Using Secondary Data

Tips for accessing and interpreting data from external sources

Michael Long
Andrew MacDonald

2/27/2018

Texas Evaluation Institute

Austin, TX

Learning Objectives

By the end of this presentation, you will understand:

- What is the definition of secondary data
- Reasons to use secondary data
- Tips for obtaining secondary data

What Is Secondary Data?

Secondary data refers to data that already exist independent of your evaluation.

- Secondary data may also consist of data collected by somebody outside your organization, such as schools, government agencies, or partner organizations
  - For example, student records, census data, or unemployment insurance claims.

- Secondary data may be data you already collect through the course of implementing your own program, such as attendance records, progress reports, or training materials.

- Secondary data can be qualitative (e.g., performance reports, training materials, etc) or quantitative (e.g., sign in sheets, student test scores, etc).

ICF

Reasons to Use Secondary Data

- It is usually less expensive to use secondary data than to collect original data.
  - Even if you need to make phone calls or pay a small fee for data, these expenses are usually smaller than collecting new data. For example, you do not need to develop new instruments, train data collectors, etc.
- Secondary data is often more complete than primary data because it is already collected.
  - For example, if your program seeks to promote stable housing, then data from a public housing agency may be more complete and accurate than surveying individuals about their living situation.

ICF

Reasons to Use Secondary Data

- If you are carrying out an impact evaluation, using secondary data may help you obtain a comparison group.
  - For example, suppose your intervention seeks to improve students' reading skills. It may be easier to obtain data on students who do not receive your intervention if you use data that schools already collect.
  - You may be able to form a "virtual" comparison group by using national data sets, such as data from the Census or American Community Survey.

Cautions With Secondary Data

- **By definition, secondary data were not collected for your evaluation, and therefore may measure different constructs than you need to measure.**
  - For example, in the previous slide we described how you may be able to piggyback off data already collected by schools, housing agencies, or other potential partners. However, these data may not measure the same constructs in your theory of change.
    - For example, suppose your program deploys AmeriCorps members to run a college application counseling program, which you posit will cause more students to apply to college. You may be able to get school test and/or administrative data, but these data do not really measure the indicators you are interested in.

Cautions With Secondary Data

- Secondary data may not provide sufficient information to answer your research questions or to run statistical tests.
  - For example, depending on how and why it was collected, secondary data may not include information about respondents' age, gender, or socio-economic status, which limits your ability to control for these factors in your analysis.

Cautions With Secondary Data

- **Secondary data can be difficult to gain access to, if you do not have a close relationship with the organization collecting or storing the data**
  - It can take a significant amount of time to negotiate an agreement to get access to data—don't assume it can happen quickly
- **Don't assume that you will receive data in a format that is easy to use, understand, or analyze**
  - For example, government datasets are often complex
  - Consider working with an outside evaluator or researcher
- **Think in advance about whether you are going to want to combine secondary data with data you are collecting yourself—and how you can make that happen**
  - If your partner provides you with data on your participants but doesn't include names or identifying information, will it be helpful to you?

ICF

Tips for Obtaining Secondary Data

- **Plan ahead**
  - In some cases, your organization (or your partners) may already be collecting data that you could use for evaluation purposes
    - You may already ask people who attend events to sign-in. Could you standardize your sign-in sheets to ensure you are consistent in the way you collect attendance? Could you ask people to provide other basic information when they sign in?
    - Talk to your partners to ask them what data they already collect, and whether it might be useful for evaluative purposes.
- **Consider publicly-available datasets**
  - There are a number of government and private data sets that are available to the public.
    - In some cases these data are free, in other cases you have to pay a fee.
    - Potentially useful datasets include:
      - Student-level data on postsecondary enrollment from the National Student Clearinghouse
      - Consumer credit and borrowing data from TransUnion credit reporting agency
      - Health data from the Centers for Disease Control and Prevention and the National Center for Health Statistics
      - Income data from the Social Security Administration

ICF

Tips for Obtaining Secondary Data

- **Consider publicly-available datasets**
  - There are a number of government and private data sets that are available to the public.
    - In some cases these data are free, in other cases you have to pay a fee.
    - Potentially useful datasets include:
      - Student-level data on postsecondary enrollment from the National Student Clearinghouse
      - Consumer credit and borrowing data from TransUnion credit reporting agency
      - Health data from the Centers for Disease Control and Prevention and the National Center for Health Statistics
      - Income data from the Social Security Administration

Tips for Obtaining Secondary Data

- **Talk with your partners**
  - If you collaborate with other organizations, ask them what data they collect while administering their program, and whether they would be willing to share it with you.
  - Some organizations, such as schools or hospitals, may have standard forms or processes for obtaining data they collect
  - In addition to administrative data they themselves collect, your partners may also have access to additional non-public datasets, such as credit or health data.
- **Obtain a data use agreement**
  - When obtaining data from a partner, it is important to set establish the parameters in writing. If you are using data from an institution such as a school district, hospital, or government agency, it is likely they will have a standard agreement.
  - These agreements generally establish how you may or may not use data, how to protect individual privacy, where you can store the data,

ICF

Data Sharing Agreements

- **Dictate what data will be shared between organizations and how those data will be used**
- **These agreements typically include:**
  - A detailed description of what data points will be shared
  - The format in which the data will be provided
  - How the receiver of the data will use the information
  - How and when data will be provided (and how often, if applicable)
  - How the receiver of the data will keep the information confidential and safe
  - In what ways the receiver of the data is allowed to share the information with others (i.e., what they are allowed to report)
  - How the receiver of the data will dispose of the information after a designated amount of time
- **Sample agreements are available online, or from outside evaluators or research organizations**

Resources

- **Using Administrative Data for Randomized Evaluations, J-PAL: https://www.povertyactionlab.org/sites/default/files/resources/2017.02.07-Admin-Data-Guide.pdf**