The UCSF Population Health Data Initiative:
New Research Resources

Rita Hamad, MD, PhD
PHDI Faculty Lead for Research Resources
Philip R. Lee Institute for Health Policy Studies
Department of Family & Community Medicine
The Population Health Data Initiative

PHDI supports efforts to provide data and other research resources for population health, health equity, and health services research at UCSF.
Cataloguing data resources

Facilitating data access and licensing

Enhancing computational resources
A Guide to Population Health Data

**The Problem:** Population health, health equity, and health services datasets are hard to find!

**The Team:** Collaboration between PHDI and the UCSF Library
Population health researchers seek to understand how biological, behavioral, environmental, social, and structural factors affect the health of populations. They also study how to design interventions to improve health and reduce health disparities at the population level.

This guide catalogues public and private population health and health services data sets for research in population health, health services research and health equity.
Population Health and Health Services Research Datasets

Administrative

- All Payor Claims Database
  Patient claim data for insured patients only
- Center for Consumer Information and Insurance Oversight - $
  Affordable Care Act datasets
- Chronic Conditions Data Warehouse - $
  Medicare and Medicaid beneficiary, claims and assessment data across the continuum of care
- Dartmouth Atlas Project - $
  Metrics quantifying the temporal and regional patterns for healthcare spending and utilization
- Health and Medical Care Archive - $
  Health department profiles, cost and access to health care, substance abuse and chronic health conditions
- Healthcare Cost and Utilization Project (HCUP) - $
  Hospital Care Data. UCSF researchers interested in this dataset should request a consultation with CTSI
- Health Professionals Shortage Areas
  County level data on health staffing shortages
- Health Resources and Services Administration Data Portal
  Data for improving healthcare to people who are geographically isolated, economically or medically vulnerable.
Want help accessing/using data?

- Check out the guide at guides.ucsf.edu/pophealthdata

- To learn more about study design, data selection, or data acquisition – request a CTSI population health data consult

- To suggest a dataset to add to the guide, or get help managing or analyzing data – request a library data science consultation
The Problem: “I used the data guide and/or found a great dataset, but how do I get access to it?”
Standard Operating Procedure: New Data

Usual approach to getting access to a new dataset:

- Submit an inquiry with the data owner
- Figure out:
  - How to store data
  - A data access and security plan
  - How to pay $$$ for the data
- Convince the data owner your team will responsibly use the data
- Sign data use agreement
- Wait for contracts, purchasing, data delivery
- …
Why Repeat Efforts?

Usual approach is necessary for truly new datasets, but...

... someone else on campus has probably done the hard work for many datasets!

Why pay more money and sign new contracts for data that another UCSF researcher has already paid for...?
The PHDI Solution

PHDI aims to solve this by providing:
- A centralized source for **finding** new datasets
- A simplified process for **getting access** to data
- UCSF **site licenses** for commonly used datasets
- Efforts to **standardize** data use agreements
Data Access Solution

- We have launched a pilot of our new data access process!
  → access to IBM MarketScan Research Data
    (formerly known as TruvenHealth)
- Open to all researchers at UCSF
- Free to access
  - Free to publish for unfunded/intramurally funded studies
  - Publication fee per study for extramurally funded projects
IBM MarketScan Claims Data

- IBM MarketScan Research Databases provide de-identified enrollment, inpatient, outpatient, and outpatient pharmacy commercial claims data
  - Over 50 million covered lives across the US
  - Fully longitudinal at the patient level from 2010-2018
  - Useful for cost, utilization, and epidemiological studies
  - Large enough to be representative of the US commercial insurance population
Phase 1: Data Access Model

- PHDI has purchased a site license for research studies
- Providing streamlined access approval process
- Improvements to data storage and compute infrastructure
- Some programming support to help researchers make initial cuts of the data and consult on data use
Phase 2: Scaling our Data Access Model

- Using this pilot process as the prototype for a new population health data access model
- Phase 2 will refine workflow and include other popular data:
  - Medicare claims sample
  - AHRQ HCUP Hospital Discharge Abstracts (e.g., SID, SEDD, NIS, KID)
  - American Hospital Association Surveys
Next Steps – How do I get more details?

- If you have a research project that would benefit from the use of IBM MarketScan data or any of these other data sets:
  - Email pophealth@ucsf.edu
  - Visit https://pophealth.ucsf.edu
The Berkeley Restricted Data Center

Provides access to dozens of governmental surveys and administrative data sets
• Geographic Codes for all NCHS Surveys
• National Health and Nutrition Examination Survey (NHANES)
• National Health Care Surveys
  ▪ National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS)
  ▪ National Hospital Discharge Survey (NHDS)
  ▪ National Nursing Home Survey (NNHS) and National Nursing Assistant Survey (NNAS)
  ▪ National Home and Hospice Care Survey (NHHCS) and National Home Health Aide Survey (NHHAS)
  ▪ National Survey of Residential Care Facilities (NSRCF)
  ▪ National Study of Long-Term Care Providers (NSLTCP)
  ▪ National Hospital Care Survey (NHCS)
• National Health Interview Survey (NHIS)
• National Survey of Family Growth (NSFG)
• State and Local Area Integrated Telephone Survey (SLAITS)
  ▪ National Survey of Children's Health (NSCH)
  ▪ National Survey of Children with Special Health Care Needs (CSHCN)
• NCHS Data Linkage Activities
  ▪ Linked Mortality Data Products
  ▪ Linked Medicare Enrollment and Claims Files Data
  ▪ Linked Medicaid Enrollment and Claims Data
  ▪ Linked Social Security Benefit History Data
  ▪ Linked Housing Assistance Program Files Data
• National Vital Statistics System (NVSS) Data Release and Access Policy
• National Maternal and Infant Health Survey
• Drug Involved Mortality Restricted Variables
• Redacted Death Certificate Literal Text File
Find out more and apply for data access:

https://www.cdc.gov/rdc
The Berkeley Restricted Data Center

Standard RDC access fees: $7,500 per project per year

For UCSF researchers: FREE

To find out more, email pophealth@ucsf.edu or request a CTSI population health consultation:

https://consult.ucsf.edu/phhs
A Guide to Population Health Compute

The Problem: “I have my dataset, I’m ready for analysis… how do I get started?”
Standard Operating Procedure: Compute

Usual approach to finding a compute environment for data analysis:

- Contact the MyResearch team…
  - but your data set is too large for the platform

- Contact UCSF IT data center…
  - but you don’t have the time or money to bring up a dedicated server environment

- Consider cloud computing…
  - but you don’t have secure cloud computing skills
The PHDI Solution

PHDI and Academic Research Systems (ARS) solution:
- Expanded compute power of MyResearch to support (much) larger data sets
- Ensuring large data computational tools (Stata, R) are available

Available Now!

Need to upgrade your MyResearch account? Need a new account?
-- submit a Service Now ticket

Questions about research computing? -- email SomTechManagers@ucsf.edu
Conclusion: Contact Us!

PHDI is working to improve access to population health research resources.

We are open to feedback on how to maximize the usefulness of these resources for the UCSF community.

**Please email us if you have data you think could be used by others on campus!**
Acknowledgements

PHDI Faculty Leads:
Kirsten Bibbins-Domingo, Vice Dean for Population Health and Health Equity
Claire Brindis, Director of Philip R. Lee Institute for Health Policy Studies

PHDI Team Members:
Robert Thombley, Data architect
Kristin Chu, Director SOMTech
Ariel Deardorff, Data science librarian

Financial Support:
Dan Lowenstein, Executive Vice Chancellor and Provost
Talmadge King, Dean, UCSF School of Medicine
Thank you! Questions?

rita.hamad@ucsf.edu

pophealth@ucsf.edu

https://pophealth.ucsf.edu