



Using Secondary Data in Your Research

**Issues related to restricted-use and
special access**

**Anastasia R. Snyder
snyder.893@osu.edu**



Goals for today

- What is secondary data? Types of secondary data.
- How have I used secondary data in my research?
- Pros and Cons
- Types of secondary data.
- Restricted-access secondary data
- Special issues you can explore (adding questions)
- Geocoding primary data



What is Secondary Data?

- Data not collected by you for your specific research
 - National surveys usually funded by government agencies
 - ACS; NLS studies; ELS studies; Add-Health; HRS: NSFG;
 - Administrative data
 - County-level birth, death and marriage records; Food Stamp allocation data; In-take data for programs; Environmental Hazard data; etc...
 - ICPSR
- Proprietary data prepared by for-profit organization
 - EMSI employment data



My uses of secondary data

- Transition to adulthood, Home Leaving/Returning, Migration and Life Events, Rural Focus
 - **NLSY 1979 & 1997 Cohorts.**
- Impact of Coal/Shale extraction Communities and Family Outcomes
 - **American Community Survey; Decennial Census; EMSI**
- Family Formation/Dissolution, Rural/Urban
 - **National Survey of Family Growth**
- Family Economic Well-Being, Rural/Urban, Race/Ethnicity
 - **Decennial Census; ACS; Current Population Surveys; Racial/Ethnic Differences**
- Food Security among Families with Children
 - **CPS Food Security Supplement Data combined with primary data**



Pros and Cons

- **Pros:**

- Forego data collection—less costly.
- Often nationally representative (generalizability)
- Publically available
- Research productivity early in career—time (longitudinal studies) and money
- Easier IRB

- **Cons:**

- Takes a lot of time to get to know a data set. Not as easy as you think.
- Usually need extensive analytic and coding skills.
- Data not designed to answer your specific research questions—(measurement)
- Some barriers to using restricted access data.



Restricted Access Secondary Data

- **Most data sets have restricted access files.**
 - Usually for contextual purposes (rural counties)
 - Some for sensitive questions
 - Some make it very difficult
 - Special computer and space requirements for housing these data (IPR space)
- **Some have extra-restricted access files: Census, BLS, CDC.**
 - Extensive proposal
 - Full federal background check
 - Researcher must go to the data
 - Can be costly



Special Issues

- **Adding questions to on-going panel studies**
 - NLSY
 - ELS
 - Panel Study of Income Dynamics
 - Maybe more
- Costly, need to get grant funding for this
- Ex. Child Trends collaboration with NLSY



Geocoding Secondary or Primary Data

- **Attaching contextual data tied to geographic space to your data.**
 - Usually involves Decennial Census Data, but can be any data that is coded appropriately
 - Ex. County-level poverty rates, educational attainment
 - Ex. School District-level variables
 - Ex. Congressional District-level variables
- **Think about this when designing primary data studies.**