

National Longitudinal Survey of Children and Youth

Guidelines for researchers and analysts using the NLSCY

Some analysts and researchers using the National Longitudinal Survey of Children and Youth (NLSCY) database have expressed a need for guidelines to help them plan their analyses and report their findings. The purpose of this document is to respond to those requests.

This document concerns research papers and reports. It deals with recommendations on what to consider when writing a paper using the NLSCY data.

Before you submit a paper for review

Methodological considerations

Before undertaking any analysis using the NLSCY data, researchers and analysts should first familiarize themselves with the complexity of NLSCY and the resulting implications for analysis. The purpose of this document is to facilitate their work by clearly identifying the key methodological issues to be considered when using NLSCY data.

This document identifies several important methodological elements to be considered when submitting a paper for review. Authors are encouraged to use these documents to ensure that they have addressed the relevant elements before submitting their research proposal or their paper.

NLSCY data can be used in many ways. The main objective of NLSCY is to allow inferences to be made about a population, using a probabilistic sample. This document has been written with this objective in mind. When NLSCY data are used with objectives other than making inferences about the population (for example a case study), some of the elements described in this document might not apply. However, for such cases, caveats provided by the author will help to put the analytical framework into perspective for the reader.

Elements of the Analytical Framework

There are six main elements to be considered in preparing a research proposal or paper using the NLSCY. These include: 1) Data sources, 2) Factors affecting the analysis, 3) Variables, 4) Type of analysis, 5) Variance estimation, and 6) Methods of analysis. These are discussed below.

1) Data Sources

All sources of data used in the analysis can be specified as follows:

a) Specify the main source of data used in the analysis.

NLSCY

Other (specify)

b) Indicate what other sources of data, if any, were used in the analysis and whether these data were included as raw data or in tabular form.

c) If the analysis was limited to a subgroup or domain, provide a description of the subgroup or domain; e.g., age groups, provinces, variables with certain characteristics, etc.

d) Specify which cycle or cycles of the NLSCY were used.

2) Factors affecting the analysis

The paper should include a description of factors that restricted or affected the analysis:

a) Description of the target population:

- Provide a clear definition of the target population of NLSCY.
- If the target population differed from the NLSCY definition, include a statement about the potential impact on the analysis.
- If comparative sources were used, include a statement about how their target populations differed from the NLSCY population.

b) Treatment of Non-response (if any)

- If some variables used in the analysis have some non-response, include a statement about the level of non-response and the impact on the analysis
- Specify how the partial non-response was handled:
 - Imputation
 - Reweighting
 - Reported as a value
 - Ignored, analysis done with the respondents only.
- Analysis of non-respondents versus respondents should be done to identify possible biases.

c) Data limitations.

- Provide the sample sizes, overall and for all sub-domains.
- Ensure that the sample sizes used in the report are sufficient both to respect confidentiality and to give reliable estimates.
- Indicate if there are any other limitations with the use of the NLSCY in your project (ex: with the variables used).

3) Variables

- Describe the variables in the NLSCY file that were used in the analysis.

4) Type of analysis

- a) Indicate the kind of study, whether longitudinal, cross-sectional, or both.
Note that if both kinds were included in the analysis, the target population may differ from one type to the other.
- b) Specify the kind of survey weights used, whether longitudinal, cross-sectional or both.
If estimates for both cross-sectional and longitudinal populations were reported, ensure that the appropriate weights were used for each analysis.
- c) If survey weights were not used, include an explanation of why not.
It is unlikely that the use of survey weights is irrelevant to the analysis.

5) Variance estimation

- a) Describe the method of estimating precision when making inferences, including the measurement of the variances and/or coefficients of variation (CV) used:
 - Approximations using the CV look-up tables (available for cycles 1 to 4)
 - Use of the NLSCY Excel Interface with CVs for many domains of interest
 - Use of the bootstrap weights with the Bootvar program, SUDAAN, or some other program that incorporates the bootstrap weights
 - No estimation of variance or coefficient of variation was done.
Note: that this would imply that no statistical inferences can be made in the paper.
 - Use of other software (specify software: _____)
Note: very few software programs are capable of handling the complex survey design of the NLSCY when estimating the variance.

6) Methods of analysis

- a) Present a description of all analytical methods used.
- b) Describe the statistical techniques used to determine whether the estimates were statistically significant.
- c) Include confidence intervals based on appropriate variance calculation.

Summary

A reviewer of your paper, who has access to the same data as you do, should be able to reproduce perfectly your results and reach the same conclusions, given the methodology you have used is sound and well explained in your paper.