NEBRASKA ACADEMY FOR
METHODOLOGY, ANALYTICS & PSYCHOMETRICS

Navigating Research Data Repositories to Support Secondary Analysis Research

HyeonJin Yoon, PhD

Nebraska Methodology Application Series
March 28, 2025

Overview

- Introduction
- Accessing Data Sets
- Using Data Sets
- Grant and Training Opportunities
- Grant Tips
- Additional Resources

Introduction

MAP ACADEMY

Introduction

- Three broad purposes of research (Babbie, 2020)
  - Explore: Gain an initial, rough understanding of a phenomenon when little is known.
  - Describe: Define and characterize the phenomenon. Identify patterns or trends within the subject.
  - Explain: Discover and report relationships among different aspects of the phenomenon. Answer the questions “What’s happening?” and “Why?”
- Primary research: Researchers design studies and collect original data
- Secondary research: Researchers analyze existing data archives.

Babbie, E. (2020). *The practice of social research* (15th ed.). Cengage Learning.

Secondary Analysis Research (SAR)

- Uses pre-exiting datasets to address new questions
- Requires a broad knowledge based and up-to-date understanding of the field
- Most SAR use quantitative data, but some qualitative data (e.g., interviews, open-ended survey responses) can also be used.

Why Secondary Analysis Research?

- Many existing datasets can be used to address important and interesting new research questions
  - Make strategic use of existing data to address a diverse range of topics
  - Can be combined with primary data collection studies to demonstrate generalizability and external validity

- Saves Time and $$
  - Data have already been collected
  - Large with high response rates and collected using sound scientific sampling procedures
  - Researchers can have access study difficult-to-reach populations and explore under-researched areas

- “Open source” approach to science

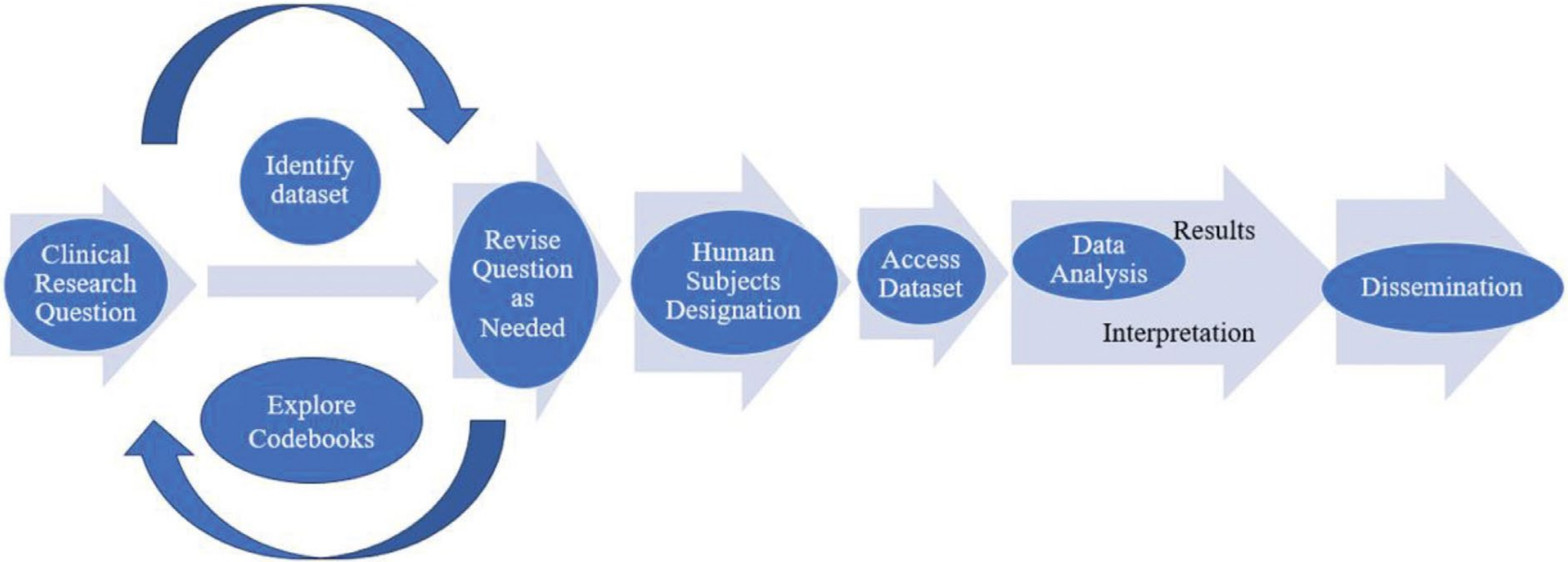
- A way to publish without a senior author

- Fundable: many institutions (Dept. of Ed; NIH; NSF) have calls for secondary analysis grants

Disadvantages

- Breadth vs. Depth
  - May ask about lots of topics; but may rely on single items to capture constructs
  - Representative sampling strategies; but may include limited numbers of subgroup populations
- Lack of control: Researchers have no control over study population, variables or study design
- Measurement issues and concerns
  - Potentially rely on single survey items vs. full measures
- Considerable investment of time to learn how to analyze data sets

Steps in Secondary Analysis Research



(Kelly et al., 2024)

Kelly, M. M., Martin-Peters, T., & Farber, J. S. (2024). Secondary Data Analysis: Using existing data to answer new questions. *Journal of Pediatric Health Care*, *38*(4), 615-618.

MAP ACADEMY

Accessing Data Sets

MAP ACADEMY

Types of Data Sets

- Federal vs. Researcher Uploaded
  - Federal agencies collected and uploaded: e.g., National Center for Educational Statistics (NCES) (e.g., High School Longitudinal Study, Schools and Staffing Survey)
  - Researcher Uploaded: E.g., ICPSR data repository

- Public Use vs. Restricted Use
  - Public: disclosure risk to research participants is minimal
  - Restricted: retain confidential data but requires controlled conditions for accessing them

Find Relevant Data Sets

- Read the literature and learn about existing data sets

- Search for metadata (i.e., data about datasets) such as access policies, sampling design, variables, existing publications

Finding Data Sets

- Search portals
  - University of Michigan's Inter-University Consortium for Political and Social Research (ICPSR): http://www.icpsr.umich.edu/
  - Institute of Education Sciences (National Center for Education Statistics): https://nces.ed.gov/pubsearch/licenses.asp
  - https://nces.ed.gov/datalab/
  - DATA.GOV https://data.gov/
  - ODUM INSTITUTE DATA ARCHIVE: https://odum.unc.edu/archive/
  - Substance Abuse and Mental Health Services Administration (SAMHSA): https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health
  - Roper Center of Public Opinion Research at the University of Connecticut: http://www.ropercenter.uconn.edu/
  - National Archive of Computerized Data on Aging: http://www.icpsr.umich.edu/icpsrweb/NACDA/

ICPSR

Help | Log In

Find data, studies, variables, and related publications

Home | Find Data | Share Data | Membership | Summer Program | Teaching & Learning | Data Management | About

Find Data

Search | view all

search tips

Browse

Topics / Series / Thematic data collections / Data-related publications

- Studies with online analysis
- Self-published data, including replication datasets
- Studies with learning guides

Statistics

21,599 studies

6,770,151 variables

Most Popular Search Terms

https://www.icpsr.umich.edu/web/pages/ICPSR/index.html

MAP ACADEMY

Home | Find Data | Share Data | Membership | Summer Program | Teaching & Learning | Data Management | About

The Expanding Children's Early Learning (ExCEL) Quality Study: Improving Preschool and Early Care and Education (ECE) Instruction through Curricula, Coaching, and Training, United States, 2019-2021 (ICPSR 39222)

Version Date: Nov 6, 2024 Cite this study | Share this page

Principal Investigator(s):
Michelle Maier, MDRC

Series:

- MDRC Series

https://doi.org/10.3886/ICPSR39222.v1

Version V1

Analyze Online (SDA) | Download | Access Restricted Data

At A Glance | Data & Documentation | Variables | Data-related Publications | Export Metadata

281 | 0

https://www.icpsr.umich.edu/web/ICPSR/studies/39222

MAP ACADEMY

Active for Life: Translation of Physical Activity Programs for Mid-Life and Older Adults, 2003-2007 [United States] (ICPSR 24723)

Version Date: Feb 14, 2024 Cite this study | Share this page

Principal Investigator(s):
Sara Wilcox, University of South Carolina

https://doi.org/10.3886/ICPSR24723.v3

Version V3 (see more versions)

Analyze Online (SDA) | Download | Access Restricted Data

At A Glance | Data & Documentation | Variables | Data-related Publications | Export Metadata

Project Description

Alternate Title

Evaluation of Active for Life: Increasing Physical Activity Levels in Adults Age 50 and Over, 2003-2007

731 Downloads * Usage Report
* past three years

21 Data-related Publications

Notes

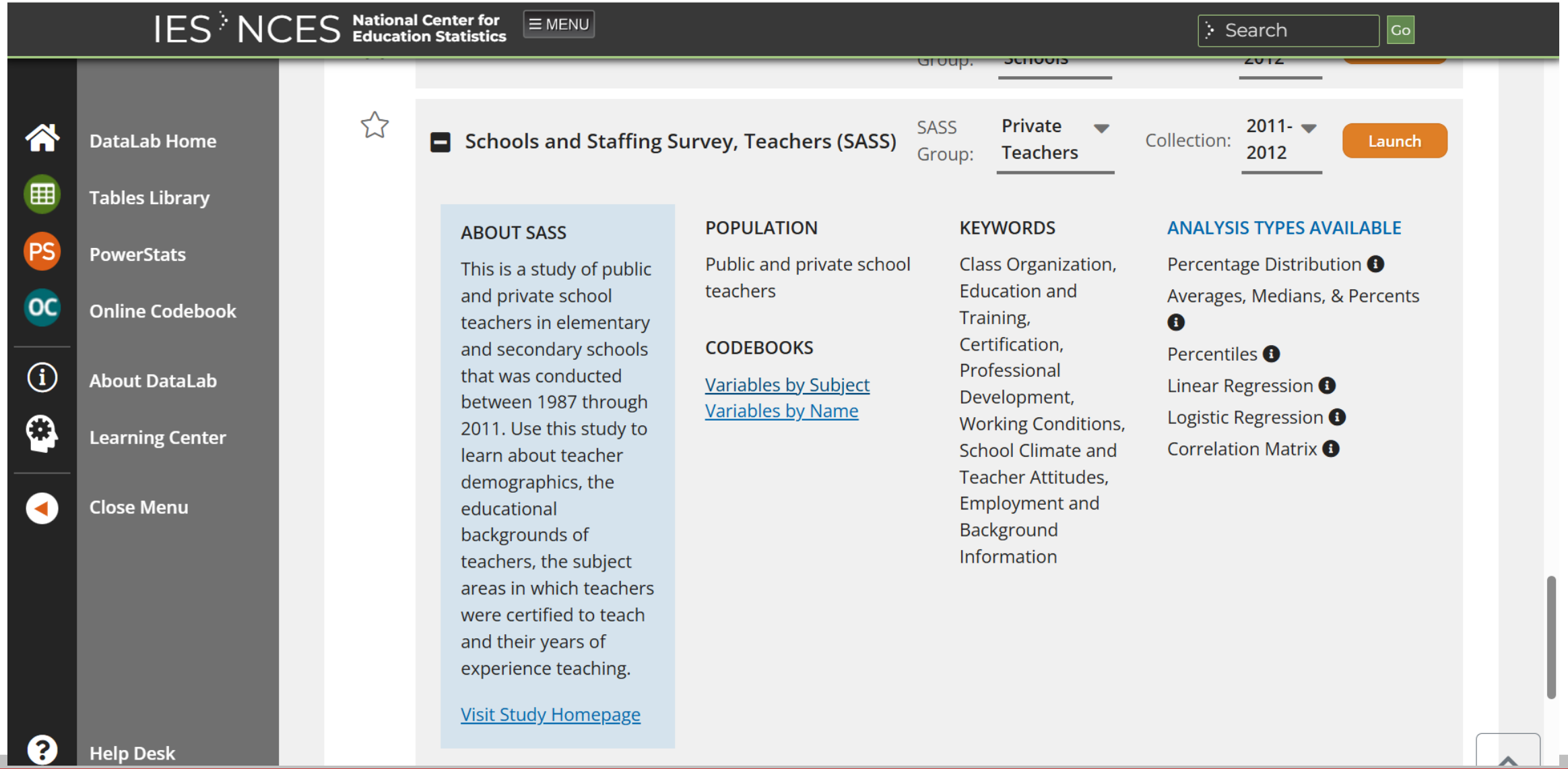
- The public-use data files in this collection are available for access by the general

https://www.icpsr.umich.edu/web/ICPSR/studies/24723/summary

MAP ACADEMY

https://nces.ed.gov/datalab/

MAP ACADEMY


IES NCES National Center for Education Statistics
MENU
Search
Go
DataLab Home
Tables Library
PowerStats
Online Codebook
About DataLab
Learning Center
Close Menu
Help Desk
Schools and Staffing Survey, Teachers (SASS)
SASS Group:
Private Teachers
Collection:
2011-2012
Launch
ABOUT SASS
This is a study of public and private school teachers in elementary and secondary schools that was conducted between 1987 through 2011. Use this study to learn about teacher demographics, the educational backgrounds of teachers, the subject areas in which teachers were certified to teach and their years of experience teaching.
Visit Study Homepage
POPULATION
Public and private school teachers
CODEBOOKS
Variables by Subject
Variables by Name
KEYWORDS
Class Organization, Education and Training, Certification, Professional Development, Working Conditions, School Climate and Teacher Attitudes, Employment and Background Information
ANALYSIS TYPES AVAILABLE
Percentage Distribution
Averages, Medians, & Percents
Percentiles
Linear Regression
Logistic Regression
Correlation Matrix

Schools and Staffing Survey (SASS)

Publications & Products | **Data Products** | **Staff**

- SASS Homepage
- Overview
- Questionnaires
- Methods & Procedures
- Research Issues
- Search SASS/TFS Table Library
- Online Bibliography Search Tool
- DataLab
- Join NewsFlash
- Contact

Related Study

National Teacher and Principal Survey

The Schools and Staffing Survey (SASS) was conducted by NCES seven times between 1987 through 2011. SASS was an integrated study public and private school districts, schools, principals, and teachers designed to provide descriptive data on the context of elementary and secondary education. SASS covered a wide range of topics from teacher demand, teacher and principal characteristics, general conditions in schools, principals' and teachers' perceptions of school climate and problems in their schools, teacher compensation, district hiring and retention practices, to basic characteristics of the student population. After 2010–11, NCES redesigned SASS and named it the National Teacher and Principal Survey (NTPS) to reflect the redesigned study's focus on the teacher and principal labor market and on the state of K-12 school staff. NCES first conducted NTPS in 2015–16. (Learn more about NTPS here.)

https://nces.ed.gov/surveys/sass/

MAP ACADEMY

UNL Resources

- Central Plains Federal Statistical Research Data Center (CPRDC)

https://business.unl.edu/research/central-plains-federal-statistical-research-data-center/

- Houses restricted use data files
  - Census Demographic Data
  - Census Economic Data Linked Business and Household Data
  - Public Health Data

Example Data Sets

- IES Early Childhood Longitudinal Studies (ECLS): https://nces.ed.gov/ecls/
- The National Educational Longitudinal Surveys. For example:
  - http://nces.ed.gov/surveys/hsb/
  - http://nces.ed.gov/surveys/nels88/
- The National Longitudinal Study of Adolescent to Adult Health (Add Health): https://dataverse.unc.edu/dataverse/addhealth
- Family and Child Experience Study (FACES) https://www.icpsr.umich.edu/web/ICPSR/series/236
- The National Longitudinal Survey of Youth: http://www.bls.gov/nls/
- Monitoring the Future: http://www.monitoringthefuture.org/
- Fragile Families & Child Wellbeing Study: https://fragilefamilies.princeton.edu/documentation
- National Survey of Family Growth: http://www.cdc.gov/nchs/nsfg.htm
- The Youth Risk Behavior Surveillance System: http://www.cdc.gov/healthyyouth/yrbs/data/index.htm
- NIH Adolescent Brain Cognitive Development (ABCD) Study: https://abcdstudy.org/
- National Survey on Drug Use and Health (NSDUH) https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/datafiles

Restricted Use Licenses

- Data are restricted and require an application process
  - Agree to follow strict legal and electronic requirements for maintaining data confidentiality
- Example Requirements
  - Investigator and research staff information
  - Research description
  - Confidential data security plan
  - Restricted data use agreement
  - IRB approval
  - Payment & Time
  - Submission and/or review of products

Security Plans

- Goal: to ensure that the restricted data are stored securely and are accessible only to the people listed in the application

- Securing data
  - External hard drive
  - Non-networked computer
  - Local virtual or physical enclave on an isolated network
  - High security location / secure room

Using Data Sets

MAP ACADEMY

Study the Dataset

- User's manual/Technical manuals
- Training materials/videos
  - https://nces.ed.gov/training/datauser/#/
- Codebook
- Interview protocols and questionnaires
- Website (errors, updates, new waves of data)

https://nces.ed.gov/ecls/dataproducts.asp#K-5
https://nces.ed.gov/surveys/sass/

Key Considerations for Data Analysis

- How were variables measured – does it fit your intended research questions and analysis?
- What are the recommended methods for handling complex sampling?
- How were composite variables constructed?
- How were missing data handled and how are missing values coded?
- Did certain questions/measures change over time?
- Did respondents change over time?
- Do variables need to be recoded to account for missing data, skip logic, reverse-scoring, different scales/units?
- Do variables need to be combined?

Document Your Study

- Create an abridged codebook and data set that are specific to your project
- Use syntax for all steps
- Some funding agencies require you to share the data and/or syntax

Use Appropriate Analytic Methods

- Sampling weights ensure that the results generalize to the target population
- Alternative variance estimation methods adjust standard errors to account for complexities such as clustering, stratification, and sampling without replacement from a finite population
- Appropriate use of plausible values accounts for measurement uncertainty and make inferences about population parameters.

Grant and Training Opportunities

MAP ACADEMY

Grant Opportunities

- IES Exploration Grants
  - https://ies.ed.gov/funding/20rfas.asp
- AERA-NSF Grants Program
  - http://www.aera.net/Professional-Opportunities-Funding/AERA-Funding-Opportunities/Grants-Program
- National Institute of Health
  - https://grants.nih.gov/grants/guide/rfa-files/RFA-DA-26-055.html
- Health Resources and Services Administration (Maternal and Child Health Secondary Data Analysis Research (MCH SDAR)
  - https://www.grants.gov/search-results-detail/355633

Training Opportunities

- IES Webinars/Trainings
  - https://ies.ed.gov/whatsnew/conferences/
  - https://nces.ed.gov/training/datauser/

- ICPSR YouTube Video Resources
  - https://www.youtube.com/channel/UC4__DY4PxAg4Fjspbubd-Pw

Grant Tips

MAP ACADEMY

Secondary Research Datasets

- Funding programs typically specify the recommended secondary datasets that applicants should use for grant eligibility
- *"The data set can originate from one or multiple sources, including (1) federal data bases, (2) federally supported national studies, (3) international data sets supported by federal funds, or (4) statewide longitudinal administrative data systems (SLDS) enhanced through federal grants.."*
- *"Many national data resources, including important longitudinal data sets, have been developed or funded by NCES, NSF, the U.S. Department of Labor, the U.S. Census Bureau, the National Institutes of Health, or other federal agencies. International datasets such as PISA, PIAAC, TIMMS, and others are supported. If international data sets are used, the study must include U.S. education."*

Research Topics/Questions

- Fundings programs also specify topics and research questions of interest

  "The Grants Program encourages proposals across the life span and contexts of education and learning of relevance to STEM policy and practice. The research may focus on a wide range of topics, including but not limited to such issues as student achievement in STEM, analysis of STEM education policies, contextual factors in education, educational participation and persistence (pre-kindergarten through graduate school), early childhood education and development, postsecondary education, and the STEM workforce and transitions..."

Provide Evidence of Access to Data Set(s)

- E.g., “Prior to receiving funding, applicants must provide documentation that they have permission to use the data for the research project” (AERA, 2020)
- E.g., “Describe your access to any necessary datasets. Include Letters of Agreement, data licenses, or existing Memoranda of Understanding…to document that you will be able to access those data for your proposed use” (IES, 2019, p. 24)

American Educational Research Association (AERA). (2020). *Call for research grant proposals*. AERA grants program. http://www.aera.net/Professional-Opportunities-Funding/AERA-Funding-Opportunities/Grants-Program/Research-Grants

U.S. Department of Education, Institute of Education Sciences (IES). (2019). *Request for applications: Education research grants* (CFDA Number: 84.305A). https://ies.ed.gov/funding/pdf/2020_84305A.pdf

Provide Evidence of Familiarity With Data Set(s)

- Highlight publications and presentations using secondary data sets
  - Ideally ones that used this particular data set
  - But even ones that use a different data set demonstrate familiarity with the complexities of using secondary data sets
- Highlight training
  - Specific to this particular data set
  - More generic training (e.g., courses in survey design and statistics)

Provide Detailed Information About Data Set(s)

- Some RFAs require specific information about the study variables
  - E.g., “Provide a categorized list of the variables from the NCES, NSF, or other data set(s) that will be used in this research project” (AERA, 2020)
- Specify analytic sample size
  - May need to provide information about variable-level response rates/missing data
  - Conduct sensitivity analyses to demonstrate adequate power to detect a reasonably sized effect given this sample size

American Educational Research Association (AERA). (2020). *Call for research grant proposals*. AERA grants program. http://www.aera.net/Professional-Opportunities-Funding/AERA-Funding-Opportunities/Grants-Program/Research-Grants

Provide a Specific Plan for Analyzing the Data

- Allow sufficient time in project timeline for data preparation
- Describe measurement plan (e.g., creating composite variables, evaluating reliability and validity evidence)
- Highlight feasibility of merging multiple data sets by identifying common unique identifier
- Address complex sampling design (e.g., sampling weights, alternative variance estimators, subpopulation analysis)

Additional Resources

MAP ACADEMY

Additional Resources

- Davis-Kean, P. E., & Jager, J. (2017). Using secondary data analysis. In D. Wyse, N. Selwyn, E. Smith, & L. E. Suter (Eds.), *The BERA/SAGE handbook of educational research: Two volume set* (pp. 505-522). London, UK: SAGE Publications.
- Koziol, N. A. (2014, December). *Analyzing data from complex sampling designs: An overview and illustration*. Presented at the Nebraska Academy for Methodology, Analytics and Psychometrics 2014-2015 Methodology Application Series, Lincoln, NE. http://mapacademy.unl.edu/training/video-presentations-page.php?id=5230dbcc8948aeb09444ba528abab8ce
- Koziol, N., & Arthur, A. (2011, December). *An introduction to secondary data analysis*. Presented at the Nebraska Academy for Methodology, Analytics and Psychometrics 2011-2012 Methodology Application Series, Lincoln, NE. http://mapacademy.unl.edu/training/video-presentations-page.php?id=d2b184c528399f3aefe2c91b01714681
- Trzesniewski, K. H., Donnellan, M. B., & Lucas, R. E. (Eds.). (2011). *Secondary data analysis: An introduction for psychologists*. Washington, DC: American Psychological Association.
- Vartanian, T. P. (2011). *Secondary data analysis*. New York, NY: Oxford University Press.

NEBRASKA ACADEMY FOR
METHODOLOGY, ANALYTICS & PSYCHOMETRICS

Thank you!
Questions??

HyeonJin Yoon, hyoon5@unl.edu

mapacademy.unl.edu