

# Leveraging Administrative Data to Assess Long-Term Effects



Janet Currie  
March, 2009

# Lack of Longitudinal Data Makes Assessing Long Term Outcomes Difficult

- 1 Collecting longitudinal data is costly
- 1 Long term outcomes cannot be immediately assessed
- 1 Attrition may present a serious problem

# Existing Data Can Offer a Solution

- 1 Can add questions to existing data sets
- 1 Can merge new information to existing data sets
- 1 Can merge several administrative data sets



## Example of Adding Questions:

1. Garces, Thomas, and Currie (2002) had special questions on early childhood education added to the 1995 Panel Study of Income Dynamics. Adult respondents  $\leq 30$  were asked whether they had ever been enrolled in Head Start and whether they had attended any other preschool.
2. Smith (2007) examines retrospective questions added to PSID on health when  $< 16$ .



## Pros:

- 1 Pro: The PSID is a long panel with rich contextual information.
- 1 Pro: The PSID offers a large sample, and also samples of siblings.
- 1 Pro: Given this data we were able to assess the impact of Head Start on Young Adults, Smith examines the impact of early health problems.



Cons:

- 1 Con: Retrospective data may be reported with error.
- 1 Con: Only outcomes already in the data can be assessed.

# Respondent Error in Retrospective Data Can be Assessed

- 1 Compare reported participation rates to known rates
- 1 Compare distributions of participant characteristics to known distributions from other data.
- 1 Investigate whether there is more “forgetting” further in the past
- 1 These tests are not conclusive but offer some reassurance of data quality.



## Merging new data to existing data sets

- 1 Generally done by geographic area and date, so requires geocodes.
- 1 Many examples:
  - i Ludwig and Miller (2007) exploit the fact that the Office of Economic Opportunity initially offered the 300 poorest counties assistance in applying for Head Start. Using the NELS, they show that children in counties with assistance were much more likely to have attended Head Start than children in otherwise similar counties.

A decorative header consisting of five circles in a row. The first, third, and fifth circles are solid light purple, while the second and fourth circles are white with a light purple outline.

# Ludwig and Miller go on to show:

- 1 Using Vital Statistics data they show that county-level mortality rates were lower in counties whose Head Start enrollment rates were higher due to the OEO assistance.
- 1 Using Census data they find that education is higher for people living in areas with higher former Head Start enrollment rates. Unfortunately, the Census does not collect county of birth ( measurement error, possibility of selective migration).

# Merging Administrative Data Bases

- 1 Requires personal identifiers.
- 1 Example: Currie, Stabile, Manivong, and Roos (2008) merge data from Canadian public health insurance records to welfare rolls and data from the educational system.
- 1 Using this data, it is possible to ask whether health problems in early childhood are related to future welfare use or lower educational attainment.



## Pros and Cons:

Pro: Large sample

Pro: Objective indicators of health and of outcomes available for full sample

Pro: Sibling comparisons possible

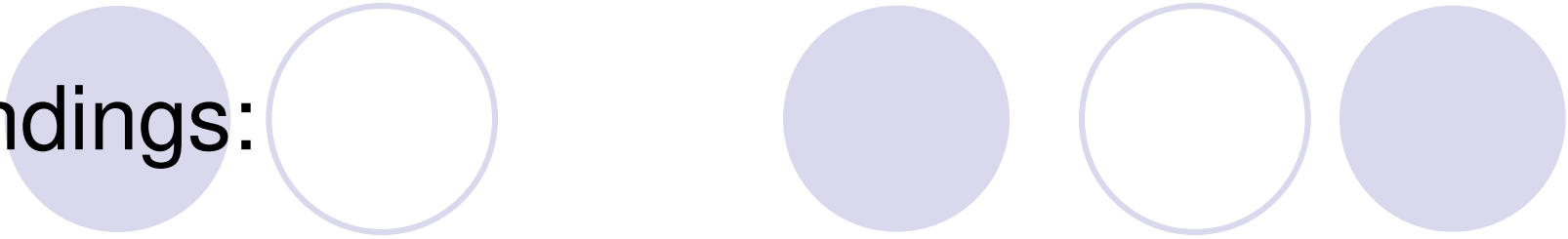
Pro: Long follow up period

Con: Limited background information

Con: Health measures depend on utilization of care

Con: Outcomes limited to those in data bases that can be merged in.

Findings:

- 
- 1 Major physical health conditions at age 0-3 are predictive of poorer educational attainment and welfare use.
  - 1 This is mainly because poorer physical health conditions at 0-3 predict poorer future health.
  - 1 Mental health conditions at early ages are independently predictive of poorer outcomes.

## Other examples:



- 1 Black, Devereux, Salvanes (QJE 2005,2007) use Norwegian data to look at long-term effects of birth weight, birth order, and family size on educational outcomes. Requires linking birth and education registries.
- 1 Doyle (2008) links child protection and criminal justice data bases in Illinois. Shows that on the margin, children who remain at home rather than being placed in foster care are less likely to be incarcerated later.



# The Major Challenge:

- 1 Privacy concerns are making it harder to obtain data just as it is actually becoming more feasible to link it. E.g. public use vital statistics natality data no longer report county of birth.

# Potential Solutions



- 1 Creators of large scale data sets need to be sensitive to the fact that their data may well be useful for questions they have not yet envisaged.
- 1 In order for data to be used in this way, it is essential to retain information that can be used for linkage.
  - 1 If it is undesirable to keep personal identifiers, geographic identifiers at a fine unit of disaggregation (e.g. Census tract or zip code) should be kept.

# Explore Methods for Making Sensitive Data Available

- 1 Suppress small cells, and/or data on rare outcomes in public use files. E.g. NCHS data sets such as NHANES and NHIS ought to report state at least for large states.
- 1 Add a small amount of noise to public use files (or data swapping to prevent identification of outliers).
- 1 Data use agreements (e.g. NLSY, NELLS).
- 1 Creation of de-identified merged files.
- 1 Secure data facilities.



# Summary

- 1 Many secrets are currently “locked up” in existing data that researchers do not have access to.
- 1 Exploring ways to make these data available might in many cases be a more cost-effective way to answer questions than carrying out new data collections.

# References

- 1 Black, Sandra E., Paul J. Devereux, and Kjells G. Salvanes. "From the cradle to the labor market? The effect of birth weight on adult outcomes," Quarterly Journal of Economics, 2007.
- 1 Black, Sandra E., Paul J. Devereux, and Kjells G. Salvanes. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Educations," Quarterly Journal of Economics, 2005.
- 1 Doyle, Joseph. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care," Journal of Political Economy, 116(4), August 2008.
- 1 Currie, Janet, Matthew Neidell and Johannes Schmeider. "Air Pollution and Infant Health: Lessons from New Jersey," forthcoming Journal of Health Economics.
- 1 Currie, Janet, Mark Stabile, Phongsack Manivong, Leslie L. Roos, "Child Health and Young Adult Outcomes," NBER WP 14482, Nov. 2008.
- 1 Garces, Eliana, Duncan Thomas, and Janet Currie. "Longer Term Effects of Head Start," The American Economic Review, v92 #4, Sept. 2002.
- 1 Smith, James P. "The Impact of Social Economic Status on Health over the Life Course," Journal of Human Resources, 42 #4, Fall 2007: 739-764.