



# Sampling Alternatives: History and Current Activity

**Lester R. Curtin, PhD**

Senior Statistician  
Division of Health Examination Statistics  
National Center for Health Statistics  
Centers for Disease Control and Prevention  
U.S. Department of Health and Human Services

April 19, 2011

# Outline



- History of Sample Design Decision
- Convenience Samples
- Probability versus convenience
- Provider survey overview
- Household survey overview
- Why is there an issue
- Current Activity
- Some basic evaluation issues to consider
- Criteria for next Decision



# History of Sample Design “Decision”



- Oct. 2002 WESTAT White paper/Report
  - Household Model
  - Office/Provider Based model
  - Center Based (Not Probability)
- Battelle Task Order White Papers/Reports
  - Hybrid Options
- NCS Sample Design workgroup
- NICHD Panel Review/Workshop (David Savitz, chair)
  - Household Model (Key: Preconception Hypotheses)
  - NICHD Decision for HH probability sample
- Federal Advisory Committee Reviews
- National Academy of Sciences Review



# Convenience sample



- Inexpensive, easier field work
- Relationship under study homogeneous with respect to population (or control variation)
- Strictly speaking, no basis for evaluation of reliability and ability to generalize
- Potential for model based evaluation, but what population
- Many examples where extremely valuable research results based on convenience samples
- Many examples of misleading results (Ellenberg, 2009)



# Probability versus Convenience



- Robert Michael and Colm O'Muircheartaigh (2008)
  - Data use and analytic objectives
  - Multiple objective studies
- NICHD Decision
  - Robust inferential design - probability sample
  - Pre-preg women, coverage – HH design
- Recognized Pitfalls to track
  - Cost and operational feasibility
  - Initial response rate and attrition



# Provider Sample Design Issues Overview



- Defining and coverage of Target population
- No pre-pregnant women
- Creating the Sample Frame (differ by area)
- Measures of size, birth counts (county of occurrence)
- Provider response rates
- Sample (mother/child) unit response rates
- Stratification variables (Geographic coverage) within PSU/Site
- Quota Sampling within specified classifications
- Analytic issues – weights, NR bias adjustment



# Household model



- Well defined area frame
- Every birth with a known probability of selection
- Demographic coverage
- Geographic coverage (residence of mother)
- Cluster sample for cost, data linkage
- 25 percent pre-conception
- 90 percent first trimester
- Problems: cost, response rates, complexity



# Why are alternative strategies under consideration



- Early Field Results from household sample in Vanguard Sites
  - Rates for screening and listing
  - Enumeration/enrollment Response rates
  - Potential field/operation changes
- Cost estimates
  - Current Household Sample estimate
  - Preliminary Provider based sample estimate
- Study Goal: Maintain a “representative” sample
  - Possibility of different sample design model in different sites
  - Is a “good” convenience sample better than a “bad” probability sample



# Survey to Sites Preliminary Assessment:



- Information used to complete form (data base to phone calls)
- How to count new pregnancies
  - Electronic medical records
  - Electronic billing records
  - Chart reviews
  - Phone calls to practices
  - Fiscal reports
  - State records (including birth certificates)
- Difficulty – range not difficult to extremely difficult to not feasible
- Big range in estimates of level of effort (e.g. chart review 300 hours to 1,000 hours)



# Current Field work



- 7 vanguard sites – continued household
- 10 sites: enhanced household
  - NHANES 78% RR (area variability)
  - New York City Exam Survey -
  - Arkansas State Exam Survey 28% RR
- 10 Sites: provider
- 10 sites: Hi/Low (convenience) sample



# Alternative recruitment Sample Design and Analytic issues



- Probability sample does not provide valid inference if not truly representative
- If no within PSU geographic clustering
  - linking geographic and contextual variables
  - Collecting Environmental samples for exposures
- Some limitation in analytic methods (HLM)
- Consistency in sampling and non-sampling error structure between PSU/sites



# Criteria for Choice of sample plan



- Total Cost (operations, staffing, remuneration)
- Feasibility
  - Frame/Birth Count Development
  - Staffing Needs
- Representativeness
  - Sample Frame Coverage
  - Response rates (Prov 83%, Direct 77%, HH 57%, EHH 52%)
  - Control of any systematic bias
- Sampling efficiency - ability to control some sources of variation
  - Within site geography
  - Characteristics of mothers

