"Inferring Sibling Relatedness from the NLSY Youth and Children Data: Past, Present, and Future Prospects"

Joseph Lee Rodgers, University of Oklahoma

Amber Johnson, Oregon Social Learning Center

David Bard, University of Oklahoma

This document represents an outline of the presentation proposed for the PAA 2006 meetings. The presentation is currently completed in a PowerPoint format, and the outline below is a subset of that presentation.

"Inferring Sibling Relatedness from the NLSY Youth and Children Data: Past, Present, and Future Prospects"

Remarkable Features of the NLSY-Children & YA Data

- **Probability sample, & sampling weights**
- **Longitudinal Structure**
- **Cross-generational structure**
  - **Some info on parents of NLSY79-Youth**
  - **NLSY-Youth info – 1979 to present**
  - **NLSY-Children info – 1986 to present**
  - **A few children born to NLSY-YA**
  - **====➔ Four generations of linked info**

    (excellent info on two generations)

- **Cross-Period comparison sample (NLSY97)**
- **Tens of thousands of variables, in many different domains**
- **Within-family information – In other words, there are lots of siblings**

Uses of NSLY-sibling data

- **Study what goes on across families**
- **Study what goes on inside a family, *across families***
- **Study special types of siblings**
  - **Twins**
  - **Male-male siblings, female-female siblings**
  - **Half-siblings, step-siblings**

- **Study special types of family relatedness besides siblings**
  - **Cousins, Half-cousins**
  - **Aunt-Nieces, Aunt-Nephews**

Twins in the NLSY-C and NLSY-YA

- **Twin sample sizes**
  - **121 total twin pairs**
    - **83 DZ (fraternal) twins**
    - **22 MZ (identical) twins**
    - **16 unknown (probably mostly MZ)**
- **Age structure**
  - **1980 & before – 7 pairs**
  - **1981 – 1990 – 68 pairs**
  - **After 1990 -- 46 pairs**
- **Fairly small for twin dataset – but, it's a probability sample of twins – those are unusual**

Cousins in the NLSY-C and NLSY-YA

- **How are they identified?**
  - **NLSY-Youth mothers who are sisters have children who are cousins**
  - **NLSY-Youth mothers who are half-sisters have children who are half-cousins**
  - **There are even NLSY-Youth mothers who are themselves identified as cousins – their children are quarter-cousins**
- **Sample sizes of cousins**
  - **56 half-cousin pairs (whose mothers lived together as cousins, in same household in 1979)**
  - **2027 cousin pairs (not so selected) – these have been very useful – and moderately used -- in past NLSY research**

Siblings!!!

- **The NLSY-Youth was conducted using a random sample of households in 1979**
  - **Lots of siblings there**
- **The NLSY-Children are all of the children ever-born to females in the NLSY-Youth**
  - **Lots of siblings there, too**
- **The NLYS-YA are all of the**

  NLSY-Children who have

  reached aged 15 or greater

  - **Lots of older siblings there**
  - **And obviously some overlap**

Full- versus Half-Siblings

- **NLSY does not distinguish Full-Siblings**

  from Half-Siblings (at least not yet)

  - **For some purposes, that's important**
    - **Behavior genetic research**
    - **Blended-family research**
  - **There is information in the data files that allows this distinction to be made**
    - **All NLSY-C siblings are full or half-siblings, because they share a biological mother**
    - **We know when many of them lived together with their biological fathers**
- **We have written algorithms to make this assignment, for both NLSY-C (including NLSY-YA) and also NLSY-Youth (others have for NLSY-C, too)**
- **Algorithms are not perfect, because information they act on is not perfect**

- **But they're pretty good, and very usable for research purposes**
- **Weaknesses – missing data, uncertainty, internal contradictions – just like in direct questions about sibling relatedness**
- **There have been around twenty or so articles published using our NLSY-YA and NLSY-C sibling links**
- **The sibling linking files are available for any researcher just by asking -- send me an e-mail**

- **Structure of kinship linking file:**
- **Other structures are useful, and can be created with data management procedures**

Sibling Sample Sizes

- **In our 2000 NLSY-C data, there are:**
  - **2,111 half-sibling pairs**
  - **5,583 full-sibling pairs -- many of which are overlapping – if that bothers you, there are:**
  - **2,174 sibling pairs who are the first two birth orders in their family (no overlap, and all adjacent birth orders)**
- **These are currently being re-done for the 2002 data, and 2004 will follow shortly**
  - **But these sample sizes won't increase very much**
- **A different format of sample size is in Table 1.6 in your packet – it separates out NLSY-C and NLSY-YA (but doesn't separate FS & HS)**

Top 10 Ways to Use Sibling Structure in NLSY Research

1. Age and Gender Matching – Longitudinal structure can be used to identify siblings at the same age – e.g., to create brother-sister pairs age matched to study gender differences

2. Control for unobserved heterogeneity (selection bias) – Selection can be (partially) accounted for by using siblings

3. Control for automatic genetic similarity

4. To study environmental influences (both nonshared and shared)

5. To study genetic influences

6. To study within-family processes; birth order, sibling spacing, etc.

7. To study more exotic sibling relationships – half siblings, adopted siblings, cousins who live in the same household

8. To study twins in an ecologically valid setting – twins have been used with other kinship categories, but no one has focused exclusively on the NLSY twins

9. To study kinship relationship created from cross-generational siblings

10. To study sibling interactions and relationships – Many have studied siblings using the NLSY, but there are still many innovative and interesting studies that could be done

Conclusion

- **The numbers inside the NLSY files are effective representations, using the best-available social science, or real-world processes**

- **The past NLSY-Y, NLSY-C, and NLSY-YA data are fixed**
- **But our designs aren't – they must be highly creative, innovative, and directed exactly toward the research questions we're addressing**
- **Sibling designs often fit all these categories -- they allow both methodological and substantive creativity, they allow design innovations, and they capture processes of great interest to social science researchers**
- **So, at the bottom line, the numbers all come back to Siblings**