

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 the research proposal. It gives the reader recommendations on different methodological aspects to consider when submitting a research proposal using NLSCY as a source of data.

Before you submit a research proposal 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 research proposal. Authors are encouraged to use this document to ensure that they have addressed the relevant elements before submitting their research proposal.

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 to be used in the analysis can be specified as follows:

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

NLSCY
Other (specify)

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

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

d) Specify the cycle or cycles of the NLSCY to be used.

2) Factors affecting the analysis

The research proposal should include a description of factors that may restrict or affect the analysis:

a) Description of the target population:

- Provide a clear definition of the target population of NLSCY.
- If the target population differs from the NLSCY definition, include a statement about the potential impact on your analysis.
- If comparative sources will be used, include a statement about how their target populations differ from the NLSCY population.

b) Treatment of Non-response

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

c) Data limitations.

- Provide the sample sizes, overall and for all sub-domains, where this information is known.
Sample sizes will be needed that are sufficiently large both to respect confidentiality and to give reliable estimates.
- Indicate if any other limitations are foreseen with the use of the NLSCY in your project.

3) Variables

- Provide a preliminary list of the variables in the NLSCY file to be used in the analysis.
- Indicate both predictor and outcome variables to be considered, to the extent that this is known.
Note that extensive information about variables can be learned before accessing the master files by studying questions in the questionnaires (on the Statistics Canada web site) or examining variable lists in the data files (via the Data Liberation Initiative at university libraries)

4) Type of analysis

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

5) Variance estimation

- a) Various methods are available for estimating precision when making inferences, including the measurement of the variances and/or coefficients of variation (CV). The research proposal should include some indication of the approach to be used, if possible. Options include:
- Approximations using the CV look-up tables (available for the first 4 cycles)
 - 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
Note: that this would imply that no statistical inferences are being made.
- 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 planned analytical methods.

b) Describe the statistical techniques to be used to determine whether the estimates are statistically significant.

c) Plan to include confidence intervals based on appropriate variance calculation in the analysis.