

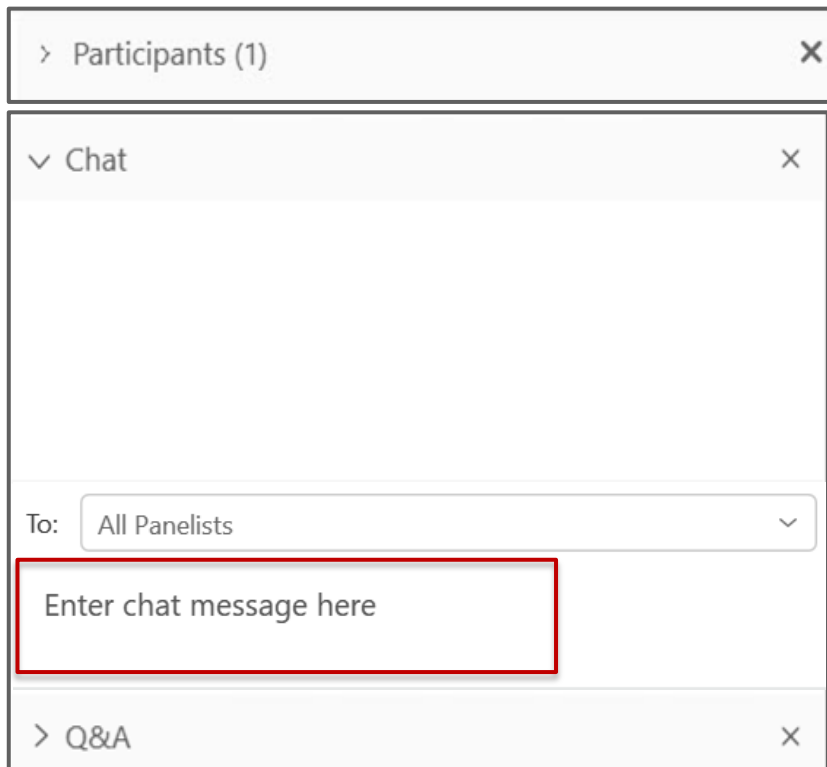
Introduction to Publicly Available Resources for Cancer Prevention and Control Research

Presentation for T-32 Fellows

Richard P. Moser, PhD

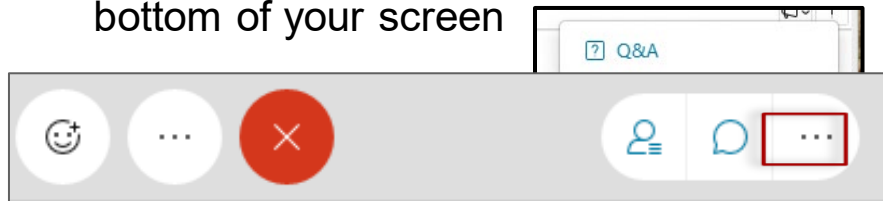
Karen Emmons, PhD

Using WebEx and Webinar Logistics



The screenshot shows the WebEx interface with two panels. The top panel is titled 'Participants (1)' and has a close button (X). The bottom panel is titled 'Chat' and also has a close button (X). Inside the Chat panel, there is a 'To:' dropdown menu set to 'All Panelists'. Below this is a text input field with the placeholder text 'Enter chat message here', which is highlighted with a red rectangular border. At the bottom of the Chat panel is a 'Q&A' section with a right arrow and a close button (X).

- All lines will be in listen-only mode
- Submit comments at any time using the **Q&A** feature and select All Panelists
- You may need to activate the appropriate panel using the menu option found at the bottom of your screen



- This webinar is being recorded

Webinar Series Organizers



Richard P. Moser, Ph.D.

Training Director and Research Methods
Coordinator
NCI Division of Cancer Control and
Population Sciences



Karen Emmons, Ph.D.

Professor Social and Behavioral
Sciences
Harvard T.H. Chan School of Public
Health

Welcome!

Purpose:

- To introduce publicly available data resources that are relevant to cancer prevention and control
- To help fellows consider use of these resources to answer new and innovative questions that fall within NCI priorities



Tools for Cancer Research

- Expand skills
- Enhance career opportunities
- Caveats
 - Not meant to be comprehensive
- Interactive: your input needed



“Give me six hours to chop down a tree and I will spend the first four sharpening the axe.”

Abraham Lincoln



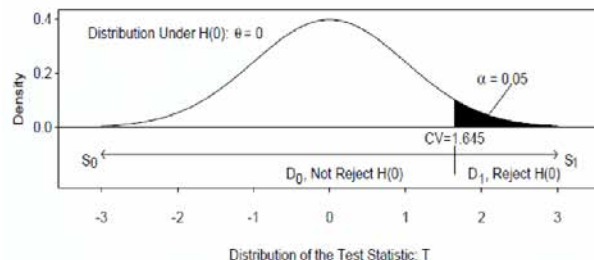
Agenda

Note: We'll stop after each section for Q & A and input from attendees

1. *Using Secondary Health Data*
2. *Broad, Population-Level Data*
3. *Youth-Focused Data*
4. *Cancer-Focused Data*
5. *Data Collection Tools*
6. *Research Methods/Data analysis Tools*
7. *Data Visualization Tools*
8. *Other Resources*

Using Secondary Health Data

- Hypothesis generation/testing
- Publications
 - Check on websites/Pubmed for current publications
- Strengths:
 - Large samples
 - Many different constructs
 - Population estimates (national, census regions; possibly state/county/census tract)
 - Testing trends over time (where possible)
 - Used for data linkages
- Limitations:
 - Non-experimental
 - Constructs measured by fewer items
 - May require special statistical software/methods/permission
 - Sensitive information
 - Geocodes
 - Accounting for complex samples



Broad, Population-Level Data



Population

Households, families, adults and children

Method

Face to face interview

Content

Health conditions and behaviors, access to and use of health services

Physical activity, smoking, alcohol use

Cancer supplements

-Screening in 2021, 2023...

-Physical activity environment in 2020, 2024...

CAM

Immunization/vaccination usage

Tobacco and E-cig Use

Food Security

Data

n~40,000 households (~87,000 individuals)

Initiated in 1957

Note

To see list of core/rotating/supplements:
<https://healthcaredelivery.cancer.gov/nhis/>

New: Interactive data query system



National Health and Nutrition Examination Survey (NHANES)

Population

Children and Adults

Method

Face to face interview

Physical exams

Content

Chronic and Infectious Disease Prevalence

Mental health and cognitive functioning

Energy Balance

Respiratory disease

Anthropometry

Biospecimens (DNA, serum, plasma, urine)

Data

N~5,000/Year

Initiated in 1960's; Annual since 1999

2017-2020 latest available data

On-line tutorial

Note

Supplemental food survey available: What We Eat in America



Centers for Disease
Control and Prevention
National Center for
Health Statistics



Population

Adults (18+)

Method

Random Digit Dial telephone survey

State Administration

Content

Behaviors associated with chronic diseases,
injuries, and infectious diseases

Sexual behavior

Smoking

Cancer screening

Diet and exercise

Data

>400,000 subjects/year

Core questions asked of everyone and
state-specific modules

Data can be combined across states to
get national estimates

Includes D.C., Puerto Rico, U.S. Virgin
Islands, Guam

2019 are latest available data; maps also
available

Notes

SMART BRFSS—city-level and
surrounding counties estimates

American Time Use Survey



U.S. BUREAU OF LABOR STATISTICS

Population

Adolescents/adults 15 and older

Method

Self report telephone interview using 24 hour recall

Content

Estimates of activities people do (work, childcare, socializing, exercising, eating, educational, sports and religious activities), whom they were with, and the time spent doing them by sex, age, educational attainment, labor force status, and other characteristics, as well as by weekday and weekend day.

Eating and health module.

Data

n ~ 13,000 per year

Data currently available: 2003-2019

Note

Data files can be linked to Current Population Survey (CPS)

<https://www.bls.gov/tus/>

Population -Adolescents/adults 15+ 1992-06; 18+ 2007-

Method –National HH address-based frame, 8 panels

Conducted every 3-4 years by Census for NCI

64% telephone (allows cell phone if preferred); 36% in-home

Translated into Spanish and 4 Asian languages

Content: Monitor, evaluate and conduct research on Cigarette & other tobacco product usage patterns;

Cessation- attempts, intentions, & treatment

Policy- work, home, “real” price, attitudes & clinician advice

Data: ~240,000 U.S. respondents per cycle; 2018-2019 are latest data;

Notes:

National, state, sub-state estimates;

Health disparities (e.g. race/ethnicity; low SES, rural)

Economic aspects with CPS detailed occupational, & health disability data

Panel design links to other CPS data (e.g., ASEC, ATUS, Food Security, Internet Use)

Panel design allows for adding prospective Follow-Up (2002-03, 2010-2011)

Linkage to outcome data (mortality and SEER data) through NLMS



<https://hints.cancer.gov/>

Population

Non-Institutionalized US Adults (18+)

Method

2003, 2005: Random digit dial (RDD)

2008: Dual frame/dual mode

HINTS 4-5: Self-administered mail survey

HINTS 5 Cycle 3: Mixed mode

Data

14 Iterations (2003-2020)

HINTS 5 Cycle 4 data released Jan 2021

HINTS FDA 1 & 2

Content

Health communications trends and practices

Cancer information access and usage

Cancer risk perception

Mental models of cancer

Health behaviors

Caregiving

Note

Web pilot study done with HINTS 5 Cycle 3 (2019)

Tutorial on website

Medical Expenditure Panel Survey



Population

Nationally representative household survey, with additional follow-up of their medical providers (doctors, hospitals, etc.), and separate survey of employers across the United States about health insurance

Household Survey Method

Five in-person interviews

Content

Demographics, health status

Employment, insurance coverage and access to care

Health care utilization and expenditures

Data

Subset of NHIS with N= ~40,000/survey

New cohort initiated annually

Note

Has overlapping panel component

Latest data from 2019; ongoing data release schedule (see website for schedule)

Population

Goal of a cohort of 1 million (50% from ethnic/racial minority groups)

Method

From electronic health records (EHRs), biospecimens, surveys, and other measures

Content

Environment: drug exposure; medical utilization

Lifestyle: tobacco, alcohol, drugs

Biology: sodium, glucose, medical history

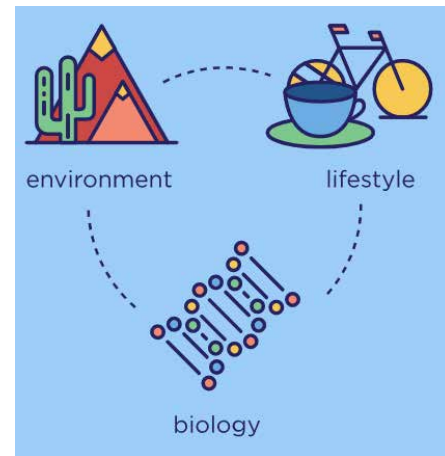
Genomic sequence data

Data

Includes physical measures; biological samples; wearables data

Note

Need to apply for data access



<https://www.researchallofus.org/>

California Health Interview Survey

Population

Adult, adolescent and child questionnaires
Very diverse racial/ethnic population



Method

Telephone survey (landline/cell phone) of all California counties; moving to mixed mode (web and telephone) for 2019

Content

Health behaviors (Diet/drug use/sexual/sun safety)
Health status
Health conditions (asthma, diabetes etc.)
Cancer history and prevention
Health insurance

Data

Latest data available is for 2018
~20,000 respondents/survey

<http://healthpolicy.ucla.edu/chis/Pages/default.aspx>

Note

Many Latino and Asian groups represented; oversamples of Koreans and Vietnamese
Fielded in five languages: English, Spanish, Chinese, Vietnamese, and Korean
Restricted geo-coded data available

Canadian Community Health Survey (CCHS)



Population

Children (12+) to adults

Coverage: 98% in the provinces; in Territories: 92% in the Yukon, 96% in the Northwest Territories and 92% in Nunavut

Method

50% in-person; 50% telephone interviews

Content

Disability, Diseases and health conditions

Factors influencing health; Injuries

Health services performance and utilization

Mental health and well-being

Prevention and detection of disease

Data

n= ~120,000; 2001, 2003, and 2005 collected biennially

2007 onward collected annually; sampling frame changed in 2015

Note

2019 latest data available

<https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3226>

Canadian Health Measures Survey (CHMS)



Population

Ages 6-79

Covers ~ 97% of Canadians

Method

In-person interviews and physical exams

Content

Nutrition, smoking habits, alcohol use, medical history, current health status, sexual behavior, lifestyle and physical activity, the environment and housing characteristics, socioeconomic variables

Blood measures; urine measures, accelerometry, nutrition;

Obesity, hypertension, cardiovascular disease;

Data

n= ~5,000; launched in 2007; biennial

Note

Latest data from 2019

Youth-Focused Data

Youth Risk Behavior Surveillance System (YRBSS)

Population

9th-12th grade students

Method

National, state, tribal, special populations

Biennial, school-based and local surveys

Content

Behaviors that contribute to unintentional injuries and violence

Sexual behaviors; STDs, HIV

Alcohol and other drug use

Tobacco use

Dietary behaviors

Physical activity



<https://www.cdc.gov/healthyyouth/data/yrbs/index.htm>

Data

~12,000-16,000 respondents per survey

Note

Data collected since 1991

Most recent data from 2019 (released August, 2020)

National Survey on Drug Use and Health



Population

12 and older

Method

In-person interview

Content

Use of tobacco products, alcohol, illicit drugs by friends and self; treatment for substance abuse and mental health disorders; perceived risk of drug use; health care access and coverage

Data

N= ~70,000/year

Conducted annually

Note

National and state-level data

Began in 1971; 2019 most recent data

Data available through ICPSR



Monitoring the Future



Population

8th, 10th and 12th grade students

Method

Multi-stage area probability sample;

On-site questionnaires completed at school; follow-up with subset of subjects

Content

Drug (prescription, illicit, OTC), tobacco, alcohol, steroid use, vaping; alcohol availability; risk perceptions of drug use; HIV/AIDS

Data

N= ~50,000/year from public and private schools

Note

Data available through ICPSR

Core and topical questions

Started in 1975; 2019 most recent data

<http://www.monitoringthefuture.org/>

Population

Adolescents (grades 7 thru 12) from 80 High Schools and 52 Middle Schools; started in 1994-95, follow-up in 2008 (ages: 24-32) and 2016-2018;

Method

In-school questionnaire and in-person interview

Content

Health conditions and behaviors; access to and use of health services; social, psychological and physical well-being; risk behaviors; **biospecimen collection**

Data

N= ~6,504

Note

Follow-up data available in waves; most recent from 2016-2018
Data now managed by ICPSR



Classification of Laws Associated with School Students

RESEARCHERS go to CLASS to:

- Download data to analyze school diet and physical activity-related laws.

Use the data to: Examine laws within and across grade levels.

- Link the data with other surveillance, health, and education data sets.
- Track changes in laws over time.
- Evaluate factors associated with implementation of laws.

POLICYMAKERS go to CLASS to:

- Monitor how state PE and nutrition policies compare to other states using the map and state report functions.

Use the data

to: Examine whether states are meeting the most current standards and recommendations for PE and nutrition.

- Evaluate how state laws compare to other states.

SCHOOL BOARDS, TEACHERS, AND THE PUBLIC go to CLASS to:

- Examine how local policies compare to state laws.

Use the data to: Examine laws within and across grade levels.

<https://class.cancer.gov/>

Policy Areas Covered In CLASS Include:

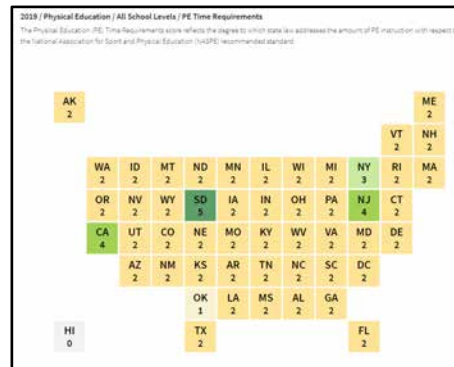
Physical Education (11 across grade levels)

- PE class time and curriculum standards
- PE teacher and staffing requirements
- Proficiency requirements
- Physical activity time requirements
- Moderate-to-vigorous activity in PE
- Assessment of health-related fitness
- Recess time
- Joint-use agreement provisions

Nutrition (19 across grade levels)

- Competitive food and beverage standards
- School meal environment and reimbursable meals
- Potable water and nutrition education requirements
- Food service director qualifications
- Coordinating or advisory council requirements
- Farm-to-School provisions and marketing restrictions
- Body mass index screening
- Provisions around parties, events, and fundraisers

- Data and maps
- Latest data from 2019



Population

Dyads of caregivers and their adolescent children (12-17) between Apr-Oct 2014

Method

Surveys; subset wore accelerometer and completed activity log

Content

Diet-related and Physical Activity related behaviors

Behavioral Correlates (sun-safety, tobacco, sleep)

Community/neighborhood/home environment

Data

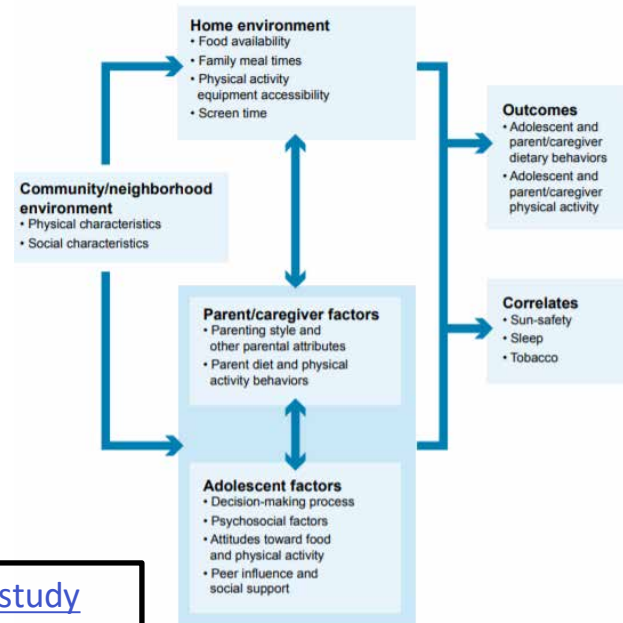
N=1,479 dyads

Note

Geodata available with neighborhood characteristics;

restricted access

<https://cancercontrol.cancer.gov/brp/hbrb/flashe-study>



Cancer-Focused Data



Population

Children to adults

Method

Data collected from cancer registries that cover ~50% of the US population; follow-up with individual cases until death

Content

Cancer incidence, prevalence, and survival data; cancer site, stage, morphology, treatment; limited demographics (age, race/ethnicity, marital status, region)

Data

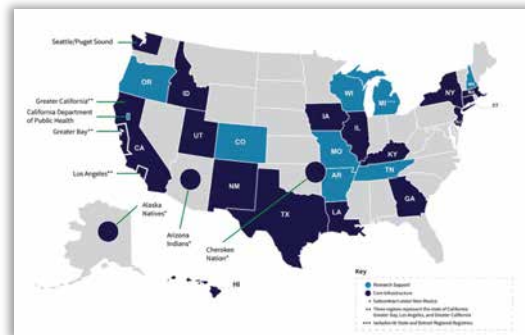
100% of cancer cases in registries (16); Six million cases with 550,000 added each year; 1973 to 2018;

Note

Need specialized software to analyze (SEER*Stat or SEER*Prep) downloaded from website;

Must sign user agreement to obtain; limited to research purposes

SEER linked to Medicare, Medicare Health Outcomes, CAHPS



PATH: Population Assessment of Tobacco and Health



Population

Household-based, nationally representative, longitudinal cohort study of adults and youth

Method

In-person interview with additional study activities

Content

Self-report

Biomarker data files

Biospecimens (urine, serum, plasma, genomic DNA)

Data

n~46,000 in wave 1; up to wave 4 of data collection (n=52,731)

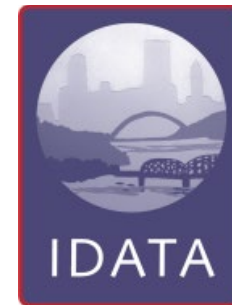
biomarker data not available for all

Note

Started in 2013

Must request access to data and outline proposed project

Interactive Diet and Activity Tracking in AARP (IDATA)



Population

Adults aged 50-74

Method

Survey and biomarkers

Self-report and wearables

Content

Physical activity self-report, physical activity objective monitors, diet, participant characteristics, biomarkers

Data

n=1082; participants assigned to 1 group (of 4) and followed a specified 12 month assessment protocol (2012-2013)

Note

All participants wore accelerometers

Must request access to data and outline proposed project through Cancer Data Access System (CDAS)

<https://cdas.cancer.gov/idata/>

NIH-AARP Diet and Health Study



NATIONAL CANCER INSTITUTE

Division of Cancer Epidemiology & Genetics

Population

50-71 adult members of AARP 1995-1996; 2004

Method

Mailed questionnaires; follow-up occurred in 2004

Content

Dietary behaviors

Lifestyle Questions

Risk Factors

Cancer and health status

Data

n~566,000 for baseline questionnaire; n~327,000 for risk factor questionnaire

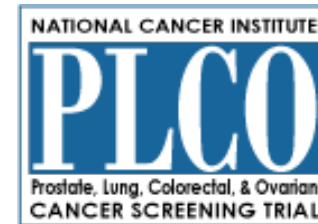
Note

Must propose potential project



<https://dietandhealth.cancer.gov/>

Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial



Population

Adults aged 55-74

Method

Population-based randomized trial design

Content

Determine effects of screening on cancer-related mortality and secondary endpoints in men and women

Biologic materials

Risk factor information

Data

n=78,234 women

n=76,704 men

Note

Initial data collected in 2006; additional data collected through 2015

Must request access to data and outline proposed project through Cancer Data Access System (CDAS)

National Lung Screening Trial



Population

Adults aged 55-74; current or former heavy smokers

Method

Screening trial

Compared two types of detection: 1) CT; 2) X-ray;

Content

Mortality

Risk

Smoking status and behaviors

Data

n=53,454

August 2002 to April 2004

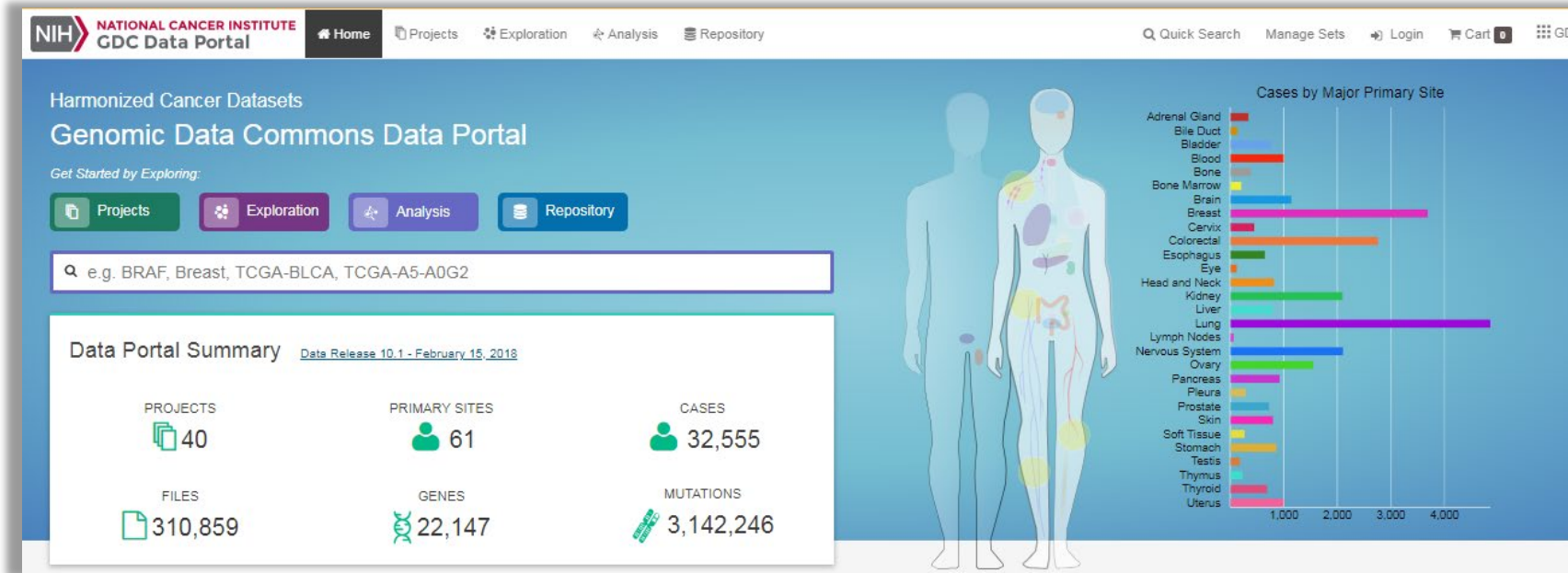
Note

Must request access to data and outline proposed project

<https://cdas.cancer.gov/nlst/>

Data Search/Repositories

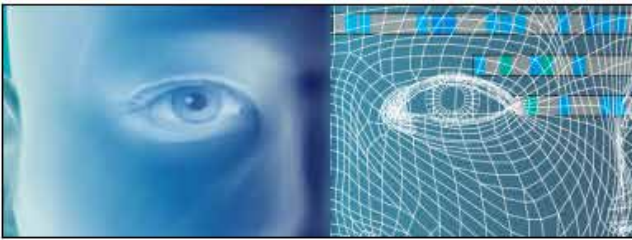
Genomic Data Commons



Available:

The Cancer Genome Atlas ([TCGA](#))

Therapeutically Applicable Research to Generate Effective Treatments ([TARGET](#))



dbGaP

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.

Purpose

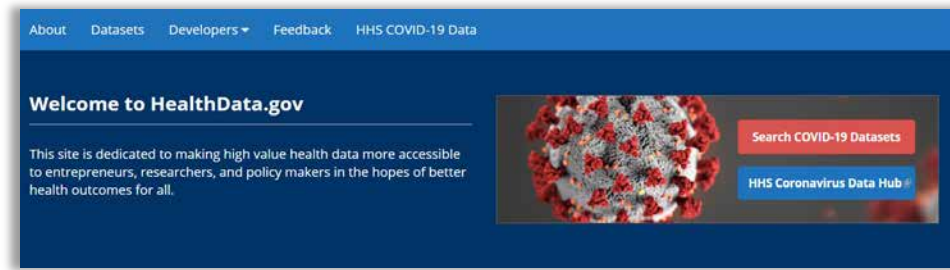
For accessing individual-level phenotype, exposure, genotype, and sequence data, and the associations between them. dbGaP assigns stable, unique identifiers to studies and subsets of information from those studies, including documents, individual phenotypic variables, tables of trait data, sets of genotype data, computed phenotype-genotype associations and groups of study subjects who have given similar consents for use of their data.

Collections

- Individual level genomic data
- Autism-omics studies
- Open translational science in Schizophrenia (OPTICS)

Note

- Data access request necessary
- Includes data browser



Purpose

Makes high value health data more accessible to entrepreneurs, researchers, and policy makers in the hopes of better health outcomes for all

Data Sources

CMS, CDC, FDA, NIH (4,523 datasets)

Dataset Types

Health, State, National, Medicare, Hospital, Quality, Community, Inpatient

Note

Feedback webpage has answers to common questions
Includes blog posts

<https://www.healthdata.gov/>

National Collaborative on Childhood Obesity Research



- Catalogue of Surveillance Systems
 - 100 public data sets
 - Related to childhood obesity research
- Health behaviors, outcomes, and determinants
- Policies and environmental factors
- Measure registry

The screenshot shows the 'Surveillance Systems' page of the NCCOR catalogue. On the left, there are filter options for 'Search' (a text box), 'Level' (checkboxes for Individual: 53, Household: 16, School: 8, Community: 17, Macro/Policy: 13, Other: 17), 'Scope' (checkboxes for Local: 37, State: 41, National: 94), and 'Key Variables' (checkboxes for Diet related: 83, Physical activity related: 51, Weight related: 40, Geocode/Linkage: 83). The main area on the right is titled 'Showing all 106 systems' and contains a table with two columns: 'System Name' and 'Compare'. The table lists 15 surveillance systems, each with a corresponding 'Compare' button.

System Name	Compare
American Community Survey	<input type="checkbox"/>
American Time Use Survey (ATUS)	<input type="checkbox"/>
Annual Agricultural Statistics	<input type="checkbox"/>
Behavioral Risk Factor Surveillance System (BRFSS)	<input type="checkbox"/>
Bridging the Gap State Snack and Soda Tax Data System	<input type="checkbox"/>
California Health Interview Survey (CHIS)	<input type="checkbox"/>
Census TIGER® (Topologically Integrated Geographic Encoding and Referencing) Files	<input type="checkbox"/>
Census Transportation Planning Products (CTPP)	<input type="checkbox"/>
Census of Agriculture (Agcensus)	<input type="checkbox"/>
Center for Nutrition Policy and Promotion Food Prices Database (CNPP/FDP)	<input type="checkbox"/>
Classification of Laws Associated with School Students (C.L.A.S.S.)	<input type="checkbox"/>
ComScore	<input type="checkbox"/>
Consumer Expenditure Survey (CE)	<input type="checkbox"/>
Consumer Reports on Eating Share Trends (CREST)	<input type="checkbox"/>
Cost of Living Index (COLI)	<input type="checkbox"/>
CropScape	<input type="checkbox"/>
Current Population Survey (CPS)	<input type="checkbox"/>

- CEDCD gives info about cohort studies that follow groups of persons over time for cancer incidence, mortality, and other health outcomes
- Searchable database with general study info
- No individual level data; cohort data only
- Can search for enrollment counts, cancer counts, and biospecimen counts
- 46 cohorts representing over 6.5 million people





Purpose

An international consortium of more than 750 academic institutions and research organizations, Inter-university Consortium for Political and Social Research (ICPSR) provides leadership and training in data access, curation, and methods of analysis for the social science research community

Data Sources

Data archive of more than 250,000 files of research in the social and behavioral sciences. It hosts 21 specialized collections of data in education, aging, criminal justice, substance abuse, terrorism, and other fields

Access

<https://www.icpsr.umich.edu/icpsrweb/>

Note

Can see collection of articles, books, reports, and other materials that have used ICPSR data

<https://www.icpsr.umich.edu/web/pages/>

Purpose

Provides maps, data, reports and dashboard to the public. The data integrates with external sources, such as the U.S. Census Bureau, providing information about HRSA's grants, loan and scholarship programs, health centers and other public health programs and services

Data Sources

The HDW manages data from more than 30 HRSA and other federal government sources

Content

- Health centers
- Health workforce
- Healthcare systems
- Rural health

Note

National/State/County-level data
Area Health Resource files, latest data 2020



<https://data.hrsa.gov/>

NIH Data Sharing Repositories

Open Domain-Specific Data Sharing Repositories:

https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html

Including:

- Genomic data commons
- Proteomic data commons

Other NIH Data Resources:

https://www.nlm.nih.gov/NIHbmic/other_data_resources.html

Including:

- Science of Behavior Change (SOBC) repository
- The Cancer Target Discovery and Development (CTD2) Network Data Portal

UC Berkeley Library



3 libraries are now open! Plus, UC Library Search is here!

ASK US LOG IN TO YOUR LIBRARY ACCOUNT HOURS AND MAPS CONNECT FROM OFF CAMPUS UC BERKELEY HOME

Berkeley Library
UNIVERSITY OF CALIFORNIA

UC Berkeley / Library Guides / Bioscience, Natural Resources & Public Health Library / Health Statistics & Data / Datasets/Raw Data

Health Statistics & Data: Datasets/Raw Data

Search 2016-2018

National & Int'l Statistics State & Local Statistics Statistics by Population Group Statistics by Health Topic

Datasets/Raw Data Using Statistics GIS/Mapping Tools Need More Help?

National Datasets

- **Berkeley Mortality Database**
"This database contains life tables for national populations and, whenever available, the raw data used in constructing these tables." Data is provided for the United States, France, Japan, and Sweden. Note: The BMD has been replaced by the Human Mortality Database (see below for link). HMD data are superior for most purposes, however some items within the BMD are not yet available within the HMD or elsewhere.
- **Child Health and Development Studies (CHDS)**
The CHDS researches the way that health and disease are passed on in families, looking at genes as well as social, personal, and environmental surroundings. Researchers may apply to use the data files. CHDS is a project of the Public Health Institute and is supported with funds from the National Institutes of Health.

California Datasets

- **California Health Interview Survey**
The largest state health survey in the United States. It is a telephone survey conducted every two years on public health topics and access to health care. The web site contains the questionnaires, data files, and AskCHIS, an online data query system. (Requires one-time, free registration).
- **California Office of Statewide Health Planning and Development**
Has data available for viewing and downloading such as Hospital Annual Financial Data, Hospital Quarterly Financial Data, Utilization Data, Long-term Care Facility Financial Data, Patient Discharge Data, etc.
- **California Open Data Portal**
Provides a single source of California-generated raw data, including many health and environmental data, as well as economic and

Links to datasets:

- National
- California-based
- International

Note: Also contains links to statistical and GIS-related resources

<https://guides.lib.berkeley.edu/publichealth/healthstatistics/rawdata>

Google Dataset Search

<https://datasetsearch.research.google.com/>



Dataset Search

Search for Datasets



Try [coronavirus covid-19](#) or [global temperatures](#).

[Learn more](#) about including your datasets in Dataset Search.

The screenshot shows a search result for the 'National Health Interview Survey (NHIS) - National Cardiovascular Disease Surveillance Data'. At the top is the CDC logo. Below it, the title 'National Health Interview Survey (NHIS) - National Cardiovascular Disease Surveillance Data' is displayed. There are four blue buttons with white text: 'Explore at catalog.data.gov', 'Explore at Open Data | Centers for Dis...', 'Explore at AmeriGEOSS Community Platfo...', and 'Explore at data.wu.ac.at'. Below the buttons, it says 'Dataset updated Jul 22, 2020'. Then, 'Dataset provided by' is followed by a link to 'Centers for Disease Control and Prevention'. A 'Description' section follows, containing a paragraph about the NHIS survey, its history since 1957, and the types of data collected and analyzed.

CDC
Centers for Disease Control and Prevention

National Health Interview Survey (NHIS) - National Cardiovascular Disease Surveillance Data

[Explore at catalog.data.gov](#) [Explore at Open Data | Centers for Dis...](#) [Explore at AmeriGEOSS Community Platfo...](#)
[Explore at data.wu.ac.at](#)

Dataset updated Jul 22, 2020

Dataset provided by
[Centers for Disease Control and Prevention](#)

Description

2001 forward. The National Health Interview Survey (NHIS) has monitored the health of the nation since 1957. NHIS data on a broad range of health topics are collected through personal household interviews. Indicators for this dataset has been computed by personnel in CDC's Division for Heart Disease and Stroke Prevention (DHDSP). This is one of the datasets provided by the National Cardiovascular Disease Surveillance System. The system is designed to integrate multiple indicators from many data sources to provide a comprehensive picture of the public health burden of CVDs and associated risk factors in the United States. The data are organized by location (region) and indicator, and they include CVDs (e.g., heart failure) and risk factors (e.g., hypertension). The data can be plotted as trends and stratified by age group, sex, and race/ethnicity.

re3data.org: Registry of Repositories



- Can search by keyword or browse repositories by subject, content type, or country

Inter-university Consortium for Political and Social Research

General Institutions Terms Standards

Name of repository	Inter-university Consortium for Political and Social Research
Additional name(s)	ICPSR data archive
Repository URL	https://www.icpsr.umich.edu/web/pages/index.html
Subject(s)	Social and Behavioural Sciences Humanities and Social Sciences Life Sciences Education Sciences Empirical Social Research Political Science Social Sciences Economics Criminology Psychology Jurisprudence
Description	ICPSR maintains a data archive of more than 250,000 files of research in the social and behavioral sciences. It hosts 21 specialized collections of data in education, aging, criminal justice, substance abuse, terrorism, and other fields. ICPSR advances and expands social and behavioral research, acting as a global leader in data stewardship and providing rich data resources and responsive educational opportunities for present and future generations.
Contact	mshukait@umich.edu
Content type(s)	Standard office documents Structured graphics Scientific and statistical data formats Plain text Structured text Archived data

Data Collection Tools

Crowd-Sourced Research Participants

- Online labor market
- Primary data collection
- Fast and relatively cheap
- Large samples
- Can randomize respondents
- Longitudinal designs
- Representativeness may be concern
- Examples:
 - Mechanical Turk (Mturk)
 - Prolific



<https://www.mturk.com/>



<https://www.prolific.co/>

- Multiple uses:
- Image processing
- Data verification/clean-up
- Surveys....

On-line Panels/Survey Tools

- Oftentimes used for marketing
- Dedicated, paid users
- Primary data collection
 - Provides sample and/or
 - Data collection tool
- Survey work
 - Probability and non-probability based
- Representativeness may be concern
- Comparison studies done



Research Methods/Data Analysis Tools

Example Statistical Trainings



Free SAS® e-Learning
Training wherever you are.

Free e-learning to get you started

- Free how-to SAS tutorials
- SAS Programming 1
- Statistics 1
- SAS Programming for R Users
- SAS Platform Administration: Getting Started
- SAS Viya Enablement

Start Now

Start Now

Start Now

Start Now

Start Now

https://www.sas.com/en_us/learn/academic-programs/resources/free-sas-e-learning.html



<https://statisticalhorizons.com/>

Machine Learning

Remote Seminar

Kevin Grimm, Instructor

August 11-14

Sold Out

Structural Equation Modeling:

Part I

Remote Seminar

Paul Allison, Instructor

August 11-14

Sold Out

Mediation, Moderation, and Conditional Process Analysis: A Second Course On Demand

Andrew Hayes, Instructor

August 24-September 21

Propensity Score Analysis

Remote Seminar

Shenyang Guo, Instructor

September 10-12

Multilevel and Mixed Models Using Stata

Remote Seminar

Stephen Vaisey, Instructor

September 24-26

Latent Growth Curve Modeling

Remote Seminar

Gregory Hancock, Instructor

October 1-3

Introduction to Python for Data Analysis*

Remote Seminar

Jason Anastasopoulos, Instructor

October 1-3

Analysis of Complex Survey Data

Remote Seminar

Ann O'Connell, Instructor

October 8-10

Meta-Analysis

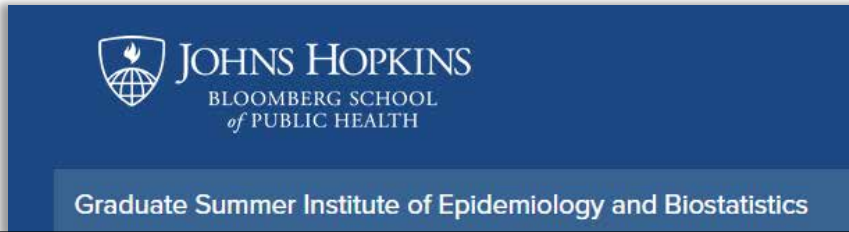
Stats/Research Methods Trainings



<http://crmdaclone.drupal.ku.edu/>



<https://si.isr.umich.edu/>



<https://www.jhsph.edu/departments/epidemiology/continuing-education/graduate-summer-institute-of-epidemiology-and-biostatistics/index.html>



<https://www.khanacademy.org/>



<https://methodology.psu.edu/>



<https://datic.uconn.edu/>

Analytic Support



Institute for Digital Research & Education
Statistical Consulting

<https://stats.idre.ucla.edu/>

- Statistical software:
 - Syntax
 - Output
- Choosing the correct statistical test
- Trainings

UPCOMING EVENTS

Remote consulting closed

July 5

Intermediate Topics in Confirmatory Factor Analysis (CFA)

August 16 @ 1:00 pm - 4:00 pm

Statistical Writing

August 23 @ 1:00 pm - 4:00 pm

[View All Events](#)

ANNOTATED OUTPUT

These pages contain example programs and output with footnotes explaining the meaning of the output. This is to help you more effectively read the output that you obtain and be able to give accurate interpretations.

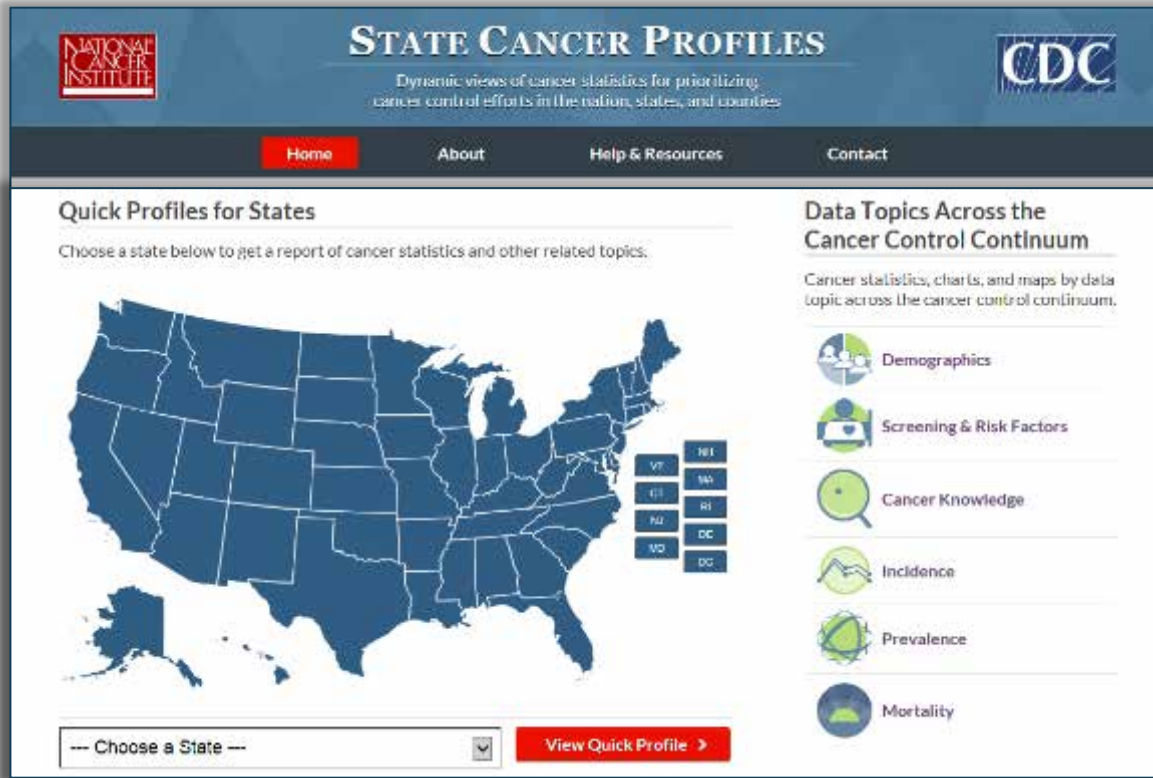
	Stata	SAS	SPSS	Mplus
Descriptive Statistics				
Descriptive Statistics	Stata	SAS	SPSS	

DATA ANALYSIS EXAMPLES

	Stata	SAS	SPSS	Mplus	R
Regression Models					
Robust Regression	Stata	SAS			R
Models for Binary and Categorical Outcomes					
Logistic Regression	Stata	SAS	SPSS	Mplus	R

Data Visualization Tools

The State Cancer Profiles Website



- A collaborative effort between NCI and CDC
- Part of Cancer Control P.L.A.N.E.T.
- Cancer-related incidence, prevalence and mortality
- Risk behaviors
- County and/or state-level

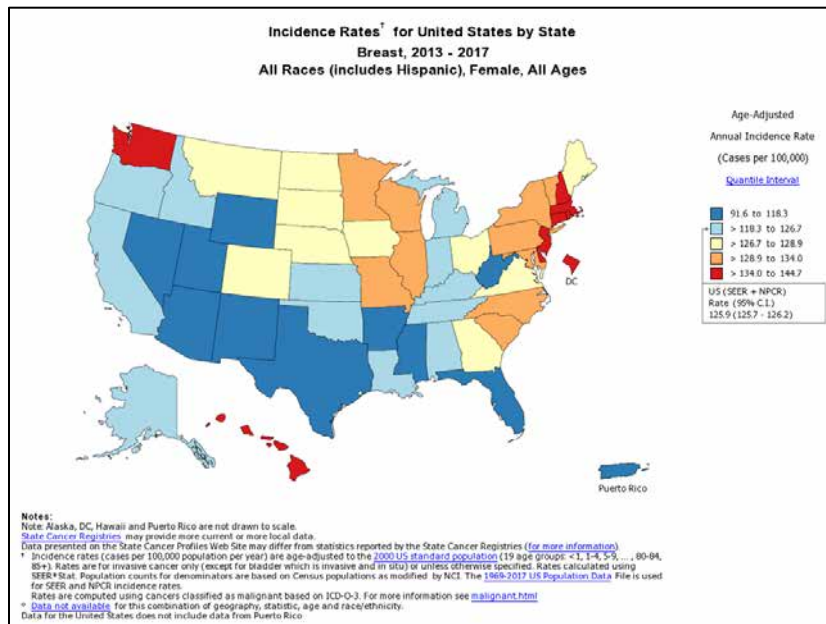
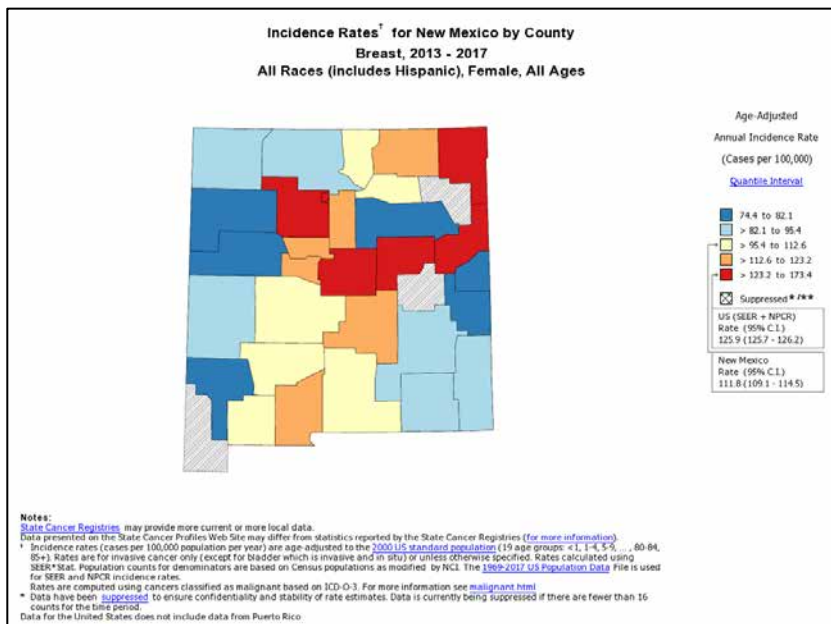
Slides provided by Zaria Tatalovich

<https://StateCancerProfiles.cancer.gov/>

Available Presentation Types by Data Topic

	 Demographics	 Screening and Risk Factors	 Cancer Knowledge	 Incidence	 Prevalence	 Mortality
Interactive maps						
Tables						
Historical trends						
5-year rate changes						
Rate-trend comparison by cancer						
Rate-trend comparison by area						

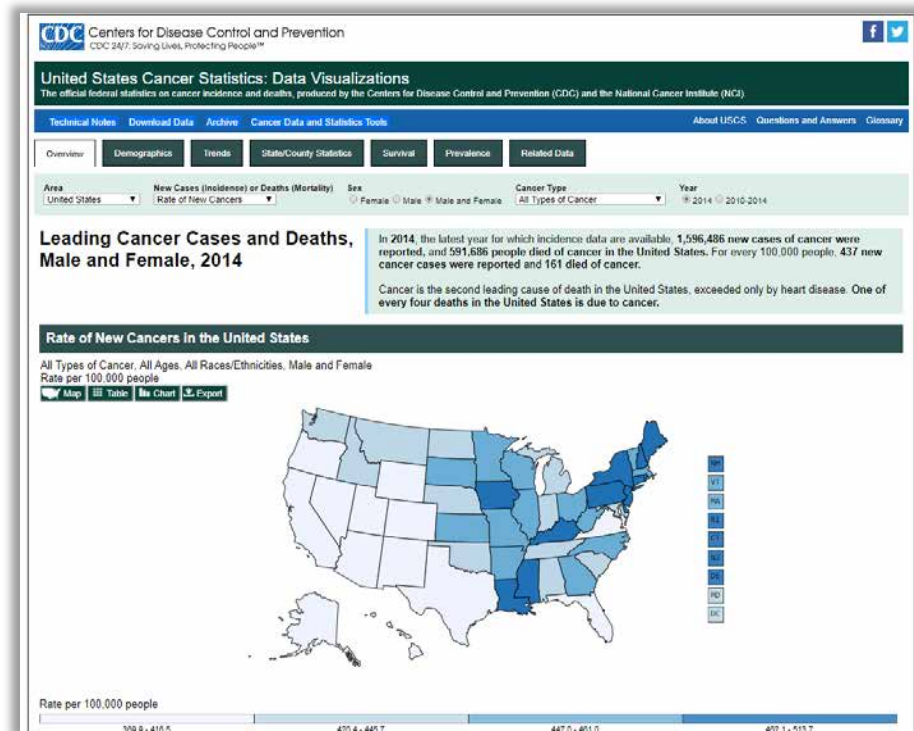
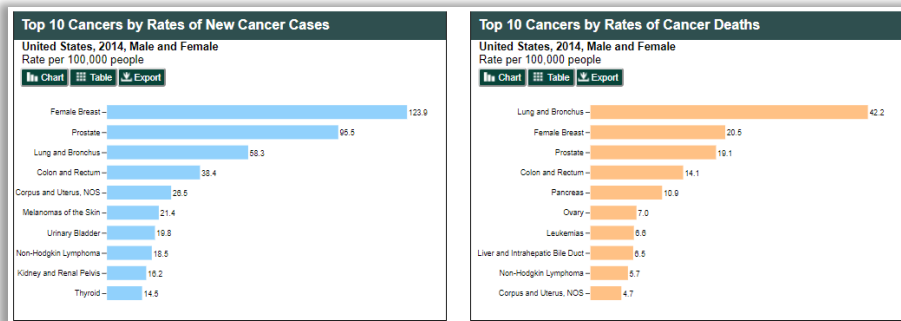
Breast Cancer Incidence in New Mexico and US by State



CDC: U.S. Cancer Statistics: Data and Data Visualizations

- Uses SEER and CDC cancer registry data that covers 100% of population
- 31 million cases (2001-2018)
- Can download data
- State/county statistics
- Incidence/prevalence/mortality
- Explore trends

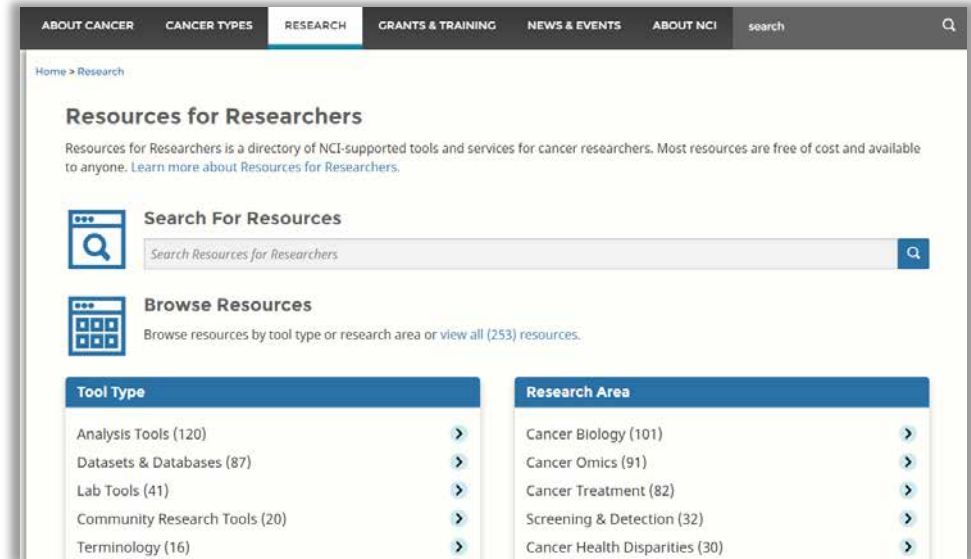
<https://www.cdc.gov/cancer/uscs/index.htm>



Other Resources

NCI Resources for Researchers

- Directory of NCI-supported tools and services for cancer researchers.
- Data, terminologies, analysis tools, including:
 - Genomic data commons
 - HINTS
 - NCI Meta-thesaurus
- Most are free



<https://www.cancer.gov/research/resources>

Bibliographies for Selected Datasets

- NHIS

- https://www.cdc.gov/nchs/nhis/nhis_bibliography.htm

- NHANES

- <https://www.cdc.gov/nchs/products/databriefs.htm?program=nhanes>

- TUS-CPS

- <https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/publications-database>

- SEER

- <https://seer.cancer.gov/publications/>

Upcoming Webinars

- Deep Dive into Available Data Resources for Specific Cancer Prevention and Control Topics – Part 1

December 9, 2021

1:00 – 3:00 p.m. ET

- Deep Dive into Available Data Resources for Specific Cancer Prevention and Control Topics – Part 2

January 12, 2022

3:00 – 5:00 p.m. ET

Webinar recordings will be available 3-4 weeks after the event:

<https://cancercontrol.cancer.gov/brp/events/tools-resources-support-cancer-prevention>



**NATIONAL
CANCER
INSTITUTE**

www.cancer.gov

www.cancer.gov/espanol