
Common Problems in Working with Data on Adolescent Sexual Risk Behaviors

March 13, 2012

**Presentation to OAH and ACYF
Teenage Pregnancy Prevention Conference**

Brian Goesling

MATHEMATICA
Policy Research

Background

Need for Self-Reported Data in TPP Studies

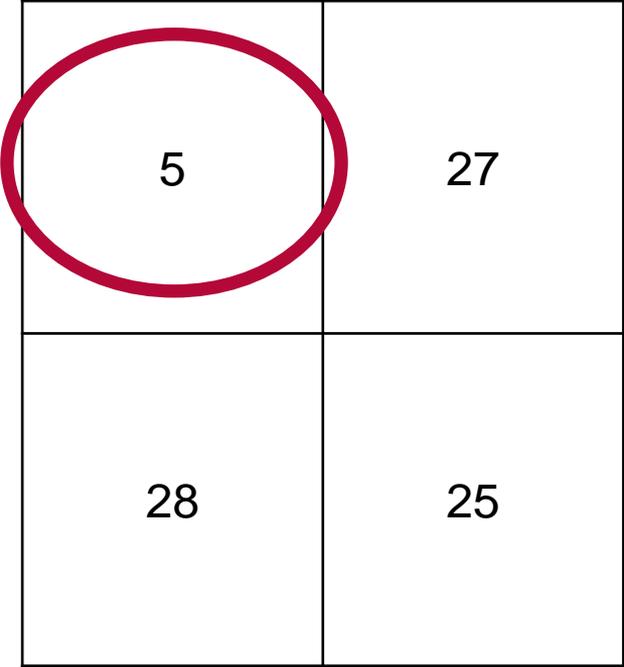
- **Only way to measure key outcomes:**
 - Sexual initiation
 - Number of partners
 - Contraceptive use
- **With some exceptions:**
 - Administrative birth records
 - Clinical tests for measuring STIs
 - Biomarkers of unprotected sexual behavior (Zenilman et al. 2005; Rose et al. 2009)

Limitations of Self-Reported Data

- **Missing data**
 - Survey nonresponse
 - Item nonresponse
- **Misreporting**
 - Recall bias
 - Social desirability bias
- **Inconsistent reporting**
 - Across items within a survey
 - Same item across longitudinal surveys

Inconsistent Reporting Within a Survey

Had Sex in Past 3 months?	Yes	5	27
	No	28	25
		No	Yes
		Ever Had Sex?	



Other Examples

- **Had sex but zero lifetime partners**
- **Fewer lifetime partners than recent partners**
- **Age at first sex older than current age**
- **More frequent contraceptive use than sexual activity**

Inconsistent Reporting Across Surveys

Had Sex at Wave II?	Yes	12.9%	29.2%
	No	54.4%	3.5%
		No	Yes
		Had Sex at Wave I?	

Source: Rosenbaum (2006) analysis of Add Health data

Other Examples

- **Declining number of lifetime partners**
- **Disappearing pregnancies or births**
- **Declining number of lifetime STIs**

Possible Reasons for Inconsistencies

- **Recall bias**
- **Social desirability bias**
- **Poor question wording**
- **Confusing skip patterns**
- **Survey fatigue**
- **Cloning**

How Do Other Surveys Address the Problem?

Youth Risk Behavior Survey (YRBS)

- **Nationally representative survey of 9th through 12th grade students**
- **Conducted every two years by CDC**
- **Sample size (2009):**
 - 158 schools
 - 16,410 students
- **Survey mode: In-school paper-and pencil interviewing (PAPI)**

YRBS Data Edits

- **Inconsistent responses are set to missing**
 - Age at first intercourse older than current age
 - Ever had sex inconsistent with age at first intercourse or number of partners
 - Fewer lifetime partners than recent partners
- **“Subverted” cases dropped if:**
 - Only few questions answered
 - Same response category selected for 15+ items
- **Fifty cases dropped in 2009 (<1 percent)**

Add Health

- **Nationally representative sample of 7th-12th graders during 1994-1995 school year**
- **Interviewed four times between 1994/1995 and 2008**
- **Data collection:**
 - **Computer-assisted personal interviewing (CAPI)**
 - **Audio computer-assisted self interviewing (ACASI)**
 - **Biological specimens for STI testing**

Add Health Skip Patterns

- **Reduces inconsistencies through use of skips**
 - If never had sex, skip immediately to next section of the survey
 - Detailed pregnancy history skipped if never had sex or no reported pregnancy
- **Facilitated by ACASI**
- **Does not account for responses across survey waves**

National Survey of Family Growth (NSFG)

- **Nationally representative survey of men and women ages 15-44**
 - Women since 1973
 - Men added in 2002
- **Sample size (2006-2010):**
 - 12,279 women
 - 10,403 men
- **Dual survey modes:**
 - Computer-assisted personal interviewing (CAPI)
 - Audio computer-assisted self interviewing (ACASI)

NSFG Skips, Edits, and Imputations

- **More extensive skips than in Add Health**
 - Assumes had sex if married
 - Partner-specific questions
- **Edit checks built into interview software**
 - “Hard” edits disallow logical inconsistencies
 - “Soft” edits prompt interviewer to question response
- **Logical imputations for most key outcomes**
 - Computer assumes answers based on past responses
 - Data file cleaned before public release

But still not foolproof...

- **Edit checks do not capture all possible inconsistencies**
 - Had sex in past twelve months...
 - But no sexual partners in past twelve months
- **Some questions repeated on CAPI and ACASI**
 - Pregnancy history
 - Ever had sex
 - Age at first intercourse
 - Number of partners

Recommendations

Design Survey to Minimize Inconsistencies

- **Weigh pros and cons of ACASI**
 - Helps minimize inconsistencies
 - But must balance against logistics and cost
- **Write good questions**
 - Shorten reference periods
 - Avoid switching reference periods
- **Make judicious use of skips**
 - Skip detailed questions on sexual activity if never had sex
 - Make formatting easy to follow

Design Survey to Minimize Inconsistencies

- **Avoid asking the same question twice**
 - Age at first intercourse
 - Date of first intercourse
- **Don't ask too many questions**
 - Prioritize primary outcomes
 - Minimize the number of secondary or exploratory outcomes

Look for Inconsistencies in the Data

- **Identify questions with high rates of inconsistent reporting**
 - Start with baseline data
 - Use findings to improve follow-up surveys
- **Flag respondents with large numbers of inconsistent responses**
- **Check for differences in reporting consistency between treatment and control groups**
 - Programs may have impacts on reporting consistency
 - Could bias study findings

Use Data Editing Procedures

- **Make logical imputations**
 - Assume no partners if never had sex
 - Assume had sex if ever been pregnant
- **Recode inconsistent responses as missing data**
- **Apply rules uniformly to all respondents**
- **Publically document procedures**
 - Published article or report
 - Online appendix
 - Supplement available upon request

Example Data Editing Procedures

Lifetime	6 months	30 days	Number of cases	Data editing rule
NO	Missing	Missing	5,412	Changed 6-month and 30-day responses to "NO."
Missing	Missing	Missing	710	No change. Cannot logically determine missing values.
YES	Missing	Missing	560	No change. Cannot logically determine missing values.
Missing	Missing	YES	192	Changed 6-month and lifetime responses to "YES."
Missing	YES	Missing	151	Changed lifetime measure to "YES."

Conduct Sensitivity Tests

- **Acknowledge subjective decisions**
- **Select best approach for primary impact analysis**
- **Check sensitivity of results to alternative approaches**
- **Publically report results of sensitivity tests**

Example Sensitivity Test

Add Health median age at first sex under different data editing procedures:

Race/ ethnicity	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7
White	17.5	17.3	17.4	17.6	17.6	17.4	17.3
Black	15.4	15.0	15.5	15.4	15.8	15.3	15.2
Hispanic	17.1	16.7	17.0	17.2	17.1	16.9	16.8
Asian	19.5	19.5	19.8	19.5	19.5	19.2	19.5

Source: Upchurch et al. (2002)

Summary and Conclusions

- **Most data problems stem from need for self-reports**
- **Example: inconsistent reporting**
- **But problems not insurmountable:**
 - **Design survey to minimize inconsistencies**
 - **Look for inconsistencies in data file**
 - **Use data editing procedures**
 - **Conduct sensitivity tests**

For More Information

- **Please contact:**
 - **Brian Goesling**
 - bgoesling@mathematica-mpr.com