NLSCY – Non-response
Non-response

• There are various reasons why there is non-response to a survey
  – Some related to the survey process
    • Timing
    • Poor frame information
    • Interviewer or field errors
  – Some related to circumstances
    • Weather
    • Language issues
    • Difficulty in tracing individuals
  – Others related to respondents
    • Unwillingness to participate
    • Unable to participate
Variety of non-response

- Total non-response
  - No information is collected
  - Insufficient information is collected
- Partial non-response (item)
  - Some individual questions were not answered
  - Some individual questions were not asked
- Partial non-response (component)
  - The NLSCY of sectioned into different groups of questions related to various topics, an entire section may be missing.
- Wave non-response
  - Where information about a respondent is available but not for every cycle of the survey due to total non-response
Dealing with non-response

• Total Non-response is measured and corrected in the NLSCY
  – Done at the cross-sectional level and reflected in the final weights
  – Cross-sectional findings are used to adjust the longitudinal weights

• Longitudinal attrition and the impact on the cross-sectional estimates for certain domains
The impact of other forms on non-response on analysis

• Either analyzing the entire dataset
  – Where a significant amount of information is missing about a variable of interest
  – Or where many variables of interest have missing data and only a minority of records have all the pieces of information

• Limiting your analysis to a subset of the population where you have reported values
  – How do you make inferences to the larger population (question of what the weighted estimates refer to)

• Complex Analytical methods
The Partition Family

Dealing with non-response in partitioned datasets

- Missing an entire component
- Missing partial information
- Missing a cycle or wave of data
How important is it?

- Maybe non-response is random.
- Maybe it's negligible
- Maybe it can be explained away
- Maybe I can get away with it
What are your options

• Report missing data as information
• Ignore missing data (limit your analysis to reported data only)
• Correct for the missing data
  – By re-weighting
  – With imputation
  – model non-response information
Get to know your non-respondents

• When you have significant non-response
  – You need to assess non-response
  – It becomes your first variable of interest
    • It's analysis like any other analysis you will do
    • Otherwise it casts doubt over every findings
Example of ignoring non respondents in your analysis

Based on the whole population...

We know that the missing information relates to this sub population...

Based on those who reported, we find that ....

Inferences are now about a sub-population only.

Relies on a good description of non-respondents | respondents
Example of imputing for non respondents in your analysis

Based on the whole population…

We know that the missing information relates to this sub population…

We compensated for this non-response by doing the following, and based on this process, we find that ….

Inferences are now about the population.

Relies on a good description of your imputation methodology
Or reweighting to compensate for non-response

- Same principle as imputation
  - Works when doing a whole components of missing values
  - Very messy in the Swiss-cheese type on non-response

- Composite methodology of imputation to adjust for local areas of non-response and re-weighting for broad areas where many variables (entire component) is missing.