The Impact of Cellphone Sample Representation on Variance Estimates in a Dual-Frame Telephone Survey

Elizabeth Ormson[1], Kennon R. Copeland[1],
Stephen J. Blumberg[2], and N. Ganesh[1]
[1]NORC at the University of Chicago
[2]National Center for Health Statistics, Centers for Disease Control and Prevention

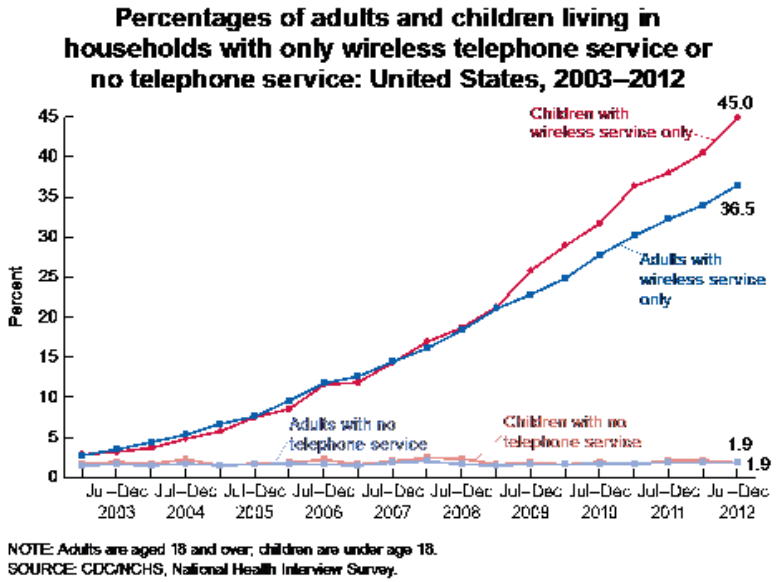
NORC at the UNIVERSITY of CHICAGO

Objectives

- Background of Issue
- Prior Research
  - Attenuation Weighting
- Adequate Cell Sample Coverage?
- Simulation Study

Background

- Current data show 89% of households (HHs) have cell telephone (Blumberg & Luke 2013)
- Cell-phone-only (CPO) HHs constitute 38.2% of population
  - CPO population has higher proportions of young adults, Hispanics, low income, renters
  - Traditional RDD landline surveys subject to bias due to noncoverage of CPO population



Inclusion of Cell Sample in Telephone Surveys

- Cell sample component added to telephone surveys
  - Cell-phone-only (CPO)
    - Dual frame with no overlap
    - Ignores potential wireless-mostly undercoverage
  - Cell-phone-only + Subset of Dual-Users
    - Address potential nonresponse bias due to dual-users not answering landline
    - CPO + Cell-Phone Mostly
    - Screen to subset of dual-users that are not contactable via landline (referred to here as cell-phone-mainly (CPMa))
  - Full cell population
    - Dual frame with overlap of dual landline/cell population
    - Reduces sample size of CPO for fixed cost

Inclusion of Cell Sample in Telephone Surveys

- Cell sample size
  - Allocation tends to be smaller than indicated from population
    - Due to higher costs of cell phone interviews
  - Cell sample often selected at higher geographic level than landline sample
    - e.g., National cell samples with state level landline
  - Leads to differential weights for cell, landline sample

Weighting Issues

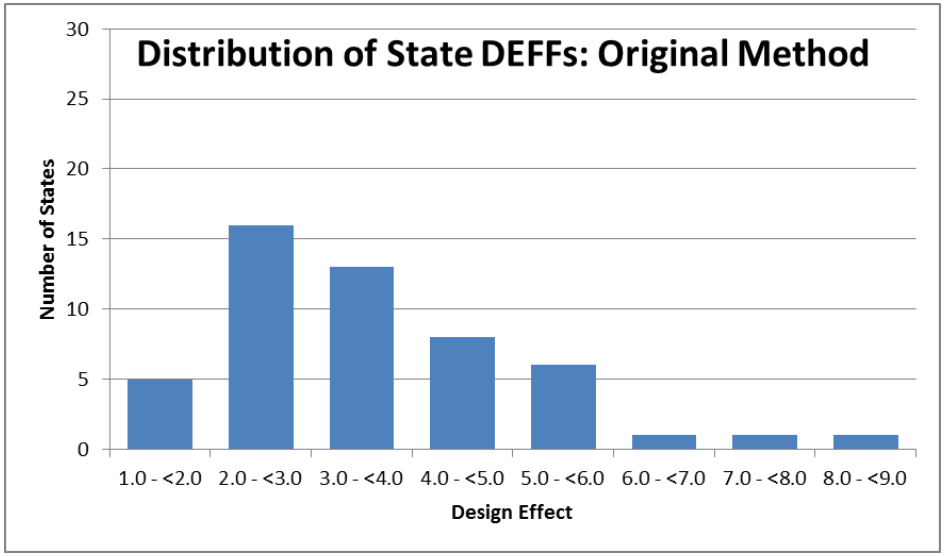
- **Differential probability of selection between landline and cell frames**
  - Large impact on variances
  - Attenuating cell sample weights can reduce variance
- Small cell sample sizes compared to landline
  - Cell sample smaller due to cost
  - Integration of population controls at different geographic levels
- Overlap of landline and cell samples
  - Full cell population overlaps with landline sample
  - Consideration of cell sample other than CPO, landline sample with cell phone

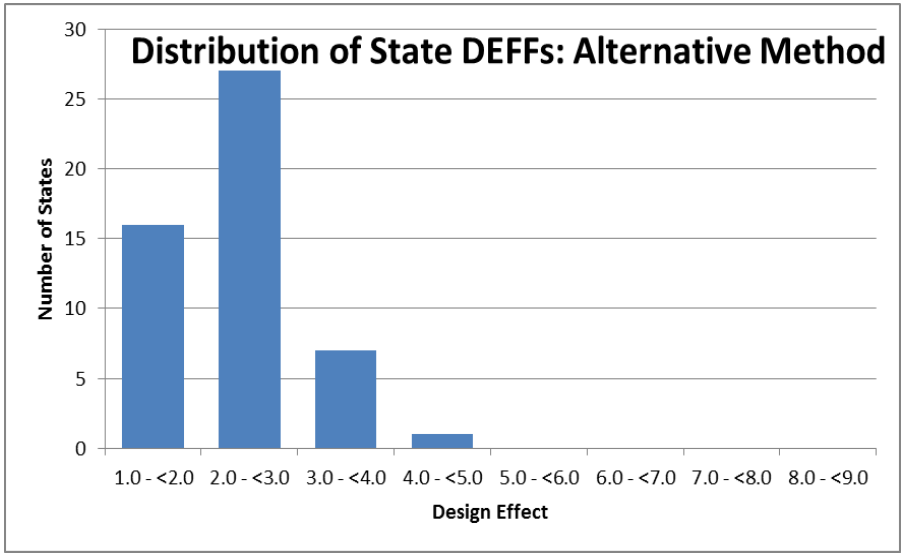
National Survey of Children with Special Health Care Needs

- National Survey of Children with Special Health Care Needs
  - Sponsored by National Center for Health Statistics
  - Data collection occurred from Q3/2009 – Q4/2010
  - Estimate prevalence of children with special health care needs (CSHCN)
  - Interviewed HHs with children <18
- Cell sample inclusion
  - Survey started as LL survey, cell added during last two quarters of data collection
  - Smaller representation than population distribution dictates
  - Screened for CPO and Cell-Mainly
  - 14.7% of sample released from cell, yielding 9.4% of completes from cell sample

Comparison of Methods: Prevalence Rate

- Created an alternative weighting method through attenuation to reduce variability of estimates
- Alternative method resulted in a smaller DEFFs for CSHCN prevalence rate
- Did introduce bias to estimates





Child Level Prevalence of CSHCN

- State Estimates of CSHCN Prevalence
  - 38.2% reduction in median DEFF
    - DC had the largest reduction, from 8.4 to 2.8
  - Slight reduction in median State Child Prevalence Rate

		Min	Median	Max
State Prevalence	Original	11.1	16.2	21.1
	Alternative	10.6	16.0	19.8
DEFF	Original	1.5	3.4	8.4
	Alternative	1.7	2.1	4.9
Absolute Differential Bias	Original	0.00	0.00	0.00
	Alternative	0.02	0.56	1.56
RMSE	Original	0.42	0.78	1.29
	Alternative	0.58	0.85	1.72

Key Indicators of NS-CSHCN Survey: Inadequate Insurance

- State Estimates of Inadequate Insurance
  - Results are more variable for key indicators than for prevalence due to smaller sample sizes
  - 37.9% reduction in median DEFF
    - ND had the largest reduction, from 8.9 to 2.0
  - Slight increase in median State Inadequate Insurance estimate
  - Reduction in median RMSE

		Min	Median	Max
Inadequate Insurance	Original	23.6	34.0	43.9
	Alternative	25.4	34.2	45.0
DEFF	Original	1.6	2.9	8.9
	Alternative	1.5	1.8	5.3
Absolute Differential Bias	Original	0.00	0.00	0.00
	Alternative	0.09	1.44	8.10
RMSE	Original	2.11	2.91	5.32
	Alternative	1.95	2.70	8.45

National Survey of Children's Health

- National Survey of Children's Health
  - Sponsored by National Center for Health Statistics
  - Data collection occurred from Q1/2011 – Q1/2012
  - Estimate general health and well-being of children
  - Interviewed HHs with children <18
- Cell sample inclusion
  - Cell Sample included during all 5 quarters of data collection
  - Smaller representation than population distribution dictates
  - "Take-all" approach
  - 33.4% of completes from cell sample
    - 16.2% from Cellphone Only

National Survey of Children's Health

- National Survey of Children's Health
  - A total of 95,677 completed surveys
    - 31,972 (33%) from cell phone sample
    - Of the cell phone completes, 48% come from Cellphone Only
    - Of the landline completes, only 5% come from Landline Only

		Dual Users			
	Cellphone Only	Cellphone Mostly	Mixed	LL Mostly	LL Only
Cell Sample	16.2%	5.7%	10.3%	1.2%	
Landline Sample		10.4%	42.9%	9.6%	3.6%
Combined	16.2%	16.1%	53.2%	10.8%	3.6%

National Survey of Children's Health

- Unclear if 16% CPO was adequate
- Ran both attenuated and non-attenuated weighting methods to compare
- Non-attenuated showed similar design effects as attenuated method
- Table shows results for one survey outcome
  - Multiple propensity model inclusion levels were compared
  - Non-Attenuated method performed just as well

	ind1_1: Child Health Status				
	Non Attenuation	Attenuated, Cut-off=0.5		Attenuated, Cut-off=0.6	
	DEFF	DEFF	Bias	DEFF	Bias
Minimum	1.76	1.55	-0.88	1.58	-0.94
Median	2.26	2.07	0.00	2.12	-0.01
90th Percentile	2.79	2.50	0.61	2.59	0.47
Maximum	3.97	3.43	0.94	3.47	0.67
National DEFF	6.69	6.19	-0.01	6.40	-0.09
National Estimate	84.16	84.14		84.06	

National Survey of Children's Health

- Similar results seen across other survey outcomes
- Final delivery used the non-attenuated weights

	Ind4_8: Medical Home Status				
	Non Attenuation	Attenuated, Cut-off=0.5		Attenuated, Cut-off=0.6	
	DEFF	DEFF	Bias	DEFF	Bias
Minimum	1.63	1.57	-1.2	1.57	-1.65
Median	2.04	1.84	0.05	1.91	0.05
90th Percentile	2.57	2.18	0.56	2.21	0.56
Maximum	3.64	3.02	1.38	3.04	1.45
National DEFF	5.33	4.88	-0.01	4.96	-0.07
National Estimate	45.73	45.73		45.66	

Discussion

- Balance between bias and variance of weighted estimates is a significant consideration
  - Determine degree of dampening so as to minimize MSE
    - Run simulations to determine optimization point
    - Degree of dampening may be somewhat subjective, and balance between bias and variance
    - End user should drive degree of dampening
  - User views of bias vs. variance
    - End user may be more concerned with bias of the estimates than the confidence intervals around the estimates
    - May require deeper explanation of attenuation approach for user to understand why it is acceptable

Discussion

- Attenuating cell sample weights is one way to account for large differences in weights between samples
  - Increases bias, reduces variance
  - Can reduce overall RMSE even with addition of bias
- Technique may not be applicable to all surveys
  - Threshold has not been tested for minimal amount of cell sample needed to use this approach
  - **Unclear how much cell sample is needed to use non-attenuated method**

Simulation

- Simulation Set up
  - Want to mimic cell sample inclusion ranging from very low up to the population target for CPO of 38%
  - Constraint is distribution of telephone status type within each sample frame
    - 48.4% of cell sample was CPO
    - 5% of landline sample was LLO
  - Determine landline completes, then manipulate cellphone completes to meet varying degrees of cell sample inclusion
    - 10% to 38%, by 1% increments

Simulation

- Simulation Set up
  - Determined largest cell sample would call for 8,814 landline completes.
    - Randomly sample landline completes to attain target sample
  - Keep landline sample stable and manipulate cell completes to achieve remaining distributions
    - Total completes range from 40,789 down to 11,105

Cell Phone Only Target	Cell Phone Only	Cell Sample-Dual Users	LL Sample-Dual Users	Landline Only	LL Sample	Cell Sample	Total Completes
38	38.0	40.4	20.4	1.2	8,814	31,972	40,789
30	30.0	31.9	36.0	2.1	8,814	14,309	23,122
23	23.0	24.4	49.7	2.9	8,814	7,956	16,770
16	16.0	17.0	63.3	3.7	8,814	4,342	13,156
10	10.0	10.6	75.0	4.3	8,814	2,291	11,105

Weighting Methodology

- Non-Attenuation Weighting Method:
  - Standard Weighting process
    - Baseweights
    - Nonresponse Adjustments
    - Raking to Population Control Totals
  - Each Telephone Status category will be weighted to represent itself
    - I.e., CPO sample weighted to represent CPO population

Weighting Methodology: Attenuation

- Attenuate cell sample weights
  - Reduce variability and improve precision of estimates
- Leverage the landline sample
  - Model CPO status using Logistic Regression
    - NS-CSHCN and NSCH have rich demographic information from interviews such as education, income, number of people in household, etc.
  - Identify landline cases that are similar to CPO cases based on model (“Proxy CPO”)
  - Potential for smallest bias, variance will depend on number of “Proxy CPO” cases that are identified

Weighting Methodology: Attenuation

- Combine proxy landline and CPO samples
  - Cell sample weights attenuated with the goal of minimizing the Mean Square Error (MSE) which combines Bias and Variance
  - Use compositing approach for combining CPO, Landline sample to derive estimate for cell only population

$$\hat{Y}_{CO} = \kappa \hat{Y}_{C(CO)} + (1-\kappa)\tilde{Y}_{L(\mathrm{CO})}$$

Where,

$$\kappa = \frac{\mathrm{var}\left(\tilde{Y}_{L(\mathrm{CO})}\right) + Bias^2\left(\tilde{Y}_{L(\mathrm{CO})}\right)}{\mathrm{var}\left(\hat{Y}_{C(\mathrm{CO})}\right) + \mathrm{var}\left(\tilde{Y}_{L(\mathrm{CO})}\right) + Bias^2\left(\tilde{Y}_{L(\mathrm{CO})}\right)}$$

- $\tilde{Y}_{L(\mathrm{CO})}$ refers to a synthetic estimate of Y_{CO} based upon data from the landline sample

Further Research

- What is ‘Adequate’
  - Is there a clear point where the attenuation method performs better?
    - MSE=Variance + Bias^2
    - Confidence interval around estimates
- State Level Estimates
  - Effect of using smaller sample sizes
- Generalize results
  - Adult and Household Surveys
  - Means and Totals

Elizabeth Ormson
ormson-elizabeth@norc.org

Thank You!

NORC at the UNIVERSITY of CHICAGO

insight for informed decisions™