

Public Access Datasets

There are numerous free and publicly accessible data sources available to study HIV related research questions. Most of the following are public access; however some may be limited or restricted data access only. Consult the website or contact person for further information. Please remember that no matter where you obtain your data, you should be aware of IRB requirements for all of your research.

Name	Description	Link
Agency for Healthcare Research and Quality	AHRQ provides a range of data resources in the form of online, searchable databases. Data are provided on topics such as the use of health care, the costs of care, trends in hospital care, health insurance coverage, out-of-pocket spending, and patient satisfaction. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	http://www.ahrq.gov/research/data/dataresources/index.html
AIDSVu	AIDSVu is an interactive online map illustrating the prevalence of HIV in the United States.	http://aidsvu.org/
American College of Surgeons National Trauma Data Bank (NTDB)	The National Trauma Data Bank® (NTDB®) is the largest aggregation of U.S. trauma registry data ever assembled.	https://www.facs.org/quality%20programs/trauma/ntdb
American Hospital Association Annual Survey	A web-based national source of proprietary hospital and health system data collected and verified by the American Hospital Association. AHA DataViewer content includes exclusive national survey data (e.g., AHA Annual Survey of Hospitals, AHA Annual Survey of Hospitals—IT Supplement), proprietary AHA membership data and much more.	http://www.aha.org/research/rc/stat-studies/data-and-directories.shtml
Behavioral Risk Factor Surveillance System	BRFSS is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services.	http://www.cdc.gov/brfss/

Cancer Genome Atlas / NCI	Search, download, and analyze data sets containing clinical information, genomic characterization data, and high level sequence analysis of the tumor genomes.	https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp
CDC Wide-ranging Online Data for Epidemiologic Research (WONDER)	WONDER online databases utilize a rich ad-hoc query system for the analysis of public health data. Reports and other query systems are also available.	wonder.cdc.gov
CDC's National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP).	The NCHHSTP Atlas gives you the power to access data reported to CDC's National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP). Use HIV, STD, hepatitis, and TB data to create maps, charts, and detailed reports, and analyze trends and patterns. (Also has a variety of other resources, slidesets, and maps.)	http://www.cdc.gov/nchhstp/Atlas/
Center for AIDS Prevention Studies (CAPS)	Survey Instruments and Scales (A resource of free, validated instruments)	http://caps.ucsf.edu/resources/survey-instruments
Comprehensive Hospital Abstract Reporting System	The Comprehensive Hospital Abstract Reporting System (CHARS) is a Department of Health system used to Identify and analyze hospitalization trends, establish statewide diagnosis related group (DRG) weights, a way of comparing hospital stays across all hospitals, identify and quantify health care access, quality, and cost containment issues.	http://www.doh.wa.gov/ForPublicHealthandHealthcareProviders/HealthcareProfessionsandFacilities/DataReportingandRetrieval/HospitalInpatientDatabaseCHARS
Demographic and Health Surveys, the DHS Program, USAID	The DHS program has collected, analyzed, and disseminated accurate and representative data on population, health, HIV, and nutrition through more than 300 surveys in over 90 countries.	http://dhsprogram.com/data/

dbGaP (NIH GWAS Database of Genotypes and Phenotypes)	The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans.	http://www.ncbi.nlm.nih.gov/gap
Fatality Analysis Reporting System (FARS)	FARS is a nationwide census providing NHTSA, Congress and the American public yearly data regarding fatal injuries suffered in motor vehicle traffic crashes.	http://www.nhtsa.gov/FARS
Forum for Collaborative HIV Research	A list of dozens of HIV cohort databases for collaborative study. A result of the 17th International Workshop on HIV Observational Databases. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	http://www.hivforum.org/projects/informationreference-services/collaborations-and-databases/495-hiv-cohorts-and-databases-2011
General Social Survey	Since 1972, the General Social Survey (GSS) has provided politicians, policymakers, and scholars with a clear and unbiased perspective on what Americans think and feel about such issues as national spending priorities, crime and punishment, intergroup relations, and confidence in institutions.	http://www3.norc.umd.edu/GSS+Website/
Health and Retirement Study	A longitudinal panel study that surveys a representative sample of approximately 20,000 Americans over the age of 50 every two years. Supported by the National Institute on Aging (NIA U01AG009740) and the Social Security Administration, the HRS explores the changes in labor force participation and the health transitions that individuals undergo toward the end of their work lives and in the years that follow.	http://hrsonline.isr.umich.edu/

Health Information National Trends Survey	HINTS collects data about the use of cancer-related information by the American public. These data provide opportunities to understand and improve health communication.	Hints.cancer.gov
Health Services Research Information Central	Extensive listing of public access data sources from the US Federal government. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	https://www.nlm.nih.gov/hsriinfo/datasites.html#164Databases/Repositories
HIV/AIDS Surveillance Database	HIV/AIDS Surveillance Database contains epidemiological information for developing countries presented at international and regional conferences on HIV/AIDS as well as additional material from other sources. The database contains over 21,000 individual data records from over 2,700 publications and presentations. The Center for International Research has been compiling HIV seroprevalence information contained in journals, articles and public presentations since 1987.	http://www.ciesin.org/datasets/hivaids/hivaids-home.html
HIV Prevention Trials Network D01	Vaccine Preparedness Study/Uninfected Protocol Cohort. One of the first HIV Prevention Trials Network studies.	http://www.hptn.org/network_information/public_datasets.htm
Hospital Compare	Hospital Compare is a consumer-oriented website that provides information on how well hospitals provide recommended care to their patients. This information can help consumers make informed decisions about health care. Hospital Compare allows consumers to select multiple hospitals and directly compare performance measure information related to heart attack, heart failure, pneumonia, surgery and other conditions.	https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html
HIV/AIDS Surveillance Database	HIV/AIDS Surveillance Database contains epidemiological information for	http://www.ciesin.org/datasets/hivaids/hivaids-home.html

	developing countries presented at international and regional conferences on HIV/AIDS as well as additional material from other sources.	
Inter-University Consortium for Political and Social Research (ICPSR)	Numerous datasets The Social Science Variables Database (SSVD) enables ICPSR users to examine and compare variables and questions across studies or series. The SSVD currently includes over 4 million variables, representing about 76% of ICPSR's holdings that have quantitative data described in statistical syntax.	http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/subject.jsp
MACS Public dataset	Multicenter AIDS Cohort Study is one of the longest standing cohort studies of the natural history of HIV in men.	https://statepi.ihsph.edu/macspdt.html
Medicare Healthcare Cost Report Information System (HCRIS)	Medicare-certified institutional providers are required to submit an annual cost report to a Medicare Administrative Contractor (MAC). The cost report contains provider information such as facility characteristics, utilization data, cost and charges by cost center (in total and for Medicare), Medicare settlement data, and financial statement data.	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Cost-Reports/?redirect=/CostReports/
National Ambulatory Medical Care Survey	The National Ambulatory Medical Care Survey (NAMCS) is a national survey designed to meet the need for objective, reliable information about the provision and use of ambulatory medical care services in the United States.	http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm
National Center for Health Statistics data linkage activities	Extensive cross dataset linkages including NDI, CMMI, etc. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm

National Health and Nutrition Examination Study (NHANES)	NHANES is NCHS' most in-depth and logistically complex survey, operating out of mobile examination centers that travel to randomly selected sites throughout the country to assess the health and nutritional status of Americans.	http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm
National Health Interview Survey	The National Health Interview Survey (NHIS) is the principal source of information on the health of the civilian noninstitutionalized population of the United States and is one of the major data collection programs of the National Center for Health Statistics (NCHS) which is part of the Centers for Disease Control and Prevention (CDC).	http://www.cdc.gov/nchs/nhis/nhis_questionnaires.htm
National Home and Hospice Care Survey	The National Home and Hospice Care Survey (NHHCS) is one in a continuing series of nationally representative sample surveys of U.S. home health and hospice agencies.	http://www.cdc.gov/nchs/nhhcs/nhhcs_questionnaires.htm
National Hospital Ambulatory Medical Care Survey	The National Hospital Ambulatory Medical Care Survey (NHAMCS) is designed to collect data on the utilization and provision of ambulatory care services in hospital emergency and outpatient departments and in ambulatory surgery centers.	http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm
National Hospital Discharge Survey	The National Hospital Discharge Survey (NHDS), conducted from 1965 to 2010, is a national probability survey designed to meet the need for information on characteristics of inpatients discharged from non-Federal short-stay hospitals in the United States.	http://www.cdc.gov/nchs/nhds/nhds_questionnaires.htm
National Immunization Survey	The National Immunization Surveys are a group of phone surveys used to monitor vaccination coverage among children 19-35 months, teens 13-17 years, and flu	http://www.cdc.gov/nchs/nis/data_files.htm

	vaccinations for children 6 months-17 years.	
National Nursing Home Survey	The National Nursing Home Survey (NNHS) is one in a continuing series of nationally representative sample surveys of United States nursing homes, their services, their staff, and their residents.	http://www.cdc.gov/nchs/nnhs/nnhs_questionnaires.htm
National Nursing Assistant Survey	The National Nursing Assistant Survey (NNAS) is the first national study of nursing assistants working in nursing facilities in the United States.	http://www.cdc.gov/nchs/nnas.htm
National Survey of Ambulatory Surgery	The National Survey of Ambulatory Surgery (NSAS) is the only national study of ambulatory surgical care in hospital-based and freestanding ambulatory surgery centers (ASCs).	http://www.cdc.gov/nchs/nsas/nsas_questionnaires.htm
National Survey of Family Growth	The National Survey of Family Growth, or NSFG, was initially designed to be the national fertility survey of the United States. So its focus was on factors that help to explain trends and group differences in birth rates, such as contraception, infertility, sexual activity, and marriage.	http://www.cdc.gov/nchs/nsfg/nsfg_questionnaires.htm
NIH Data Sharing Repositories	NIH-supported data repositories that make data accessible for reuse. Most accept submissions of appropriate data from NIH-funded investigators (and others), but some restrict data submission to only those researchers involved in a specific research network. Also included are resources that aggregate information about biomedical data and information sharing systems. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

National Latino and Asian American Study (NLAAS)	The CPES Combined Dataset combines the data from three epidemiological surveys: the National Latino and Asian American Study (NLAAS), the National Survey of American Life (NSAL) and the National Comorbidity Survey – Replication (NCS-R).	http://www.multiculturalmentalhealth.org/nlaas.asp
National Automotive Sampling System (NASS) and General Estimates System (GES)	NASS is composed of two systems - the Crashworthiness Data System (CDS) and the General Estimates System (GES). These are based on cases selected from a sample of police crash reports. CDS data focus on passenger vehicle crashes, and are used to investigate injury mechanisms to identify potential improvements in vehicle design. GES data focus on the bigger overall crash picture, and are used for problem size assessments and tracking trends.	http://www.nhtsa.gov/NASS
National Center for Education Statistics	The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education.	https://nces.ed.gov/
National Election Studies	To serve the research needs of social scientists, teachers, students, policy makers own surveys on voting, public opinion, and political participation.	http://www.electionstudies.org/
National Epidemiological Survey on Alcohol and Related Conditions (NESARC)	In 2001—2002, the National Institute on Alcohol Abuse and Alcoholism (NIAAA) conducted the first wave of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), the largest and most ambitious survey of this type conducted to date.	http://pubs.niaaa.nih.gov/publications/AA70/AA70.htm
National Longitudinal Survey (NLSY)	The National Longitudinal Surveys (NLS) are a set of surveys designed to gather information at multiple points in time on the labor market activities and	http://www.bls.gov/nls/

	other significant life events of several groups of men and women.	
National Survey of Children's Health	The National Survey of Children's Health (NSCH) has been conducted three times between 2003 and 2012. It provides rich data on multiple, intersecting aspects of children's lives—including physical and mental health, access to quality health care, and the child's family, neighborhood, school, and social context.	http://childhealthdata.org/learn/NSCH
National Survey of Children with Special Health Care Needs	The National Survey of Children with Special Health Care Needs (NS-CSHCN) was conducted three times between 2001 and 2010. It was designed to take a close look at the health and functional status of children with special health care needs in the U.S.—their physical, emotional and behavioral health, along with critical information on access to quality health care, care coordination of services, access to a medical home, transition services for youth, and the impact of chronic condition(s) on the child's family.	http://www.childhealthdata.org/learn/NS-CSHCN
National Technical Information Service	Federal resource for thousands of open access datasets. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	http://www.ntis.gov/
Organ Procurement and Transplantation Network	All data within the OPTN database is collected via an online Web application called UNet sm . Transplant professionals from hospitals, histocompatibility (tissue typing) laboratories, and organ procurement organizations located across the country use the application.	https://optn.transplant.hrsa.gov/
Publicly Available Databases for Aging-Related Secondary	This document provides snapshots of selected publicly available data collections supported in whole or in part by the National Institute on Aging Division	https://www.nia.nih.gov/research/dbsr/publicly-available-databases-aging-related-

Analyses in the Behavioral and Social Sciences	of Behavioral and Social Research (BSR) to promote understanding of aging populations both domestically and throughout the world. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	secondary-analyses-behavioral-and-social
PROMIS	Patient Reported Outcomes Measurement Information System (PROMIS) is a system of highly reliable, precise measures of patient-reported health status for physical, mental, and social well-being. PROMIS tools measure what patients are able to do and how they feel by asking questions. Numerous datasets listed; these may not all be non-human subjects eligible; consult with SPH or IRB first.	http://www.nihpromis.org/science/publicusedata
Roper Center for Public Opinion Network	The Roper Center for Public Opinion Research, currently located at Cornell University, is one of the world's leading archives of social science data, specializing in data from public opinion surveys. The Center's mission is to collect, preserve, and disseminate public opinion data; to serve as a resource to help improve the practice of survey research; and to broaden the understanding of public opinion through the use of survey data in the United States and abroad.	http://ropercenter.cornell.edu/about-the-center/
School Health Policies and Programs Study (SHPPS)	The School Health Policies and Practices Study* (SHPPS) is a national survey periodically conducted to assess school health policies and practices at the state, district, school, and classroom levels.	http://www.cdc.gov/healthyyouth/data/shpps/data.htm
State and Local Area Integrated Telephone Survey	SLAITS allows us to produce state-level data on such topics as the health of	http://www.cdc.gov/nchs/slait/nsch.htm

	children with special needs, to meet the data needs of our colleagues in HHS' Maternal and Child Health Bureau and elsewhere. (Includes National Survey of Children's Health, National Survey of Children with Special Health Care Needs, and numerous other specialized surveys on various health conditions.	
Survey of Consumer Finances (SCF)	The 2013 Survey of Consumer Finances (SCF) is the most recent survey conducted. Below are links to the bulletin article, historical bulletin tables, full public dataset, extract dataset, replicate weight files, and documentation.	http://www.federalreserve.gov/econresdata/scf/scfindex.htm
U.S. Bureau of the Census	Our surveys provide periodic and comprehensive statistics about the nation, critical for government programs, policies, and decision making.	http://www.census.gov/data.html
U.S. Bureau of Labor Statistics	Extensive US labor statistics	http://www.bls.gov/data/
The Voluntary HIV-1 Counseling and Testing Efficacy Study Group	Individual or couple participants were randomly assigned HIV-1 VCT or basic health information. At first follow-up (mean 7.3 months after baseline) health-information participants were offered VCT and all VCT participants were offered retesting. Sexually transmitted infections were diagnosed and treated at first follow-up. The second follow-up (mean 13.9 months after baseline) involved only behavioural assessment, and all participants were again offered VCT. 3120 individuals and 586 couples were enrolled. Lancet. 2000 Jul 8;356(9224):103-12.	http://caps.ucsf.edu/resources/datasets
Vital statistics	Extensive US vital health statistics.	http://www.cdc.gov/nchs/data_access/vitalstatsonline.htm

Youth Risk Behavior Surveillance System	YRBSS monitors six types of health-risk behaviors that contribute to the leading causes of death and disability among youth and adults.	http://www.cdc.gov/healthyyouth/data/yrbs/index.htm
World Health Organization	The GHO data repository provides access to over 1000 indicators on priority health topics including mortality and burden of diseases, the Millennium Development Goals, non-communicable diseases and risk factors, epidemic-prone diseases, health systems, environmental health, violence and injuries, equity among others.	http://www.who.int/gho/database/en/