Development of Statistical Sampling Strategies and Optimal Design Considerations for National Children's Study Exposure Assessments

Warren Strauss, ScM[1] and Louise Ryan, PhD[2]

[1] Statistics and Information Analysis, Battelle Memorial Institute, Columbus, OH and [2] Harvard School of Public Health, Boston, MA

(Conducted under contract from the National Children's Study to Battelle Memorial Institute)

Introduction

The National Children's Study requires the collection of a wide range of exposure data over a number of years. Because of the size and complexity of the Study, an efficient exposure assessment strategy is required for data collection while maintaining data quality. An efficient strategy is one that considers costs, participant burden, data quality, and sample preservation for future analyses. This research evaluated the use of validation samples for introducing efficiency in the National Children's Study data collection effort.

Validation samples have the potential to allow the National Children's Study to capitalize on less precise or accurate measures of exposure for the majority of the cohort while still preserving the ability to assess the impact of "true" exposure on the health outcomes of interest.

What is Validation Sampling?

- A validation sample is a small sample that is designed to provide information related to the bias or error introduced by using alternative measures of exposure.
- The information gathered from the validation sample is designed to allow for appropriate statistical adjustments to the data collected in the larger cohort to address bias and error.

Example Uses of Validation Samples Within the National Children's Study

Larger Cohort	Collect exposure information using low-cost, low-precision methods across the cohort	Use of a single biomarker for most respondents, even when there is an anticipated high degree of within-person temporal variability
Validation Samples	Conduct detailed environmental assessment	Include people with multiple measures over time

- When reasonable surrogate measures of exposure are available, such as lower cost, less detailed, or less accurate measures, it may not be necessary to collect the "ideal" exposure information for the entire cohort.
- This efficiency results in reduced costs, reduced subject burden, and increased Study feasibility.

Conceptual Model for Validation Samples

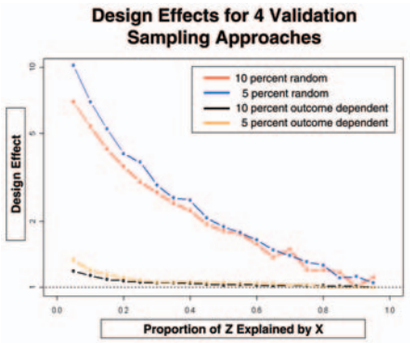
- Y is the health outcome of interest, X is the "gold standard" exposure measure, and Z is the less precise measure of exposure.
- X is measured on a small subset of the cohort (validation sample), while Y and Z are measured on the full cohort.
- This allows leveraging the information contained in the small validation sample to draw inferences on what the exposure is in the Z sample and the effect of the "true" exposure (X) on outcome (Y).
- There are 3 general methods for selecting the subset of Study participants for the validation sample:
  1) Outcome dependent sampling (depending on Y)
  2) Covariate dependent sampling (depending on Z)
  3) Random sampling (no prior information).

Estimating Design Effects from Validation Sampling

- To measure the loss of statistical efficiency as a result of using the validation sampling approach, we compute a **design effect**.
- Design effect = ratio of the variance of the estimate of the relationship between *Y* and *X* under the validation sampling approach versus the corresponding variance under an approach that measures *Y* and *X* on the entire cohort.
- The magnitude of the design effect allows us to assess the magnitude of the loss of information resulting from the use of a validation sample as opposed to collecting the true exposure information for the entire cohort.
- The degree of loss depends on a number of factors, including: the availability and accuracy of a surrogate (less expensive/detailed) measure of exposure, the strength of the exposure/outcome relationship, the methods used in selecting the validation sample, and the size of the validation sample.
- All validation sampling approaches have Design Effects > 1 indicating some loss of efficiency.

Example Results/Outcomes

- The figure displays design effects as a function of the strength of the relationship between X and Z for validation samples selected using random and outcome dependent approaches. Validation sample sizes were fixed at 5 and 10 percent of the original cohort size.
- As seen in the figure, provided there is some reasonable surrogate measure for the exposure of interest, there can be relatively little loss of statistical efficiency (e.g., design effects less than 2.0 when the portion of the variability in Z that is explained by X is greater than 0.50 with random sampling).
- Outcome dependent designs are much more efficient with design effects < 1.5.
- Small losses of statistical efficiency are likely outweighed by reductions in costs and burden associated with exposure assessment and the ability to draw unbiased inference on important relationships.



Implications for the National Children's Study

- Validation sampling provides a statistical basis to correct for bias or error in exposure assessment when investigating relationships.
- Allows the National Children's Study to collect less detailed measures of exposure for the majority of the cohort while preserving the ability to assess the impact of the "true" exposure on disease (which also limits the burden on participants).
- Can be used to address the need for pre/peri-conception exposure information – a small portion of the Study could be recruited as the pre-conception validation sample that undergoes all desired pre/peri-conception data collection, while the remainder of the cohort is recruited later (e.g., during pregnancy) with pre/peri-conception exposure information assessed retrospectively.
- Key issues for future research include developing optimal validation sampling designs tailored to specific National Children's Study hypotheses for integration into the Study protocol, and the identification of appropriate surrogate measures and their relationship with measures of true exposure.

Implementing the National Children's Study:
Scientific Progress, Challenges, and Opportunities

Study Assembly Meeting

November 29–30, 2005 • Washington, DC

THE NATIONAL CHILDREN'S STUDY

HEALTH GROWTH ENVIRONMENT