

**Table 1: List of potential study data sets.**

Short Name (Acronym)	Full Name	Sponsor	Total N	Dates Available	Sample Selection	Data Restrictions	Data Summary	Website
<b>ACS</b>	American Community Survey	US Census Bureau	Varies	2005 – 2014	US Survey Sample	Publicly Available	Ongoing survey that provides vital information on a yearly basis about US population.	<a href="#">ACS</a>
<b>All of US Research Program</b>	All of US Research Program	NIH	>500,000	2017 – present	US diverse cohort	Publicly available with registration	Ongoing cohort with vast survey, HER, and biologic data	<a href="#">AoU</a>
<b>BRFSS</b>	Behavioral Risk Factor Surveillance System	Center for Disease Control (CDC)	Varies (~70,000 observations/year)	1984 - Present	US landline random sample	Free and publicly available	Collects prevalence data among adult U.S. residents regarding their risk behaviors and preventive health practices	<a href="#">BRFSS</a>
<b>CDC-Wonder</b>	Wide-Range Online Data for Epidemiological Research (Wonder)	(CDC)	Varies	Varies	US - Population based	Free and publicly available	Various public health data available. Data ranging from AIDS, births, to cancer.	<a href="#">CDC-Wonder</a>
<b>CMF</b>	Compressed Mortality File (CMF)	CDC	Varies	1968 – 2014	US-Population based	Free and publicly available	County-level national mortality and population database spanning the years 1968-2014.	<a href="#">CMF</a>
<b>DHS</b>	Demographic and Health Surveys	U.S. Agency for International Development (USAID)	Varies by country	1984 - Present	Survey samples for various countries	Free and publicly available	Nationally (from multiple countries) representative samples (usually between 5,000 and 30,000 households).	<a href="#">DHS</a>
<b>HRS</b>	Health and Retirement Study	NIH National Institute on Aging	~20,000	1998 – present	US Longitudinal survey sample (Adults 50+)	Publicly available (must register)	Examines temporal changes in work and health among older adults	<a href="#">HRS</a>
<b>NHAMCS</b>	National Hospital Ambulatory Medical Care Survey	CDC	Varies (large)	1992 -2022	US population based	Publicly available	NHAMCS collected data about medical services provided in hospital emergency and outpatient departments. NHAMCS data can answer questions about hospital-based medical care for patients in these settings	<a href="#">NHAMCS</a>

							(ambulatory medical care). The Health Information National Trends Survey (HINTS) collects nationally representative data routinely about the American public's use of cancer-related information. The survey:	
<b>HINTS</b>	Health Information National Trends Survey	NCI-NIH	~5000 participants per year	2003 – Present	US Survey Population based	Publicly available	Provides updates on changing patterns, needs, and information opportunities in health Identifies changing communications trends and practices Assesses cancer information access and usage Provides information about how cancer risks are perceived Enables researchers to test new theories in health communication	<a href="#">HINTS</a>
<b>SEER</b>	Surveillance, Epidemiology, and End Results Program	NCI-NIH	100,000+ Cases per year	1969 – Present	Sample of 21+ US cancer registries	Publicly available	The Surveillance, Epidemiology, and End Results (SEER) Program provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. SEER is supported by the Surveillance Research Program (SRP) in NCI's Division of Cancer Control and Population Sciences (DCCPS)	<a href="#">SEER</a>
<b>UK Healthcare Data</b>	UK Center for Clinical and Translational Science	University of Kentucky	Varies	2004- Present (UK Healthcare)	Patients within UK healthcare system	UK faculty and students		<a href="#">UK CCTS</a>

**Table 1 Continued: List of potential study data sets.**

Short Name (Acronym)	Full Name	Sponsor	Total N	Dates Available	Sample Selection	Data Restrictions	Data Summary	Website
<b>HCUP Databases</b>	Healthcare Cost and Utilization Project	US Department of Health & Human Services	Varies (~7 million hospital stays/year)	1997 – 2013	Pre-2012: Sample of US hospitals from which all discharges were retained Post-2012: Sample of discharge records from all HCUP-participating hospitals	1) Free and publicly available.  2) Purchasable and more specific data.	Can identify, track, and analyze national trends in health care utilization, access, charges, quality, and outcomes.	<a href="#">HCUP Simple</a>  <a href="#">HCUP Databases</a>
<b>NGHIS</b>	National Historical Geographic Information System	University of Minnesota	Varies	1790 - Present	US Census and other surveys	Free and publicly available	Free online access to summary statistics and GIS boundary files for U.S. censuses and other nationwide surveys from 1790 through the present.	<a href="#">NHGIS</a>
<b>NHANES</b>	National Health and Nutrition Examination Survey	CDC	Varies	1959 - Present	US Survey Sample	1) Free and publicly available.  2) Purchasable and more specific data.	Program of studies designed to assess the health and nutritional status of adults and children in the United States.	<a href="#">NHANES</a>
<b>NEISS</b>	National Electronic Injury Surveillance System	US Consumer Product Safety Commission	Varies – obtained from >100 US Hospitals	1979 - Present	National probability sample	Free and publicly available	Information extracted from medical charts, including patient demographics (i.e., age, sex, and race) and injury information.	<a href="#">NEISS</a>

**Table 2: List of cohort studies that require study proposals and PI approval (min 8 weeks for most) before use in a potential study.**

Cohort Name	Total <i>N</i>	Dates of enrollment	Years of Follow-up	Race (%)	Sex (%)	Age at enrollment	Region/ Area	Sample Selection	Data Restrictions (Web Link)
<b>Atherosclerosis Risk in Communities Study (ARIC)</b>	15,792	1987 – 1989	2014 (27)	Black (27.3) White (72.6)	Male (44.8)	45 – 69	1) Suburban Minneapolis, MN 2) Jackson, MS 3) Forsyth County, NC 4) Washington County, MD	Random	Must complete ancillary study proposal: See: <a href="#">ARIC</a>
<b>Rochester Epidemiology Project (REP)</b>	>493,606	1966 - present	Present (49)	Black (4.8) White (85.7)	Male (48.9)	All Ages	Southeastern, MN – Rochester, Olmsted County, MN	Population-based	Must complete proposal, “REP Collaborator” appointment through Mayo Clinic: See: <a href="#">REP</a>
<b>Health, Aging, and Body Composition (Health ABC) Study</b>	3,075	March 1997 – July 1998	-	Black White	-	70 – 79	1) Memphis, TN 2) Pittsburgh, PA	Random	Must complete ancillary study proposal See: <a href="#">Health ABC</a>
<b>The Cardiovascular Health Study (CHS)</b>	~5,888	1989 – 1999	1999 (10)	Black White	-	65+	1) Allegheny County, PA 2) Forsyth County, NC 3) Sacramento County, CA 4) Washington County, MD	Random	Must complete ancillary study proposal: See: <a href="#">CHS</a>

**Table 2 continued: List of cohort studies that require study proposals and PI approval (min 8 weeks for most) before use in a potential study.**

Cohort Name	Total <i>N</i>	Dates of enrollment	Years of Follow-up	Race (%)	Sex (%)	Age at enrollment	Region/ Area	Sample Selection	Data Restrictions
<b>Jackson Heart Study (JHS)</b>	5,301	2000 - present	Present (15)	Black participants only	-	35 – 84	Jackson, MS	17% Random 22% Volunteer 31% ARIC 30% Secondary Family Members	Complete ancillary study: <a href="#">JHS</a>
<b>Framingham Heart (Offspring) Study</b>	5,124	1971	2014 (43)	-	Male (48.5)	All ages (0 – 70)	Framingham, MA	Population-based	Complete research application: <a href="#">Framingham</a>
<b>REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort</b>	30,239	2003 - 2007	Present (12)	59% White 41% Black	55% Female 45% Male	45+	US Population sample	Population-based	Must Complete Proposal: <a href="#">REGARDS</a>
<b>NIH-AARP</b>	567,000	1995 – Present	Present	4% Black	50% Male	62+	US sample	Prospective cohort	Complete proposal here: <a href="#">NIH-AARP</a>

ONLINE COLLECTIONS:

<https://research.fiu.edu/irb/public-use-datasets/>

<https://www.nhgis.org>

<https://guides.library.unt.edu/datasets/public-data-sources>