HHS Public Access

Author manuscript

J Public Health Manag Pract. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

J Public Health Manag Pract. 2021 ; 27(3): E126–E142. doi:10.1097/PHH.0000000000001079.

Life in Data Sets: Locating and Accessing Data on the Health of Americans Across the Life Span

Jaron Hoani King, BS,
Department of Public Health, Brigham Young University, Provo, Utah

Mary Ann K. Hall, MPH,
Cherokee Nation Assurance, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Richard A. Goodman, MD, MPH,
Department of Family and Preventive Medicine, Emory University School of Medicine, Atlanta, Georgia

Samuel F. Posner, PhD
National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract

Context: The US government manages a large number of data sets, including federally funded data collection activities that examine infectious and chronic conditions, as well as risk and protective factors for adverse health outcomes. Although there currently is no mature, comprehensive metadata repository of existing data sets, US federal agencies are working to develop and make metadata repositories available that will improve discoverability. However, because these repositories are not yet operating at full capacity, researchers must rely on their own knowledge of the field to identify available data sets.

Program or Policy: We sought to identify and consolidate a practical and annotated listing of those data sets.

Implementation and/or Dissemination: Creative use of data resources to address novel questions is an important research skill in a wide range of fields including public health. This report identifies, promotes, and encourages the use of a range of data sources for health, behavior, economic, and policy research efforts across the life span.

Evaluation: We identified and organized 28 federal data sets by the age-group of primary focus; not all groups are mutually exclusive. These data sets collectively represent a rich source of information that can be used to conduct descriptive epidemiologic studies.

Correspondence: Samuel F. Posner, PhD, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta GA 30329 (shp5@cdc.gov).

The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (http://www.JPHMP.com).

Discussion: The data sets identified in this article are not intended to represent an exhaustive list of all available data sets. Rather, we present an introduction/overview of the current federal data collection landscape and some of its largest and most frequently utilized data sets.

Keywords

government information; metadata; open data; public health; research; storage and retrieval

The US government manages a large number of data sets, including federally funded data collection activities that examine infectious and chronic conditions, as well as risk and protective factors for adverse health outcomes. These data sources comprise collections of information across the life span, from birth to death, and can provide a large volume of key data for those who know where to look and how to use the data.[1] Students, educators, researchers, practitioners, and policy makers may be unaware of, or experience difficulty locating, many of these resources. As developers, entrepreneurs, and other groups increasingly combine data across sectors and platforms in creative ways (eg, connecting food safety data to Yelp[2]), it is also desirable to promote use of health data among nontraditional users.[3]

In 2009, the Obama administration issued a memorandum[4] addressed to federal agency executives to increase public use of government data by requiring agencies to "collect and create information in a way that supports public transparency as well as downstream, secondary information dissemination and processing by third parties, thereby making government information accessible, discoverable, and usable." While several notable early efforts have seen some success in open data provision[5–7]—Martin and colleagues define open data as those that are "publicly accessible, available in nonproprietary formats, free of charge, and with unlimited use and distribution rights"[7(pe5)]—interested parties may be unaware of the numerous data sets available for use by students, researchers, policy makers, community groups, developers, and other groups.

Concurrently, primary data collection is becoming more difficult, expensive, and less effective,[8,9] further motivating researchers to utilize available data sets.[10,11] Moreover, students often lack time or resources to collect primary data for projects, while policy makers need answers to time-sensitive questions without making the time and resource investments to collect adequate data, particularly those that might be redundant with previously collected information. The accessibility of data visualization tools and other computational techniques to combine data sources in novel ways allows for additional analysis of existing data for new insights.

Several previous works have compiled guides for individuals interested in federal health data systems.[7,12–15] We sought to identify and consolidate a practical and annotated listing of such data sets that are readily accessible by general users (ie, we did not include Medicare claims or Department of Veterans Affairs data, as these data are generally only available to academics/specialists with data-use agreements in place). This study therefore takes a broad perspective, with identification and summative discussion of 28 data sets intended to address the knowledge gap on federal data sets specific to health behaviors and outcomes across the life span. This life span approach—which we operationalize as studies that examine health

conditions, behaviors, and outcomes, organized by and across age-groups—offers a rational organizing framework and supports the identification and cataloging of federal data sets across the life span.

Methods

We sought to identify all data sets resulting from ongoing data collection activities funded by the US government as of June 2019. Criteria for inclusion were efforts that are funded/sponsored by the federal government; are national in scope (either nationally representative or with nationwide coverage); currently maintain ongoing data collection (including those that have merged with other collection activities); collect individual-level data; and measure health behaviors, risk factors, or outcomes.

The Centers for Disease Control and Prevention (CDC) Surveillance Resource Center's list of interactive data systems (https://www.cdc.gov/surveillancepractice/data.html) provided a foundation for our listing. We conducted a systematic search of data.gov (which includes resources listed on health.data.gov) and cdc.data.gov using terms such as "surveillance," "monitoring," and "health." We reviewed the National Institutes of Health's list of registries (https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries) and searched US Department of Health and Human Services (HHS) agency Web sites (see the Supplemental Digital Content Table, available at http://links.lww.com/JPHMP/A610, for the full list of sites searched). Finally, to augment our search, we conducted informal key informant interviews with epidemiologists at CDC to identify any studies or data sets not ascertained by our searches.

We organized data sets by the age-group of primary focus (not all are mutually exclusive): individuals of all ages; birth (including outcomes of assisted reproductive technology [ART] cycles) through 17 years of age; adolescents (grades 6 [approximately 12 years of age] to [approximately 18 years of age]); adults 18 to 64 years of age; and adults 65 years and older.

Results

The following 28 data sets met our inclusion criteria. We describe each in the following text in alphabetical order by age-group(s) included, and all are listed in Tables 1–5 by age-group. Each table contains the data system title, population of interest, health metric(s) and data collected, and notes on data availability and processes.

Data sources for individuals of all ages

Several data systems provide national or nationwide data that are not age specific (Table 1).

Healthcare Cost and Utilization Project—Healthcare Cost and Utilization Project (HCUP) data sets (https://www.ahrq.gov/research/data/hcup/index.html) provide nationwide data on hospital inpatient, outpatient, and emergency department visits. These systems have large sample sizes, allowing for investigation of rare events. Specific HCUP data sets include the National (Nationwide) Inpatient Sample; the Kids' Inpatient Database; the Nationwide Emergency Department Sample; the Nationwide Readmissions Database; State Inpatient

Databases; State Ambulatory Surgery and Services Databases; and State Emergency Department Databases.

National Ambulatory Medical Care Survey and National Hospital Ambulatory Medical Care Survey—The National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey (NHAMCS) (information and data sets for both are available at https://www.cdc.gov/nchs/ahcd/index.htm) quantify use of American outpatient health care services: who uses services and where; conditions and diagnoses that cause people to seek care; and treatments they receive, including medications prescribed. Both NAMCS and NHAMCS provide reliable statistics on provision and use of ambulatory medical care services.

NAMCS surveys non–federally employed office-based physicians about their patients. NHAMCS surveys emergency departments, outpatient departments, and ambulatory surgery locations about hospital emergency and outpatient ambulatory care visits.

National Health and Nutrition Examination Survey—The National Health and Nutrition Examination Survey (NHANES) (https://www.cdc.gov/nchs/nhanes/index.htm) provides national prevalence estimates for a number of diseases and related risk factors. NHANES is a cross-sectional survey of the civilian, noninstitutionalized US population conducted by the National Center for Health Statistics (NCHS).

NHANES combines a household interview with physical examinations and blood collection at a mobile examination center. NHANES data are the basis for national standards for height, weight, and blood pressure.

The survey became continuous in 1999 and collects data annually on Americans of all ages. Each year, NHANES surveys 5000 individuals.

National Health Interview Survey—The National Health Interview Survey (NHIS) (https://www.cdc.gov/nchs/nhis/index.htm), a cross-sectional household survey conducted continuously, is a primary source of information on the health of the civilian, noninstitutionalized US population. Data are collected through personal household interviews. The survey consists of several components. The Household component collects limited demographic information on all individuals living in a household. The Family component collects additional demographic information on each member of each family in the house and information on health status and limitations, injuries, health care access and utilization, health insurance, and income and assets. NHIS data are used to identify Americans' health problems, determine barriers to accessing health care, evaluate health programs, study health-related disparities, monitor progress toward national health objectives, and monitor progress toward national well-being indicators. Supplements are used to respond to emergent public health data needs.

National Hospital Care Survey—The National Hospital Care Survey (NHCS) (https://www.cdc.gov/nchs/nhcs/index.htm) replaced the National Hospital Discharge Survey (NHDS), performed annually from 1965 to 2010. NHCS describes national patterns of

health care delivery in hospital-based settings, including inpatient departments, emergency departments, and outpatient departments, including hospital-based ambulatory surgery. NHCS collects patient-level identifiers that allow episodes of care to be linked between different settings, as well as to outside databases such as the National Death Index. Because of low response rates, NHCS data are not yet nationally representative.

National Longitudinal Mortality Study—The National Longitudinal Mortality Study (NLMS) (https://www.census.gov/nlms) is unique in that it triangulates data collected through several US Census mechanisms—Annual Social and Economic Supplements from March 1973 to March 2011, Current Population Surveys for February 1978, April 1980, August 1980, December 1980, and September 1985, and one 1980 Census cohort—with NCHS death certificate data to study the association between demographic and socioeconomic factors and death rates in the United States. Because of the study design, NLMS data provide a single baseline measurement of subjects for association with mortality data.

Nationally Notifiable Disease Surveillance System—The Nationally Notifiable Disease Surveillance System (NNDSS) (https://wwwn.cdc.gov/nndss) collects information on 120 nationally notifiable infectious diseases and conditions from local, state, territorial, federal, and international public health entities. NNDSS collects basic information to aid in outbreak identification and investigation and provides basic information on the number of cases each year. For some pathogens, the system collects additional epidemiologic data. These data are summarized and published in the *Morbidity and Mortality Weekly Report* (*MMWR*; https://www.cdc.gov/mmwr/index.html).

In addition to conditions tracked by NNDSS, many—if not most—diseases affecting Americans are represented in condition-specific data sets (eg, the Supplemental Legionnaire's Disease Surveillance System, the Chronic Kidney Disease Surveillance System, and the National HIV Surveillance System).

National Program of Cancer Registries—The National Program of Cancer Registries (NPCR) (https://www.cdc.gov/cancer/npcr/index.htm) collects data on occurrence of cancer; type, extent, and location of the cancer; and type of initial treatment. Through NPCR, CDC supports cancer registries in 46 states, the District of Columbia, Puerto Rico, the US Pacific Islands, and the US Virgin Islands—in all, 97% of the US population falls within jurisdictions participating in NPCR. Data on patient demographics and tumor/cancer types are reported to the central registry by medical facilities.

National Vital Statistics System—The National Vital Statistics System (NVSS) (https://www.cdc.gov/nchs/nvss/index.htm) represents the collection of records regarding US births, deaths, marriages, divorces, and fetal deaths. Vital statistics are recorded within localities in all states, as well as a number of major cities and territories, and then compiled in the national system.

Surveillance, Epidemiology, and End Results Program—The National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program (https://

seer.cancer.gov/) collects data on demographics, diagnosis, cancer markers, treatment, and survival of patients with cancer in 20 US jurisdictions; together, these jurisdictions cover 28% of the US population but are demographically representative of the US population. SEER data include incidence and population data associated with age, sex, race, year of diagnosis, and geographic area; data sets are released each spring for the previous year.

US Census: American Community Survey—The US Census Bureau conducts a census of the entire US population every 10 years, with the goal of counting every individual living in the nation. The Census collects data on age, race, sex, and duration of residence. The American Community Survey (ACS) (https://www.census.gov/programs-surveys/acs.html) is an ongoing supplement to the Census collecting in-depth information about the US population. ACS includes questions on conditions associated with health and wellness, such as marital status/marital history, health insurance coverage, fertility, employment status, educational attainment, and disability status.

Gestation, early life, and childhood life stages

These data sets contain information of those in early life and childhood, from birth through 17 years of age, as well as outcomes of ART cycles (Table 2).

National Assisted Reproductive Technology Surveillance System—Per the Fertility Clinic Success Rate and Certification Act of 1992, all ART programs must report pregnancy success rates. This is operationalized as reporting of each ART cycle performed (ie, women who undergo multiple ART cycles in a year will be present in the data multiple times and will not be linked). The National Assisted Reproductive Technology Surveillance System (NASS) is a Web-based reporting system. NASS data are used for development of federally mandated clinic success rate reporting, as well as surveillance and research reports.

National Immunization Survey—The National Immunization Survey (NIS) (https://www.cdc.gov/vaccines/imz-managers/nis/index.html) collects information about immunizations in the first 35 months of life, as well as during childhood. NIS collects data in 2 ways: via randomly selected, dual-frame landline and cellular telephone surveys of parents and guardians, followed by a mail survey sent to health care providers identified by telephone survey participants to confirm immunization information.

NIS-Teen surveys parents, guardians, and health care providers to determine vaccine coverage for US teens 13 to 17 years of age and immunization coverage rates for vaccines recommended between 11 and 17 years of age. NIS-Child Influenza Model (CIM) is conducted each year from October through June with parents and guardians of children 6 to 18 months and 3 to 12 years of age to determine whether their children received an influenza vaccine.

National Survey of Children with Special Healthcare Needs—The National Survey of Children with Special Healthcare Needs (NSCSHN) (http://www.childhealthdata.org/learn/NS-CSHCN) provides national and state-level estimates of the general characteristics of children with special health care needs, following the definition used by the Health Resources and Services Administration's Maternal and Child Health Bureau, first proposed

by McPherson and colleagues: "…those who have or are at increased risk for a chronic physical, developmental, behavioral, or emotional condition and who also require health and related services of a type or amount beyond that required by children generally."[16] NSCSHN is currently being integrated into the National Survey of Children's Health (NSCH; see the following section).

National Survey of Children's Health—NSCH (http://www.childhealthdata.org/learn) produces national and state-level data on the health and well-being of noninstitutionalized US children 0 to 17 years of age.

Pregnancy Risk Assessment and Monitoring System—In nearly all states, the Pregnancy Risk Assessment and Monitoring System (PRAMS) (https://www.cdc.gov/prams/index.htm) surveys a random sample of mothers from the birth certificate registry to collect information about maternal health before, during, and shortly after pregnancy, as well as early infant health outcomes, and health care utilization during this time frame.

The survey has a set of core questions; states can choose to add additional items. This system is used for a wide range of programmatic activities including Title V reporting measures, surveillance, and research. In some states, PRAMS data are linked to NASS data.

Adolescent life stage

These surveys collect information on the health and health behaviors of adolescents (those approximately aged 12–17 years or grades 6–12 in school) (Table 3).

Monitoring the Future survey—Funded by the National Institutes of Health's National Institute on Drug Abuse, the Monitoring the Future (MTF) survey is administered each year to a nationally representative sample of US public and private secondary school students in grades 8, 10, and 12. Participants are asked about use of licit and illicit drugs, including alcohol, tobacco, marijuana, opiates, cocaine, inhalants, and steroids.

Youth Risk Behavior Surveillance System—The Youth Risk Behavior Surveillance System (YRBSS) (https://www.cdc.gov/healthyyouth/data/yrbs/index.htm) collects information about risk behaviors via school-based surveys of US adolescents in grades 9 to 12 at both public and private high schools. YRBSS is conducted by CDC and state, territorial, and local education and health agencies and tribal governments, and the set of data that comprise YRBSS results arise from a collection of subsurveys conducted at the national, state, territorial, tribal government, and school levels. Middle school students are also surveyed in some states. YRBSS is the most extensive federally funded risk behavior surveillance system among nonadults.

Youth Tobacco Survey—The Youth Tobacco Survey (YTS) (https://www.cdc.gov/tobacco/data_statistics/surveys/yts/index.htm) is a state-based surveillance system that focuses specifically on tobacco use among adolescents in grades 6 to 12. This survey measures knowledge and attitudes regarding tobacco use. YTS' primary focus is to aid state agencies in preventing initial use of tobacco among adolescents. The program contains technical support for states to help with analysis and dissemination of online data.

Adult life stages

The majority of data systems in this category, which gather data from adults aged 18 to 64 years, utilize complex sampling designs to be representative of the population at either the jurisdiction or national level (Table 4).

Behavioral Risk Factor Surveillance System—The Behavioral Risk Factor Surveillance System (BRFSS) (https://www.cdc.gov/brfss/questionnaires/index.htm) is the largest continuously conducted health survey in the world, involving approximately 400 000 American adults interviewed by phone each year from all 50 states. Americans are surveyed on demographic information, health-related risk behaviors (such as seatbelt and tobacco use), chronic health conditions, and use of preventive services. Like PRAMS, BRFSS contains a core set of questions to which states can add optional modules; states can also add individual questions.

Medical Expenditure Panel Survey—For the past 2 decades, the Medical Expenditure Panel Survey (MEPS) (https://meps.ahrq.gov/mepsweb/index.jsp) has played a major role in addressing the cost of health care in the United States by evaluating insurance statistics side by side with out-of-pocket health care expenses. Administered by the Agency for Healthcare Research and Quality (AHRQ), MEPS helps policy makers and researchers understand the landscape of American health care.

Unlike traditional panel studies that follow a single cohort over time, the panel for the Household Component of MEPS is a nationally representative subsample drawn from the set of randomly selected individuals who participated in the NHIS (see earlier). Respondents are surveyed 5 times over a 2-year period. A separate survey component, the Insurance Component, surveys employers directly.

National Adult Tobacco Survey—As a nationally stratified, random-digit dialed phone survey of all noninstitutionalized Americans, the National Adult Tobacco Survey (NATS) (https://www.cdc.gov/tobacco/data_statistics/surveys/nats/index.htm) provides estimates of adult tobacco use by subgroups based on gender, age, and race/ethnicity. The main goals of the survey are to provide current and reliable statistics about tobacco use. To assess adult tobacco use, CDC surveys adults about their tobacco use, quit attempts, and success rates.

NATS has been conducted annually since 2009. Although the survey continues to take place, the most current data set available represents surveys from 2013 to 2014. Most states maintain independent statistics, with similar metrics that can be obtained from individual state health departments.

National Survey of Family Growth—To maintain accurate national statistics about family life including marriage, pregnancy, contraception use, and divorce, the National Survey of Family Growth (NSFG) (https://www.cdc.gov/nchs/nsfg/index.htm) is conducted through in-person interviews. NSFG participants include men and women from 15 to 49 years of age (15–44 years prior to September 2015), sampled to be representative of the US population. Most survey questions are administered in an in-person interview; the most sensitive questions are self-administered. Topics include marriage; divorce; cohabitation;

women's pregnancy and birth history; contraceptive use; sexual behaviors that may carry risk of HIV or sexually transmitted infections, including same-sex behavior; and sexual orientation and attraction.

National Survey on Drug Use and Health—The National Survey on Drug Use and Health (NSDUH) (https://nsduhweb.rti.org/respweb/homepage.cfm) estimates the prevalence of substance abuse and mental health disorders in the United States. Participants are asked about tobacco, alcohol, and illicit drugs; this is further broken down into subcategories (eg, opioids, types of tobacco used). Participants are also asked about substance use disorders and treatment in the past year. NSDUH also examines a number of mental health and behavioral health indicators, including major depressive episodes in the past year; mental illness among adults in the past year; mental health service use in the past year; and others.

Interviewers conduct the survey in the homes of approximately 70 000 participants 12 years or older each year. The study provides statistics on risk behaviors involving substance use; NSDUH also helps policy makers understand the prevalence of mental health disorders.

Older adult life stages

These surveys focus primarily (though not exclusively) on individuals 65 years and older (Table 5).

Health and Retirement Survey—The Health and Retirement Survey (HRS) (http://hrsonline.isr.umich.edu) is a panel survey of a succession of individuals older than 50 years, organized into nationally representative age-group cohorts. The panels consist of 6-year birth cohorts; a new cohort is introduced every 6 years as it ages into the survey's target range. HRS follows individuals and their spouses from the time they enter the survey until their death. Because interviews are conducted over time with the same set of individuals, HRS allows for longitudinal study of health and health care service use.

HRS, launched in 1992, is supported by the National Institute on Aging (part of the National Institutes of Health) and the Social Security Administration. In addition to interview data, HRS collects data on biomarkers, genetic information, and psychosocial variables including perceived well-being, social support, and self-related beliefs.

Medicare Current Beneficiary Survey—The Centers for Medicare & Medicaid Services (CMS) conducts the Medicare Current Beneficiary Survey (MBCS) (https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/index.html) to examine health outcomes, health care expenditures, and sources of payment for individuals of all ages who receive Medicare: individuals 65 years and older (and a small number of younger individuals with disabilities or end-stage renal disease). Survey respondents are a nationally representative sample of Medicare beneficiaries. MCBS has been continuously administered for more than 25 years.

The MCBS Survey File contains data on beneficiary demographics, household characteristics, access to care, satisfaction with care, usual source of care, health insurance

timeline (shows types of insurances, the coverage eligibility, and what is covered), health status and functioning, and other topical survey sections such as medical conditions, health behaviors, preventive services, interview characteristics, beneficiary knowledge of the Medicare program, residence timeline, facility characteristics, and beneficiary income and assets. The MCBS Survey File is complemented by the MCBS Cost Supplement File, which links cost and utilization data for survey beneficiaries with Medicare claims data. To capture the full range of medical services delivered to Medicare beneficiaries, the MCBS Cost Supplement File collects data from participants on other health care utilization and costs that are not covered by Medicare.

National Health and Aging Trends Study—The National Health and Aging Trends Study (NHATS) (https://www.nhats.org/scripts/aboutNHATS.htm) surveys a nationally representative sample of Medicare beneficiaries 65 years and older. Interviewers conduct in-person interviews regarding well-being as defined by distinct areas ranging from disability and daily living assistance received to social interaction and involvement. The sample is periodically cycled, which allows researchers to study the population as a whole or to track individual aging trends among participants.

Discussion

Our article highlights the importance of ongoing investments in data catalogs to ensure that information about data sets and their potential for use—such as that described in this article—is more readily accessible and usable to a broader audience including researchers from other disciplines, community groups, developers, entrepreneurs, and other potential users.

Organizing and accessing data are an ongoing challenge. A primary limitation of our study and compilation—the likelihood that there are data sources that were not detected or that are redundant with those listed here and elsewhere—reflects the absence of an easily accessible, standardized catalog of federal surveillance activities with similarly categorized indicators. Several US government initiatives (eg, https://www.data.gov, https://www.healthdata.gov, and https://clinicaltrials.gov) are intended to provide researchers and the public with access to data sources such as those identified here, as well as studies on specific topics and clinical trials. Examples of additional attempts to meet these needs are the ongoing efforts by HHS[7,17–19] to make more efficient and creative use of data, including the ReImagine HHS initiative and HHS' Data Initiative, following the previously noted guidance from the Obama administration.[3,4] In addition to the discrete, data system–specific sources we have presented, there are ongoing efforts to improve transparency and access to federal open data sources more broadly. A broader understanding of currently funded federal health and public health surveillance efforts can decrease unnecessary data collection costs by allowing researchers and policy makers to more fully utilize extant governmental efforts. In addition, efficiency of federal health surveillance data collection will likely improve as redundancies and duplicated data are identified and remediated.

The data sets identified in this article are not intended to represent an exhaustive list of all available data sets but rather a set of some of the US government's largest and most frequently accessed data sets for increased use by students, researchers, and others. As

noted, our study concentrated on identifying and describing federal data collection efforts that are consistent with our selection criteria. However, our approach did not ascertain and include other data sources, systems, or compilations that are not national in scope or for other reasons. Examples of such data systems that were not included are several infectious disease surveillance systems that collect information on outbreaks or specific diseases and data systems that collect information from a limited number of jurisdictions on birth defects, violent deaths, and environmental hazards.

Promotion of open data may be of benefit to data collectors who plan for and invest in open sharing of the data they collect or maintain. For example, Martin and Begany[20,21] found that early innovators acting in response to release of open government data cited a number of anticipated benefits, including more efficient public health operations; improved data quality, timeliness, and usefulness; and promotion of government transparency and fairness as benefits of data sharing. Even where data cannot be open in the technical sense (eg, person-level data with identifiable information), making information about the data more accessible and open versions when feasible may enhance use among a larger community of users.

Innovative use of data resources to address novel questions is an important need for research and policy in public health, as well as many other fields. This report has identified and characterized a multitude of data sources for use in health, behavior, economic, and policy research across the life span. These data sets collectively represent a rich source of data for descriptive epidemiologic studies, informing policy making efforts, and other purposes. Given the concurrent trends of provision of open data and declining survey response rates, we suggest that increasingly it will be important to (1) encourage data collectors and repositors to consider open data from planning through the analysis and dissemination stages of data collection, and (2) promote open data resources and encourage interested parties to request access to data as appropriate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr Erika Martin, PhD, MPH, of the Rockefeller College of Public Affairs and Policy at the University at Albany, State University of New York, for her thoughtful and generous review.

The findings and conclusions of this report do not reflect the official views of the Centers for Disease Control and Prevention.

References

1. Bryan J, Kim J, Shi Q. Identifying and using secondary datasets to answer policy questions related to school-based counseling: a step-by-step guide. In: Carey JC, Harris B, Lee SM, Aluede O, eds. International Handbook for Policy Research on School-Based Counseling. Cham, Switzerland: Springer International Publishing; 2017:153–181.
2. Code for America. San Francisco promotes its restaurant inspection data on Yelp to improve food safety. https://www.codeforamerica.org/featured-stories/san-francisco-promotesits-restaurant-inspection-data-on-yelp-to-improve-food-safety. Accessed September 3, 2019.

3. Wold C. In Plain Sight: Is Open Data Improving Our Health? Oakland, CA: California Health Care Foundation; 2015. https://www.chcf.org/wp-content/uploads/2017/12/PDF-InPlainSightOpenData.pdf. Accessed September 3, 2019.
4. Obama B. Transparency and Open Government. Washington, DC: White House; 2009. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government. Accessed September 3, 2019.
5. Martin EG, Helbig N, Birkhead GS. Opening health data: what do researchers want? Early experiences with New York's Open Health data platform. J Public Health Manag Pract. 2015;21(5):E1–E7.
6. Martin EG, Shah NR, Birkhead GS. Unlocking the power of open health data: a checklist to improve value and promote use. J Public Health Manag Pract. 2018;24(1):81–84. [PubMed: 28257411]
7. Martin EG, Law J, Ran W, Helbig N, Birkhead GS. Evaluating the quality and usability of open data for public health research: a systematic review of data offerings on 3 open data platforms. J Public Health Manag Pract. 2017;23(4):e5–e13. [PubMed: 26910872]
8. Czajka JL, Beyler A. Declining Response Rates in Federal Surveys: Trends and Implications(background paper). Washington, DC: Mathematica Policy Research; 2016.
9. Kohut A, Keeter S, Doherty C, Dimock M, Christian L. Assessing the Representativeness of Public Opinion Surveys. Washington, DC: Pew Research Center; 2012.
10. Lohr SL, Raghunathan TE. Combining survey data with other data sources. Stat Sci. 2017;32(2):293–312.
11. Miller PV. Is there a future for surveys? Public Opin Q. 2017;81(S1):205–212.
12. Thacker SB, Qualters JR, Lee LM; Centers for Disease Control and Prevention. Public health surveillance in the United States: evolution and challenges. MMWR Surveill Summ. 2012;61(suppl):3–9.
13. Goodman RA, Posner SF, Huang ES, Parekh AK, Koh HK. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. Prev Chronic Dis. 2013;10:E66. [PubMed: 23618546]
14. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Health, United States, 2016: With Chartbook on Long-term Trends in Health. Hyattsville, MD: National Center for Health Statistics; 2017. DHHS Publication No. 2017–1232.
15. Blewett LA, Call KT, Turner J, Hest R. Data resources for conducting health services and policy research. Annu Rev Public Health. 2018;39:437–452. [PubMed: 29272166]
16. McPherson M, Arango P, Fox H, et al. A new definition of children with special health care needs. Pediatrics. 1998;102(1 Pt 1):137–140. [PubMed: 9714637]
17. National Academies of Sciences, Engineering, and Medicine. Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy. Washington, DC: National Academies Press; 2017.
18. National Academies of Sciences, Engineering, and Medicine. Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps. Washington, DC: National Academies Press; 2018.
19. US Department of Health and Human Services. The Data Initiative. Washington, DC: US Department of Health and Human Services; 2018. https://www.hhs.gov/cto/initiatives/data-initiative/index.html. Accessed September 3, 2019.
20. Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. J Am Med Inform Assoc. 2017;24(2):345–351. [PubMed: 27497796]
21. Begany GM, Martin EG. An open health data engagement ecosystem model: are facilitators the key to open data success? In: Erdelez S, Agarwal NK, eds. Proceedings of the Association for Information Science and Technology. Hoboken, NJ: Wiley; 2017: 621–623.

Implications for Policy & Practice

- Innovative use of data resources to address novel questions is an important need for research and policy in public health and many other fields. This report identifies a range of data sources for health, behavior, economic, and policy research efforts across the life span. Provision and use of such data should be promoted and encouraged by both data collectors and those who seek to utilize the data.
- Use of existing data resources is a viable option for researchers in an environment where data collection is increasingly expensive and difficult. In both public and private sectors, use of existing data sets is becoming much more common in describing and studying health and behavior. Current efforts across HHS and other parts of the US government are directed at more efficient and creative use of data, including the ReImagine HHS initiative and HHS' Data Initiative.
- A broader understanding of currently funded federal health surveillance efforts can help decrease unnecessary data collection costs by more fully utilizing current governmental efforts. In addition, efficiency of federal health surveillance data collection will likely improve as redundancies and duplicated data are identified and remediated.
- Promotion of open data and accessible data catalogs targeting users at different levels of expertise may be of benefit to data collectors who plan for and invest in open sharing of the data they collect or maintain.

TABLE 1

Data Sets Focused on Americans of All Ages

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
Healthcare Cost and Utilization Project (HCUP)	All	Different HCUP data sets have different designs and sampling procedures, detailed at https://www.hcup-us.ahrq.gov/tech_assist/sampledesign/508_compliance/index508_2018.jsp	Hospital discharge	Hospital discharge data: all listed diagnoses and procedures, as well as patient demographics and other information (eg, discharge, payers)	Some HCUP data sets are only available for purchase. However, individuals can generate health care statistics using HCUP data at the free online HCUPnet query system (https://hcupnet.ahrq.gov/#setup), and many data analyses are available at HCUP Fast Stats (https://www.hcup-us.ahrq.gov/faststats/landing.jsp).
National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS)	All people in the United States who seek treatment	Multistage probability sample design (https://aspe.hhs.gov/report/data-health-and-wellbeing-american-indiansalaska-natives-and-othernative-americans-datacatalog/national-ambulatorymedical-care-survey-namcs)	Patient visit	Outpatient health care usage. NAMCS includes information about the physicians' practice type; patients' age, sex, race, and ethnicity; and patients' reason for visit, diagnosis, services ordered or provided, and treatments. NHAMCS collects data on patients' age, sex, race, and ethnicity; and patients' reason for visit, diagnosis, services ordered or provided, and treatments	NAMCS and NHAMCS data are readily available: NAMCS public-use files for 1993–2015 and NHAMCS public-use files for 1992–2015 can be downloaded directly from CDC'sWeb site, and there is a suite of associated research tools available at https://www.cdc.gov/nchs/ahcd/ahcd_research_tools.htm. For reasons of confidentiality, some NAMCS and NHAMCS data are not released in public-use format; instructions for proposals to the NCHS Research Data Centers for use of such data can be found at http://www.cdc.gov/rdc.
National Health and Nutrition Examination Survey (NHANES)	Civilian, noninstitutionalized US population	Multistage, national area probability survey	Individuals	Health conditions, including diseases, medical conditions, and health indicators including anemia; cardiovascular disease; diabetes; environmental exposures; eye diseases; hearing loss; infectious diseases; kidney disease; nutrition; obesity; oral health; osteoporosis; physical fitness and physical functioning; reproductive history and sexual behavior; respiratory diseases (asthma, chronic bronchitis, emphysema); sexually transmitted diseases; and vision	NHANES collects data on a wide range of health conditions, including diseases, medical conditions, and health indicators. NHANES data sets, questionnaires, and related documentation are available at the survey Web site https://wwwn.cdc.gov/nchs/nhanes/Default.aspx.
National Health Interview Survey (NHIS)	Civilian, noninstitutionalized US population	Area probability design (https://www.cdc.gov/nchs/nhis/about_nhis.htm#sample_design)	Households/individuals	Physical and mental health status; chronic conditions, including asthma and diabetes; access to and use of health care services; health insurance coverage and type of coverage; health-related behaviors, including smoking and alcohol use, and physical activity; measures of	Researchers, students, and others can download reports, data files, and associated documentation at http://www.cdc.gov/nchs/nhis.htm. For reasons of confidentiality, some data are not released in public-use format; instructions for proposals to the NCHS Research Data Centers for use of such data can be found at http://www.cdc.gov/rdc.

J Public Health Manag Pract. Author manuscript; available in PMC 2022 January 01.

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
				functioning and activity limitations; and immunizations	
National Hospital Care Survey (NHCS)	All people in the United States who seek treatment	Census of selected hospitals (https://www.cdc.gov/asthma/survey/NHCS_2018_508.pdf)	Discharge record	Inpatient, emergency, and outpatient department statistics from hospitals	NHDS data (ie, data collected through 2010) are available for free public use and download online (https://www.cdc.gov/nchs/nhds/nhds_questionnaires.htm). For reasons of confidentiality, NHCS data are not released in a public-use format; instructions for proposals to the NCHS Research Data Centers for use of NHCS data can be found at http://www.cdc.gov/rdc.
National Longitudinal Mortality Study (NLMS)	All	None	Individual	Association of demographic and socioeconomic data—age; region, state, county, and date of birth; sex; citizenship status; race/ethnicity; marital status; occupation/industry; veteran status; migration status; family income/household poverty level; household characteristics; and tobacco use (for some cohorts)—with NCHS death data (state, date, place, and the underlying cause of death)	Data are restricted; access to a public-use data file requires an approval process via the Census Bureau's Research Data Centers (https://www.census.gov/about/adrm/fsrdc/locations.html). Interested researchers can also apply for on-site access to the data (with restricted variables) through the Census Bureau NLMS Principal Investigator.
Nationally Notifiable Disease Surveillance System (NNDSS)	All	None	Disease event	Incidence of reportable conditions; disease-specific (multiple)	Weekly and annual disease reporting tables, as well as lists of resources for interpreting the data, are available for infectious diseases (https://wwwn.cdc.gov/nndss/infectious-tables.html), and links to data on notifiable noninfectious diseases (https://wwwn.cdc.gov/nndss/noninfectious.html) are available.
National Program of Cancer Registries (NPCR)	All	None	Cancer diagnosis	Cancer incidence and prevalence: occurrence of cancer; type, extent, and location of the cancer; and type of initial treatment	A number of tools are available for analysis (https://www.cdc.gov/cancer/npcr/tools.htm). Tools that do not require undergoing a researcher approval process to access include CDC's Interactive Cancer Atlas, State Cancer Facts, and overall United States Cancer Statistics, which provide statistics for cancer cases and deaths. Approved researchers can access restricted data sets with approval, such as NCHS cancer data.
National Vital Statistics System (NVSS)	All	None	Individual	Births, deaths, marriages, divorces, and fetal deaths	Free digital vital statistics files can be found at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm. Most microdata files for all deaths and births in the United States dating back to 1968 can be downloaded by the public. Data files for infant deaths linked to births (period and birth cohort–linked birth and infant death files) are also available. Physical copies of vital records are maintained locally; contact information for obtaining vital records can be found at https://www.cdc.gov/nchs/w2w/index.htm. State marriage and divorce tables can be found at https://www.cdc.gov/nchs/nvss/marriage-divorce.htm.
Surveillance, Epidemiology, and End Results (SEER) Program	All individuals residing in 20 US jurisdictions (sample is demographically	None	Cancer diagnosis	Cancer incidence and prevalence. Cancer patient demographics, diagnosis, cancer markers, treatment, and survival. Incidence and	Direct access for researchers to SEER data via the SEER*Stat statistical software requires a signed data-use agreement. However, there are multiple publicly available resources for interpreting and understanding SEER data. Some examples include SEER*Explorer (https://seer.cancer.gov/explorer/),

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
	representative of the US population)			population data by age, sex, race, year of diagnosis, and geographic area	which provides visual depictions (such as tables and graphs) of cancer data; State Cancer Profiles (https://statecancerprofiles.cancer.gov/), which include graphs and maps displaying cancer trends at the county, state, and national levels; and Cancer Query Systems (https://seer.cancer.gov/canques/), which produce and display reports of cancer statistics using information from SEER and other databases.
US Census American Community Survey (ACS)	All	Multistage sampling design	Household and individual	Conditions associated with health and wellness, such as marital status/marital history, health insurance coverage, fertility, employment status, educational attainment, and disability status, as well as a number of others	ACS data tables on Comparison and Selected Population Profiles, and Subject, Detailed, Ranking, and Geographic Comparison Tables are available at the ACS Data Tables and Tools site at https://www.census.gov/acs/www/data/data-tables-and-tools/index.php. Census and ACS data are readily searchable by geographic unit using the American FactFinder database at https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.

Abbreviations: CDC, Centers for Disease Control and Prevention; NCHS, National Center for Health Statistics; NHDS, National Hospital Discharge Survey.

TABLE 2

Data Sets Focused on Americans in Early Life and Childhood, As Well As Outcomes of ART Procedures

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
National Assisted Reproductive Technology Surveillance System (NASS)	Adult women; pregnant women; embryos/ fetuses; infants	All ART clinics are required to report to NASS. It is estimated that more than 95% of clinics report (https://www.cdc.gov/art/nass/index.html)	Cycle	Outcomes from ART. Patient demographics, patient obstetrical and medical history, parental infertility diagnosis, clinical parameters of the ART procedure, and resultant pregnancies and births	Because NASS data contain a significant amount of indirect, potential personally identifying information—patient characteristics, ART procedure information, and treatment outcomes—CDC only provides limited access to researchers under a certain set of circumstances or by permission following a proposal process (https://www.cdc.gov/art/nass/accessdata.html). However, key findings are available at https://www.cdc.gov/art/key-findings/index.html.
National Immunization Survey (NIS)	All (mostly children aged 0–17 y)	Random-digit dial telephone sample (https://www.cdc.gov/nchs/nis/reports.htm#methodology)	Individual	Immunizations in the first 35 mo of life, as well as during childhood NIS-Teen: vaccine coverage for US teens aged 13–17 y and immunization coverage rates for vaccines recommended between the ages of 11 and 17 y NIS-Child Influenza Model (CIM): influenza vaccination rates among children aged 6–18 mo and 3–12 y	Data sets and information specific to NIS-Teen are available at https://www.cdc.gov/vaccines/imz-managers/coverage/nis/teen/index.html. NIS-CIM influenza vaccination questions are also included in NIS and NIS-Teen. NIS public-use data files, data user guides, data sets, and questionnaires are available at https://www.cdc.gov/vaccines/imz-managers/nis/datatables.html.
National Survey of Children with Special Healthcare Needs (NSCSHN)	Children (aged 0–17 y)	Random-digit dial telephone sample (https://www.childhealthdata.org/docs/drc/2011–12-nschsampling-and-administration.pdf)	Individual	Health status among children identified with special needs; overall health and health status of CSHCN, including establishment of a medical home, adequate health insurance, access to needed services, and adequate care coordination	NSCSHN data are available in a number of formats, including an interactive query for browsing, a data query content map, and data snapshots and maps by health topic and state/region; see http://www.childhealthdata.org/learn/NSCSHCN/data.
National Survey of Children's Health (NSCH)	Children (aged 0–17 y)	Randomly selected addresses from noninstitutionalized households across the United States	Individual	Child and family demographics; children's physical and mental health status, including health conditions and functional difficulties; health insurance status, adequacy and type of coverage; access to and use of health care services; and several other topics	NSCH data are available at the Data Resource Center for Child and Adolescent Health (www.nschdata.org), and individuals can use the NSCH interactive data query (www.nschdata.org/browse/survey) to access data.
Pregnancy Risk Assessment and Monitoring System (PRAMS)	Recently postpartum women	Monthly random sample of women who have had a recent live birth is drawn from the state's birth certificate file	Individual	Maternal attitudes and feelings about the most recent pregnancy; content and source of prenatal care; maternal alcohol and tobacco consumption; physical abuse before and during pregnancy; pregnancy-related morbidity; infant health care; contraceptive use; mother's knowledge of pregnancy-related health issues, such as adverse effects of tobacco and alcohol, benefits of folic acid consumption, and risks of HIV infection	At the PRAMStat Data Portal (https://www.cdc.gov/prams/work-directly-PRAMStat.html), users can filter and export data (a variety of file formats are available); create custom visualizations; and view associated metadata.

Abbreviations: ART, assisted reproductive technology; CDC, Centers for Disease Control and Prevention; CSHCN, children with special health care needs; NCHS, National Center for Health Statistics.

TABLE 3

Sources of Data Focused on American Adolescents

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
Monitoring the Future (MTF) survey	US public and private secondary school students in grades 8, 10, and 12	Multistage random sampling (http://www.monitoringthefuture.org/purpose.html#Sampling)	Individual	Use of licit and illicit drugs; perceived risk, personal disapproval, and perceived availability of substances. Demographic data are collected on race/ethnicity; gender; college plans; geographic location, including region of the country and population density of the area of residence (eg, urban, rural); and socioeconomic status	MTF data tables can be accessed at https://www.drugabuse.gov/related-topics/trends-statistics/monitoring-future. Some publicly available MTF microdata from cross-sectional, in-school surveys can be obtained through the Find DataWeb page (https://www.icpsr.umich.edu/icpsrweb/NAHDAP/data/index.jsp) of the National Addiction & HIV Data Archive Program (https://www.icpsr.umich.edu/icpsrweb/NAHDAP/index.jsp).
Youth Risk Behavior Surveillance System (YRBSS)	School children (grades 9–12), middle school students (grades 6–8) included in some subsurveys in some years	Two-stage cluster sample (https://www.cdc.gov/healthyyouth/data/yrbs/faq.htm#methodology)	Individual	Risk behaviors: behaviors that contribute to unintentional injuries and violence; sexual behaviors related to unintended pregnancy and sexually transmitted diseases, including HIV infection; alcohol and other drug use; tobacco use; unhealthy dietary behaviors; and inadequate physical activity	Researchers and others can download national-level YRBSS data directly from the surveyWeb site; however, to access state, district, territory, or tribal government data files, individuals must complete an YRBSS Data Request Form (https://www.cdc.gov/healthyyouth/data/yrbs/contact.htm).
Youth Tobacco Survey (YTS)	School children (grades 6–12)	Two-stage random sample (https://www.cdc.gov/tobacco/data_statistics/surveys/yts/index.htm)	Individual	Knowledge and attitudes regarding tobacco use; exposure to media and advertising; information on enforcement of minors' access regulations and laws; presence of tobacco programs in school curricula; cessation attempts and successes; secondhand smoke exposure; and prevalence of other tobacco products	Weekly surveillance data are available in *MMWR*. YTS data and documentation are available on the survey Web site at https://www.cdc.gov/tobacco/data_statistics/surveys/yts/index.htm.

Abbreviation: MMWR, Morbidity and Mortality Weekly Report.

TABLE 4

Data Sets Focused on American Adults Aged 18 to 64 Years

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
Behavioral Risk Factor Surveillance System (BRFSS)	Adults (including those 65 y and older)	Landline: Disproportionate stratified sample; Cellular: Random sample of confirmed area code/prefix combinations (https://www.cdc.gov/brfss/about/brfss_faq.htm)	Individual	Demographic information, health-related risk behaviors (such as seatbelt and tobacco use), chronic health conditions, and use of preventive services	Because of the unique depth and breadth of the data collected, BRFSS has a number of data tools to assist users, in addition to its Data and Documentation page at https://www.cdc.gov/brfss/data_documentation/index.htm. These include the Prevalence and Trends Data Tools; theWeb Enables Analysis Tool; the Selected Metropolitan/Micropolitan Area Risk Trends project; and the Chronic Disease Indicators Tool.
Medical Expenditure Panel Survey (MEPS)	Adults	Nationally representative sample for the Medical Expenditure Panel Survey Household Component (MEPS-HC) drawn from among households responding to the previous year's National Health Interview Survey	Household and individual	Demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment	Most MEPS household public-use data files, documentation, and codebooks are available for download; some data are restricted, but researchers can request access via the AHRQ's Data Center at https://meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp.
National Adult Tobacco Survey (NATS)	Adults (including those 65 y and older)	Dual-frame random-digit dial sample (https://www.cdc.gov/tobacco/data_statistics/surveys/nats/pdfs/2014-methodology-report-tag508.pdf)	Individual	Demographic information: gender, age, and race/ethnicity. Tobacco use, quit attempts, and success rates; reasons for smoking and reasons for quitting	Data are available online for download without restriction from the NATS Web site at https://www.cdc.gov/tobacco/data_statistics/surveys/nats/index.htm.
National Survey of Family Growth (NSFG)	Individuals (aged 15–44 y)	Multistage random sample	Individual	Topics include marriage; divorce; cohabitation; women's pregnancy and birth history; men's fathering of biological children; breastfeeding; adoption and nonbiological parenting; contraceptive use; intendedness of pregnancies; sexual intercourse and number of sexual partners; family planning and related medical services; infertility and use of infertility services; attitudes on sex, parenthood, marriage, and cohabitation; men's involvement as fathers with children they do and do not live with; other sexual behaviors (besides vaginal intercourse) that may carry risk of HIV or sexually transmitted infections, including same-sex behavior; and sexual orientation and attraction	Public-use NSFG data files dating back to 1973 are available (https://www.cdc.gov/nchs/nsfg/nsfg_questionnaires.htm). Some associated files, such as those that contain interviewer observations and paradata files that describe the NSFG data collection process, are available through the NCHS Research Data Centers at http://www.cdc.gov/rdc.
National Survey on Drug Use and Health (NSDUH)	Adults (including those 65 y and older)	Multistage complex sampling design (https://www.samhsa.gov/data/sites/default/files/cbhsqreports/NSDUHmrbSampleDesign2018/NSDUHmrbSampleDesign2018.pdf)	Individual	Use of tobacco, alcohol, and illicit drugs, broken down into subcategories (eg, opioids, types of tobacco used); substance use disorders and treatment in the past year. Mental health and behavioral health indicators, including major depressive episodes in the past year; mental illness among adults in the past year; mental health service use in the past year; co-occurring major depressive episodes and substance use	NSDUH data can be accessed at http://datafiles.samhsa.gov/studyseries/national-survey-drug-useand-health-nsduh-nid13517.

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
				among adolescents; co-occurring mental health issues and substance use disorders among adults; and suicidal thoughts and behavior among adults	

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; NCHS, National Center for Health Statistics.

TABLE 5

Data Sets Focused on American Adults 65 Years and Older

Data Set	Population	Sampling Method(s)	Unit of Analysis	Health Metric(s)/Data Collected	Data Availability and Processes
Health and Retirement Survey (HRS)	Individuals older than 50 y	Multistage area probability design (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3997380)	Individual	Health and health care service use; income and wealth, including retirement savings; retirement and employment status; disability status; and family support Data on biomarkers, genetic information, and a number of psychosocial variables including perceived well-being, social support, and self-related beliefs	A wide range of HRS data products are available at https://hrs.isr.umich.edu/dataproducts. These include both public-use data and restricted/sensitive data, such as that containing biomarker information. Most HRS data are publicly available; those that are not are available only under specific contractual conditions. More information about and instructions for accessing such data can be found at http://hrsonline.isr.umich.edu/index.php?p=reslis.
Medicare Current Beneficiary Survey (MCBS)	Medicare beneficiaries (mainly individuals older than 64 y and disabled persons)	Three-stage cluster sample design (https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Downloads/MCBS2015MethodReport508.pdf\051	Individual	Health outcomes, health care expenditures, and sources of payment for individuals of all ages who receive Medicare	De-identified MBCS data from 2013 forward are available in public-use files and accompanying documentation at https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File/index.html. These data files do not include the level of detail of the MCBS limited data set. Limited Survey and Cost Supplement data set files can be accessed via a permission process detailed at https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA_-_NewLDS.html.
National Health and Aging Trends (NHATS)	Medicare beneficiaries 65 y and older	Stratified 3-stage sample design (https://www.nhats.org/scripts/sampleDesign.htm)	Individual	Physical, social, technological, and service environment, physical and cognitive capacity, use of assistive devices and rehabilitation, help received with daily activities (self-care, household, and medical), participation in valued activities, and well-being	NHATS provides access to public-use data sets categorized by rounds (ie, years of data collection); files can be accessed via a permission process detailed at https://www.nhatsdata.org/. Researchers can apply for access to additional sensitive and restricted-use data sets by following the instructions at https://www.nhatsdata.org/ResDataFiles.aspx.

J Public Health Manag Pract. Author manuscript; available in PMC 2022 January 01.