THE NATIONAL CHILDREN'S STUDY

HEALTH GROWTH ENVIRONMENT

The National Children's Study
Provider Based Sampling

Graham Kalton
Westat

May 29, 2012

Overview of the Presentation

- Provider based sampling (PBS) is described in the context of a probability sample of provider locations and pregnant women within area PSUs (counties or groups of counties).
- PBS can draw on some of the experience gained from provider based recruitment but it is a very different approach.
- The aim of PBS is to identify and enroll women as early in their pregnancies as possible.
- PBS does not cover pre-conception women.

Two Versions of PBS

Version *A*: Sample pregnant women within the PSU via prenatal provider locations. This implies sampling some locations outside the PSU.

Version *B*: Sample provider locations only from within the PSU. This implies selecting some pregnant women from outside the PSU.

- *A* is compatible with alternative area-based samples in other PSUs. *B* is not.
- *A* is more consistent with the measure of size used in sampling PSUs.

Versions *A* and *B*

- With *A*, birth certificate data for the PSU can be used to assist in constructing the frame of provider locations and measures of size, also for checking on sample coverage. For *B*, the birth certificate data from adjacent counties would also be needed.
- With *A*, there are more small provider locations because size is measured in terms of pregnant women living in the county. This problem is avoided with *B*.
- Version *A* has been chosen for the pilot study now underway.

Overview of PBS

1. Construct a list frame of prenatal care provider locations that provide care to pregnant women in the sampled PSU.
2. Populate the frame with:
   a. Estimated numbers of pregnant women from within the county seen in a year - measures of size (MOS) for use in PPES sampling.
   b. Variables that may be of use in stratified sampling.
3. Select a stratified PPES sample of providers.
4. Select a sample of pregnant women who are making their first visit to a prenatal care provider location for this pregnancy.

PBS Pilot Study

- Being conducted in three single-county PSUs:
  - Harris County TX, about 68,000 births per year
  - Jefferson County KY, about 10,500 births per year
  - Worcester County MA, about 9,500 births per year
- Select 15-25 provider locations in each PSU, and around 20 eligible pregnant women at each sampled location.
- Lists of prenatal providers are compiled by study centers, complemented by birth certificate records (Harris and Worcester only).

Sampling Frames for Provider Locations

- List frames include prenatal care provider locations from outside the PSU.
- Provider locations with very small MOS are dropped from the final frames.
- Hospitals and birthing centers are included on the list to cover:
  - Women with no prenatal care,
  - Women who receive prenatal care from a provider location not on the final frame (either dropped or missed).

Birth Certificate Data

- MOS and stratification data are collected from birth certificate data and/or from a provider location questionnaire.
- Birth certificates can provide information on the MOS for each location for PPES sampling.
- Birth certificates can also provide information on such characteristics as the percent Hispanic, percent Black, percent on Medicaid that can be used for stratification.

Provider Location Questionnaire

- Numbers of providers of different types
- Type of practice
- Number of prenatal visits in 2011.
- Numbers of first prenatal visits, and of those in the county
- Payer mix
- Race/ethnicity mix of patients
- Age distribution of patients

THE NATIONAL CHILDREN'S STUDY
HEALTH GROWTH ENVIRONMENT

Approximate Numbers of Provider Locations on the Sampling Frames

- Harris County: ~360 initially, reduced to ~250 after dropping the very small ones.
- Jefferson County: ~50 initially.
- Worcester County: ~150 initially.
- Some provider locations have such large MOS's that they are selected with certainty.
- The remainder are then allocated to strata for sample selection.
- Sampled locations that do not cooperate will be replaced by matched substitutes.

Illustrative Stratification for Provider Locations for the Harris County PSU

	% Hispanic	% Medicaid	Stratum
Outside Harris County	Low		1
	High		2
Outside the Beltway	High		3
	Low		4
Inside the Beltway	Low	Low	5
		High	6
	Medium	High	7
		Low	8
	High	Low	9
		High	10

THE NATIONAL CHILDREN'S STUDY
HEALTH GROWTH ENVIRONMENT

PPES Selection of Provider Locations

- The provider locations are to be sampled within the strata with PPES.
- The overall probability of sampling an eligible woman is to be a constant, implying that within selected provider locations women need to be sampled at rates that compensate for the differential sampling rates for the locations.

$$P(\text{woman}) = P(\text{location}).P(\text{woman}|\text{location})$$

$$= \left(\frac{aMOS_{\alpha}}{\Sigma MOS_{\alpha}}\right).\left(\frac{b}{MOS_{\alpha}}\right) = \frac{ab}{\Sigma MOS_{\alpha}}$$

THE NATIONAL CHILDREN'S STUDY
HEALTH GROWTH ENVIRONMENT

Sampling Women

- The rate of sampling women at selected location α is 1 in MOS_{α} / b, where b is the desired number of eligible to be sampled
- In Harris County, this rate is likely to range from 1 in 1.4 for locations with low MOS to 1 in 32 for locations with high MOS.
- There are various ways these rates can be applied depending on the arrangements made with the staff at the provider location.

Sampling Women

- If the location staff can provide a listing of attending women on a regular basis throughout the enrollment period, a systematic sample can be selected from that list at given intervals (e.g. weekly).
- In most cases, a sample of time intervals will be used, for example with a 1 in 4 rate:
  - One month in 4 could be selected (undesirable)
  - One week in 4 could be selected
  - One day in 4 could be selected

Sampling Women by Time Interval

- For a provider location with few eligible in-county women, the rate will be close to 1, implying selecting all women in the enrollment period.
- For a provider location with many eligible women and a rate of 1 in 32, this rate could be achieved by taking, say, a stratified sample of:
  - One day in 32 working days
  - One half-day in every 32 half-days
  - Two consecutive days in every 32 pairs of working days

Sample Size Within the Sampled Provider Locations

- The desired number of eligible women selected from each location is $b = 20$.
- Consider a location with $MOS = 80$. The sampling rate is then $MOS / 20 = 80 / 20 = 4:1$.
- Actual sample sizes will deviate from 20 due to:
  - Inaccuracy of the MOS's.
  - Variability in the numbers of eligible women sampled in the different time intervals.
- Deviations from the desired number are acceptable provided that they are within reasonable bounds.

Screening for Eligibility

- The sample of women is restricted to
  - Pregnant women making their first prenatal visits to any of the provider locations on the final sampling frame.
  - Women living in the sampled PSU.
  - Women age 18 and over.
- The sampling operation will be more efficient if location staff can accurately filter out ineligible women before the sampling is performed.
  - Screening for visits to other provider locations will not be performed by location staff.
- Eligibility will also be determined at interview.

Some of the Lessons to be Learned from the Pilot Study

- Cooperation rate for provider locations and reasons for non-cooperation.
- Response rates for sampled women:
  - Screening
  - Enrollment
- Stage of pregnancy at which enrolled
  - Proportion of women enrolled at child's birth.
- Accuracy of the MOS.
- An estimate of sample coverage.