Using Existing Data in Research

Statistics Support Virtual Conference

October 7th, 2015

Matthew A. Jones, PhD
Sunny Liu, PhD
Center for Research Quality
Agenda

• Introduction
• When to use existing data and prospective data
• Sources of existing and prospective data
• Resources and support at Walden
Introduction- Why Discuss Existing Data

• CRQ and data support.
• 20% of Walden doctoral studies are based on secondary analysis only while another 25% use secondary analysis combined with collecting new data.
• Less risk for IRB concerns and could expedite the data collection phase of research.
• Several instances where using secondary analysis would have both expedited and strengthened the capstone study.
• Novice researchers may believe it is a better study if the data is collected at a site from actual people.
When to use Existing Data

• When quality data already exists to address the research question, or
• When quality data could be generated as part of an organization’s standard procedures.
• Can be used alone or in combination with collected data.

Bottom line: Reinventing the wheel will only weaken the study/not strengthen it.
Differentiating Existing and Prospective Data

# Sources of Data for Research

<table>
<thead>
<tr>
<th></th>
<th>Quantitative</th>
<th>Non-Research/Operational Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Existing</strong></td>
<td>Quantitative data that were systematically collected for the purposes of research.</td>
<td>Quantitative data that will be collected for some other purpose, and will be reviewed systematically for use in research.</td>
</tr>
<tr>
<td></td>
<td>• Large Database • Meta-Analysis</td>
<td>• Program Impact • Client Outcomes</td>
</tr>
<tr>
<td><strong>Prospective</strong></td>
<td>Quantitative data that will be systematically collected as part of proposed research.</td>
<td>Qualitative data that were collected for some other purpose, but were reviewed systematically for use in research.</td>
</tr>
<tr>
<td></td>
<td>• Measurement Instruments • Some Surveys</td>
<td>• User transcripts • Video recordings</td>
</tr>
<tr>
<td><strong>Non-Research/Operational Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Existing</strong></td>
<td>Qualitative data that were collected for some other purpose, but were reviewed systematically for use in research.</td>
<td>Qualitative data that will be collected for some other purpose, and will be reviewed systematically for use in research.</td>
</tr>
<tr>
<td></td>
<td>• Employee data • Student records (test scores) • Administrative Data</td>
<td>• Client Records • Debriefings</td>
</tr>
<tr>
<td><strong>Prospective</strong></td>
<td>Qualitative data that will be systematically collected as part of proposed research.</td>
<td>Qualitative data that were collected for some other purpose, but were reviewed systematically for use in research.</td>
</tr>
<tr>
<td></td>
<td>• Interviews • Observations</td>
<td>• User transcripts • Video recordings</td>
</tr>
</tbody>
</table>

**Qualitative**

| **Research Data**      |              |                               |
| **Existing**           | Qualitative data that were systematically collected for the purposes of research. | Qualitative data that will be collected for some other purpose, and will be reviewed systematically for use in research. |
|                        | • Document Repository • Meta-Synthesis | • Client Records • Debriefings |
| **Prospective**        | Qualitative data that will be systematically collected as part of proposed research. | Qualitative data that were collected for some other purpose, but were reviewed systematically for use in research. |
|                        | • Interviews • Observations | • User transcripts • Video recordings |
What Are “Data for Research”?

Data used for research are different from other types of information in three key ways:

● A systematic method was or will be used to obtain the data.
● The integrity (validity) of the data is or will be evaluated before they are used or analyzed.
● The data are aligned to the research question and the anticipated analytical strategy.

**Example:** Discussing a potential problem with a couple of administrative colleagues in a worksite would not produce information that is acceptable for research. A company-wide survey of all managers and employees to understand the nature of the problem would be a source of acceptable data.
How are Data Used in Doctoral Research?

To Shape the Research Question
● *Data can provide background information and reasons for the research question, which would appear in Literature Review.*
  ○ Results from other researchers.
  ○ Statistics on the occurrence of the research problem

To Answer the Research Question.
● *Data can be used to answer the research question, as reported in Results.*

To Elaborate on the Results
● *Data can provide added insight, while not being part of the main analysis.*
  ○ Demographic variables that describe the sample.
  ○ Environmental referents (open ended questions) on a quantitative survey.
Taxonomy of Data Sources

Types of Data
● What is the nature of the data being used?
  ○ Quantitative or qualitative?

Purpose for the Data
● The data were or will be collected for what reason?
  ○ Is the present study the primary purpose or a secondary purpose of the data?

State of the Data
● Do data currently exist or will be they be collected at some future point?
Original Purpose for the Data

For Research Purposes

● *Data collected to support the scholarly research process.*
  ○ May be collected by the student or some third party, with the goal of collecting information for research purposes.

For Non-Research/Operational Purposes

● *Data collected for another purpose, but are used for scholarly research.*
  ○ May be collected by some third party, or by the student as a member of that setting.
  ○ Dual roles may need to be clarified.
State of the Data

Existing
- Data that were collected before the research question was finalized.
  - These studies are often called “secondary analyses” because the data were originally created for some other purpose (research or operations) and the current doctoral research project is a secondary purpose.

Prospective
- Data that will be collected after the research question has been finalized.
  - It is also possible for a student to perform a secondary analysis of data that has not yet been collected (i.e., when the data will be primarily collected for another purpose and its use in research is secondary).
Existing Quantitative Research Data

Quantitative data that have been collected for research purposes, which can be made available to other researchers for analysis:

- Public databases (e.g., NIH, state agencies)
- Controlled access databases
- Published research results (for literature reviews, meta-analysis)
- Archived research data from other researchers (ICPSR)
- Data from a previous study conducted by the student researcher.

Example: *A regression analysis of protective factors’ prediction of youth outcomes in a large multi-city dataset that was originally only used to report relationships between risk factors and youth outcomes.*
Prospective Quantitative Research Data

Quantitative data that will be collected for research purposes.
● Psychological tests given under Walden IRB and committee oversight.
● Surveys and perceptual measures administered under Walden IRB and committee oversight.
● Personal attributes (BMI, age) collected under Walden IRB and committee oversight.
● A student who is part of a larger research team under the oversight of another organization can use some or all of that project’s data in a secondary analysis.

Example: A causal-comparative design comparing the differences in trait anxiety and self-efficacy in doctoral students who video-chatted regularly with their chair in the first term of capstone completion, and those that not.
Resources for Existing Datasets

• Through Walden CRQ
  – [http://academicguides.waldenu.edu/researchcenter/dataresources](http://academicguides.waldenu.edu/researchcenter/dataresources)
  • ICPSR or [www.ICPSR.umich.edu](http://www.ICPSR.umich.edu)
  • Social Change Impact Report Datasets

• Public data (i.e. [www.data.gov](http://www.data.gov))

• Government and private foundation resources

• Organization owned datasets
Data from Walden CRQ Website

Data Resources & Support: Home

Existing Datasets

Often researchers are not aware of the rich amount of existing data to support their research. Within our site we want to support you in your efforts to explore these resources. We have listed some resources below to help you get started in considering and exploring secondary data.

- Tutorial on Secondary Data Analysis
- Accessing versus generating research data
- Sources of Data for Research
  A Research Primer
- Transcript: Secondary data in doctoral research

Secondary Data in Doctoral Research

Play video
ICPSR Homepage

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.

Announcements
- Council on Libraries and Information Resources (CLIR) seeks hosts for 2016-2018 postdoctoral fellows
- New Releases through 2015-09-20
- ICPSR updates Collection Development Policy

Core Partners
- MEASURES of EFFECTIVE TEACHING Longitudinal Database
- NAHDAP
- NACDA Program on Aging
ICSPR Site Search

Find Data

Find Data

More Search Options

- Browse by topic
- Browse by series
- Browse by geography
- Browse by investigator
- Browse international data
- View all studies
  - View all studies for which online analysis is available
  - View all replication datasets (more info)
  - View all studies that have learning guides
- Browse by thematic collection

Search Tips

Select a link below to reveal search tips relevant to your search.

- I'm looking for datasets related to a specific topic.
- I'm looking for datasets that contain specific variables.
- I'm looking for datasets that answer a particular research question.
- I'm looking for a specific dataset; I know the title and/or principal investigator.
ICPSR Browse by Topic

Users have two options for identifying studies related to a particular topic.

- **Topic Classifications** address topics broadly, and are meant to be a good starting point for exploring the collection.
- The **Thesaurus** contains both broad and narrow terms that are more specific to the subject matter of individual studies.
- Our **Thematic Collections** are full websites that focus on particular subject areas.

**Topic Classifications**

1. Census Enumerations: Historical and Contemporary Population Characteristics
2. Community and Urban Studies
3. Conflict, Aggression, Violence, Wars
4. Economic Behavior and Attitudes
5. Education
6. Elites and Leadership
7. Geography and Environment
8. Government Structures, Policies, and Capabilities
9. Health Care and Facilities
10. Instructional Packages
12. Legal Systems
Example: Data file for Download

National Survey on Drug Use and Health, 2012 (ICPSR 34933)

Alternate Title: NSDUH 2012

Principal Investigator(s): United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality

Summary:
The National Survey on Drug Use and Health (NSDUH) series (formerly titled National Household Survey on Drug Abuse) primarily measures the prevalence and correlates of drug use in the United States. The surveys are designed to provide quarterly, as well as annual, estimates. Information is provided on the use of illicit drugs, alcohol, and tobacco among members of United States households aged 12 and older. Questions included age at first use as well as lifetime, annual, and past-month usage for… (more info)

Series: National Survey on Drug Use and Health (NSDUH) Series

Access Notes

- These data are freely available.

Dataset(s)

WARNING: This study is over 150MB in size and may take several minutes to download on a typical Internet connection.

National Survey on Drug Use and Health, 2012 - Download All Files (2,160.1 MB) large dataset

Documentation:
- Codebook.pdf
- Documentation.pdf (method_resources)
- Questionnaire.pdf (screener)
- Questionnaire.pdf (showcards)
- Questionnaire.pdf (specs)

Data:
- SAS
- SPSS
- Stata
- R
- ASCII
- Delimited
- ASCII +
- SAS Setup
- SPSS Setup
- Stata Setup

Analyze Online:
- simple crosstab/frequency
- SDA

Quick Tables:
- Drug Use: 12-17 year olds
- Drug Use: Entire Sample
The home of the U.S. Government’s open data
Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.

GET STARTED
SEARCH OVER 187,082 DATASETS

Monthly House Price Indexes

BROWSE TOPICS

- Agriculture
- Business
- Climate
- Consumer
- Ecosystems
- Education
- Energy
- Finance
- Health
- Local Government
- Manufacturing
- Ocean
- Public Safety
- Science & Research
Controlled Access Datasets

• Controlled Access Datasets
  – Can be either research or non-research/operational data that is not publicly available (held by federal, state or local government agencies as well as NGOs, private businesses or foundations).
  – Some just might not be accessible to graduate students (may require committee chair to serve as the PI on the data access request)
  – Need to contact datasetaccess@waldenu.edu for signatory official’s approval and set up access to electronic portal (for example, eCommons).
  – Appropriate permissions/forms required by Walden and the agency holding the controlled access data.
Example: Restricted Data

Evaluation of SAFEChildren, a Family-Focused Prevention Program in Chicago, Illinois, 2006-2010 (ICPSR 33101)

Principal Investigator(s): Tolan, Patrick, University of Virginia; Henry, David, University of Illinois at Chicago; Gorman-Smith, Deborah, University of Chicago. Chapin Hall; Schoeny, Michael, University of Chicago. Chapin Hall

Summary:
Schools and Families Educating Children (SAFEChildren) is a family-focused program designed to aid families residing in high risk communities with child development during the child's transition to school. The program has the goal of building protection and impeding risk trajectories for aggression, violence, and school failure. The program utilizes multiple family groups (four to six families) combined with reading tutoring for the child. The SAFE Effectiveness Trial (SAFE-E) involved comm... (more info)

Access Notes

- One or more files in this study are not available for download due to special restrictions; consult the restrictions note to learn more. You can apply online for access to the data. A login is required to apply for access. (Instructions on YouTube.)

Dataset(s)

DSO: Study-Level Files
Data Preparation and Analysis of Existing Datasets

• Be sure to obtain necessary documentation
• You will need information on:
  – Weighting
  – Codebook
  – Crosswalk with other datasets

You may need to do some conversion to prepare data for analysis for your purposes.
Possible Challenges

- Missing data.
- Potential bias.
- Reliance on other researchers’ measures.
- Data collection process may be unknown.
- May not be a perfect fit for your research question.
- Complicated to use (sampling, waves, weight, imputation, etc.).
Support Resources

• datasetaccess@waldenu.edu
• Methodology office hours for students
  – http://academicguides.waldenu.edu/researchcenter/orqm/researchresources
• Secondary Data Analysis Tutorial
  – http://academicguides.waldenu.edu/researchcenter/dataresources
• Office of Research & Sponsored Programs (grants@waldenu.edu)
Thank you for attending!

Additional questions?

E-mail: ASCTutoring@Waldenu.edu