reducing them (7). Most imaging disparities studies focus on access and utilization, although a few examine patient outcomes (8–10). Research on disparities in imaging utilization covers the breadth of current practice with recent studies examining mammography (11–17), positron emission tomography (18,19), computed tomography (CT) and magnetic resonance imaging (20–24), pediatric radiography (25), and aggregate imaging utilization (26).

Similar to most research, the availability of quality data determines both the clinical focus of the study and the nature of the disparities examined. Researchers using data from their own practice or health-care system often have a robust set of patient demographic and insurance information, such as age, gender, ethnicity, and payer type, at their disposal, whereas researchers relying upon external data sources such as medical billing data or other third party datasets are reliant upon the specific data collected. Access to these kinds of data sources is frequently limited because of privacy concerns, availability of information technology resources, or the cost of data acquisition. Even when appropriate clinical data are available, often data on the disparities of interest, such as a specific SES measure or education level, are not. Thus, a key challenge for radiologist researchers undertaking health disparities research is access to appropriate data. However, there are a number of public datasets available for researchers to examine imaging disparities or provide SES data that are linkable to other clinical datasets. This article provides an overview of public datasets that are available for imaging disparities research and familiarizes radiologist researchers with their use.

DATA CONSIDERATIONS

There are many *public datasets*, that is, datasets that are made available to researchers and the public, available for research. However, many public datasets such as Medicare claims data or the Agency for Healthcare Research and Quality's (AHRQ) Healthcare Cost and Utilization Project (HCUP) datasets require considerable fees to purchase. This paper will focus on public datasets that are either freely downloaded or acquired with very low cost. Before exploring specific datasets, it is important for researchers to consider what kinds of data are needed for their study. Ultimately, the appropriate data depend heavily on the study's unit of analysis and the specific research question.

Unit of Analysis

The unit of analysis refers to the level of data required to investigate the research question. This entails determining whether the study question is best explored through data that have been aggregated by some dimension or individual level data. Aggregate data are often compiled by geographic divisions, for example, states, counties, or hospital referral regions. However, aggregate data can also be grouped by nongeographic categories, such as aggregating the data by physician specialty or by place of service, such as emergency department, inpatient, outpatient, and physician office, which also serve as useful units of analysis.

Individual level data are usually found within surveys, administrative claims data, or electronic medical records. Although using individual administrative claims data typically imposes significant acquisition costs, there are free national surveys available. Also, radiologists may have access to patient data through their system's electronic medical records data after successfully submitting their study design to the appropriate institutional review board. However, these data often lack socioeconomic and other population characteristics, such as household income or education level, needed to fully address certain health disparities. Rather than rely upon a specific aggregate or individual dataset, researchers can link population level datasets to their primary data to increase the value of their analyses. For example, researchers interested in whether SES is associated with imaging utilization can link county level or census block level SES data to their individual patient data as a reasonable proxy for many patient SES characteristics.

Study Question

Before embarking on a search for data, researchers must also carefully consider the data requirements of their health disparities research question. Specifically, what is the motivation of the disparities study? Is the disparities component simply one facet of a larger nondisparities research study? For example, is the researcher trying to adjust for potential disparities to ensure these patient characteristics are not influencing the primary clinical research question? Is the disparities component simply an initial descriptive table or a secondary aim of the analysis? If examining specific disparities is the primary research question, is the overall study descriptive in nature or oriented toward formal modeling and hypothesis testing? The majority of imaging-specific research questions fit the mold of "does utilization of modality X for patients with condition Y vary by population factor Z?" Recent examples include whether imaging utilization for acute ischemic stroke patients vary by insurance status (22); whether emergency department utilization of radiography for acute pediatric sinusitis vary by regional and SES (25); and whether positron emission tomography utilization for Medicare patients with cancer vary by SES (19).

Other issues to consider when designing a research question and determining the data needs to appropriately address it include the following:

1. Does the study require cross-sectional (ie, a single time period) or longitudinal (ie, following the same unit of analysis over multiple time periods) data?
2. Is the study question primarily concerned with aggregate trends or individual decision making?
3. Are there broader policy or program changes or specific interventions that may be associated with changes in health disparities? These studies will typically require more involved modeling and advanced analytical techniques.

PUBLICLY AVAILABLE DATASETS FOR HEALTH DISPARITIES RESEARCH

Table 1 provides a list of datasets that can be accessed through the Internet for disparities research. These datasets are described in detail in the following according to the unit of analysis and availability of imaging data.

Aggregate Data Sources: Population Characteristics

There are numerous publicly available data sources that contain aggregate data by different geographic levels on SES, education, income, ethnicity, health status, medical utilization and spending, and other variables valuable for health disparities research:

*   American Community Survey (27): This survey is performed by the United States Census Bureau on a continuing basis and contains geographic aggregates, as well as detailed microdata on a large number of demographic, SES, education, occupational, and housing data.
*   Dartmouth Atlas of Heath Care (28): Provides aggregate Medicare data at the state, hospital referral region, hospital service area, pediatric service area, primary care service area, county, and hospital level on patient demographics, health-care spending, and medical utilization.
*   Kaiser Family Foundation State Health Facts (29): Compiles data from different public and private sources to produce more than 800 health measures at the state level.