Dual Frame Theory Applied to Landline and Cell Phone Surveys

Survey Research Methods Section
Webinar
November 10, 2009

Presenter

J. Michael (Mike) Brick

Westat
Joint Program in Survey Methodology
University of Maryland

Outline

- Introduction to Dual Frame Telephone Surveys
- Dual Frame Theory Assuming Only Sampling Errors
- Nonresponse and Other Nonsampling Errors In Dual Frame Telephone Surveys
- Sampling and Weighting Strategies to Reduce Bias and Deal with Differential Yield
- National & State/Local Dual Frame Surveys

INTRODUCTION

Random Digit Dialing Telephone Surveys

- Popularity of RDD surveys for household sampling rose as the percentage of households with telephones increased and efficient sampling methods were developed.
  - RDD surveys were prevalent in commercial, political, and government surveys.
- Response rates and coverage rates began to fall precipitously after 2001
  - Lead to a search for alternative approaches.
- We discuss dual frame telephone surveys as a possible remedy for the loss due to coverage.

Coverage in Telephone Surveys

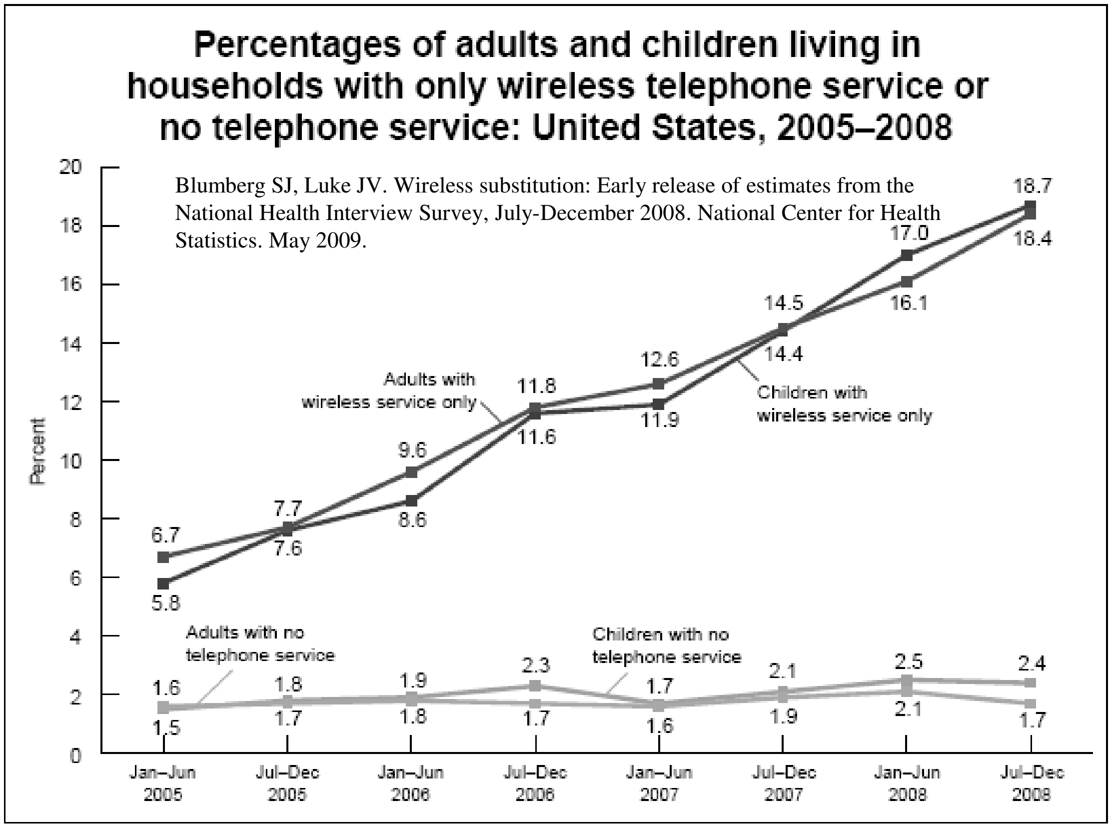


Figure 1

Blumberg and Luke (2009). Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2008.

The Wireless-Only

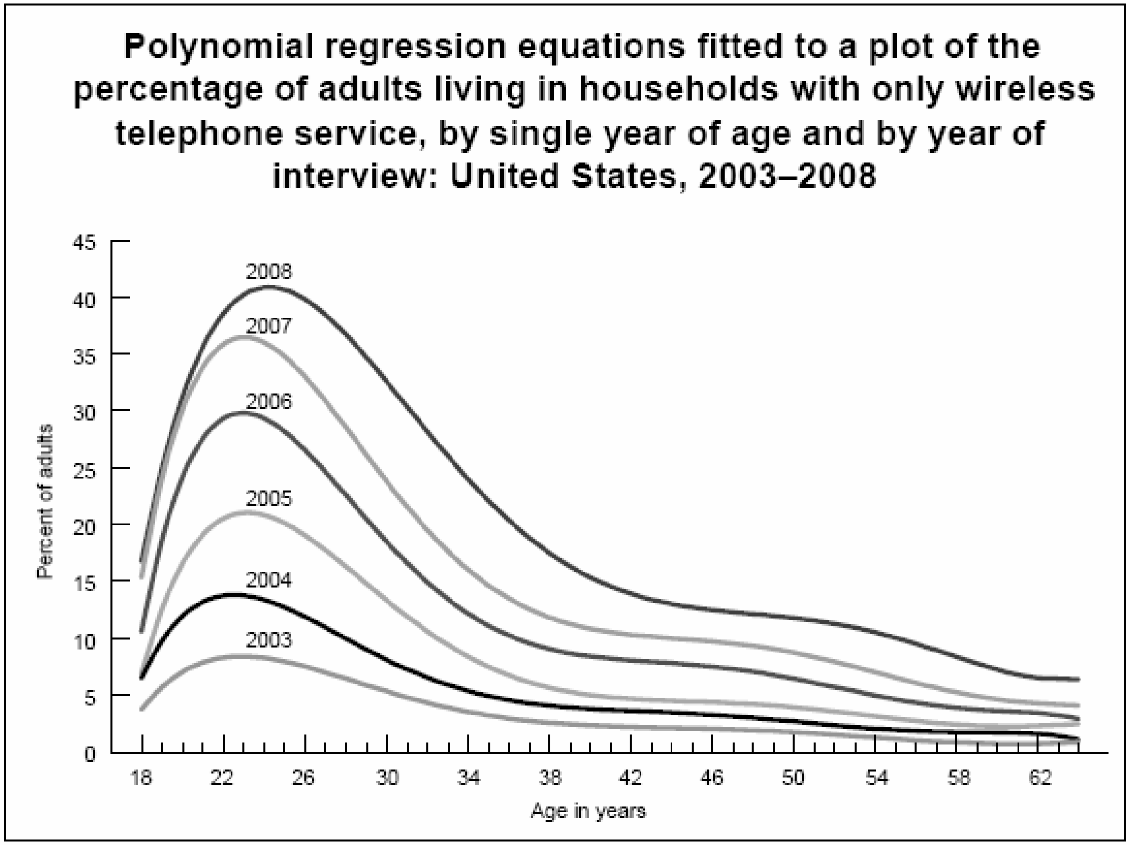
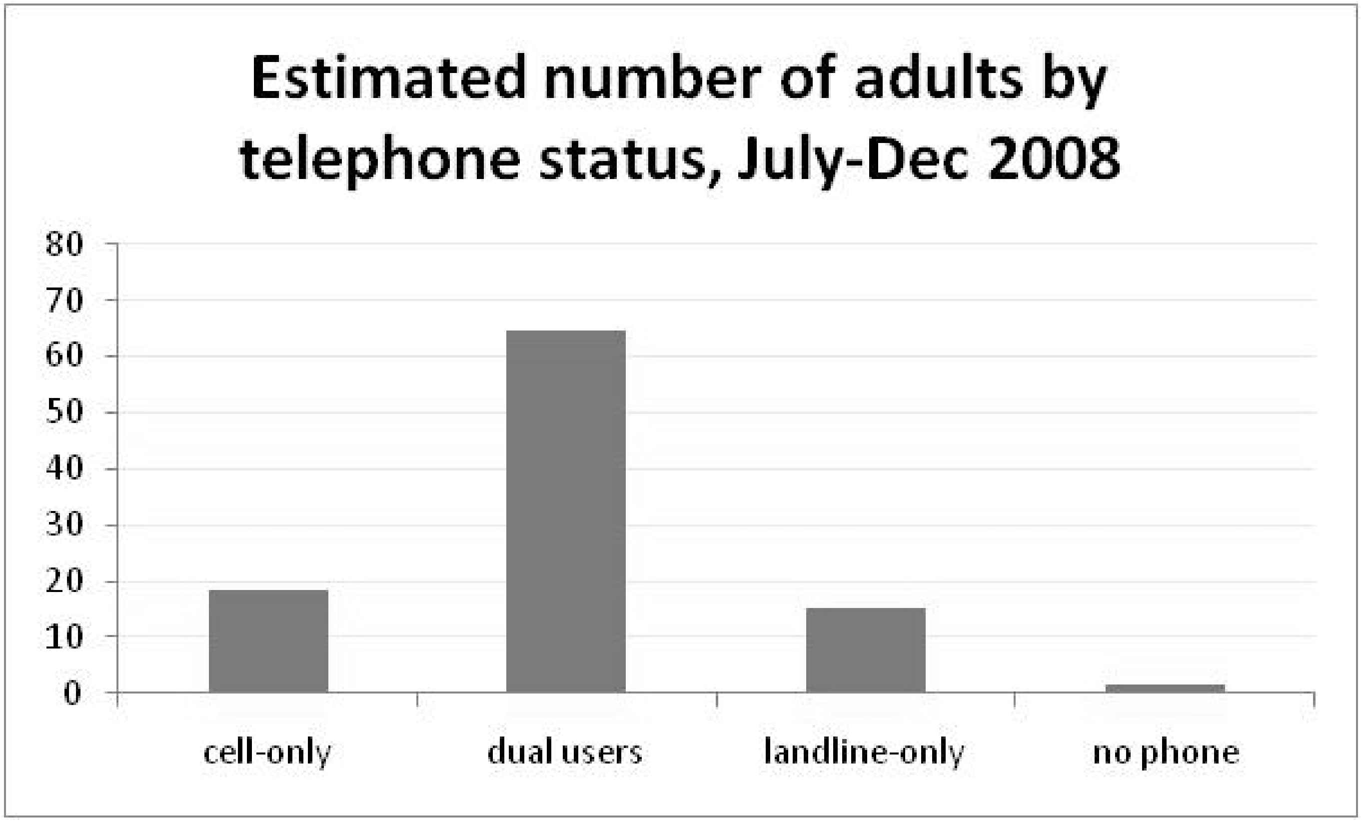


Figure 2

Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2008. National Center for Health Statistics. May 2009.

Telephone Status



Blumberg and Luke (2009). Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2008.

The Good News!

- The NHIS data show that household coverage remains high, provided we include cell phones along with landlines.

The Good News and the Bad News

- The NHIS data show that household coverage remains high, provided we include cell phones along with landlines.
- The cell phone frame has operational, ethical, cost, and statistical difficulties.
- We focus on the statistical issues associated with dual frames.

Data Collection Highlights

- TCPA regulations limit methods of calling cell phones and prohibit predictive or automated dialing, without prior consent. Manual dialing for cell surveys is permitted.
- Contact rate from the cell sample consistent across different time periods (contrast with landlines)
- Lower response rate from the cell sample than landline sample in many surveys
- Post-paid incentives may improve response rates

Other Issues for Surveying Cell Phones

- Safety and privacy concerns are greater than with landline phones since the respondents may be doing other things (like driving a car) when contacted
  - AAPOR (2008) and IRBs encourage or require a statement at the beginning of the call and the interviewer is instructed to discontinue interview if respondent is driving or otherwise engaged in activity that might cause safety concerns
- Respondents pay for their cell phone time so common practice is to reimburse for these calls; reimbursement is not required
- See Lavrakas et al. (2008)

Sampling Frame

- Sampling can be done using same Telecordia data base (previously Bellcore and even older AT&T) used to sample landlines
  - Market System Group and Survey Sampling Inc are two prominent vendors
- Information on Telecordia data base used to sample
  - More information is available for landline phones
- Households cannot be pre-identified by their telephone status (landline-only, cell-only, dual) from the frame

THEORY OF DUAL FRAME SURVEYS ASSUMING ONLY SAMPLING ERRORS

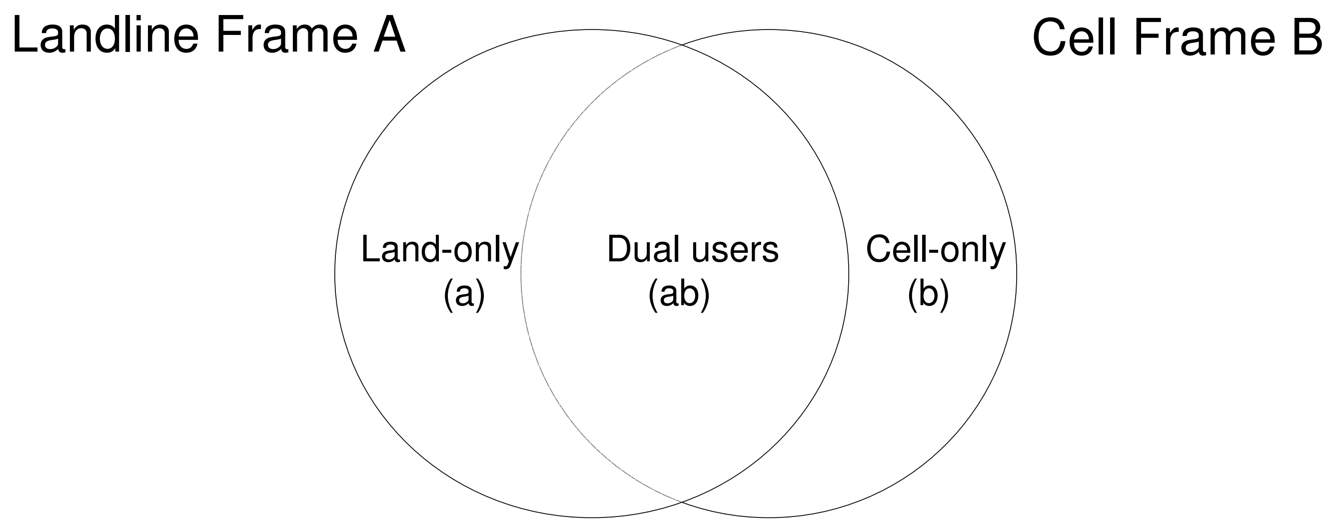
Reasons for Dual Frame Surveys

- To obtain better coverage of the population of interest.
  - No single sampling frame contains all the members of the population but two (or more) frames do cover the entire, or almost the entire, population.
- To improve the efficiency of the sample.
  - An incomplete frame contains a list of potential units that are relatively inexpensive to sample; a second, complete, frame is available but expensive to sample.

The Overlap Issue

- There is overlap of units that could be sampled from both frames.
- Many of the difficulties associated with dual frame designs are associated with methods of handling the overlap.
- The overlap must be considered in design, implementation, and estimation to get approximately unbiased estimates.
- The overlap with a dual frame design could be empty, complete, or partial.

Dual Frame Telephone Frame



Dual Frame Telephone Notation

- N_A – number of adults in landline households.
- N_a – number of adults in landline-only households.
- N_B – number of adults in cell phone households.
- N_b – number of adults in cell-only households.
- N_{ab} – number of adults in dual use households.
- n_A – number of adults sampled from land frame
- n_a – number of adults sampled from landline-only households
- n_B – number of adults sampled from cell frame
- n_b – number of adults sampled from cell-only households
- n_{ab} – number of adults sampled from land frame in dual use hhs
- n_{ba} – number of adults sampled from cell frame in dual use hhs

Dual Frame Outcome Notation

- Y_A – total of variable (y) for units in landline frame.
- Y_a – total of variable for units that are landline only.
- Y_B – total of variable for units in cell phone frame.
- Y_b – total of variable for units that are cell-only.
- Y_{ab} – total of variable for units in overlap (dual use).

Two Design Approaches

- Screening – remove the overlap so that all the units have only one chance of being in the sample used in estimation.
- General or Overlap – accept the overlap in the sample and weight the units in the overlap to account for the multiple chances of being sampled.

Screening During Data Collection

- Determine frame status during data collection, and drop sampled units from one of the frames if it is in both frames.
- In dual frame telephone surveys the screening approach that has been used is to drop units sampled from the cell frame if the household has a landline.
- This operation requires asking respondents in the cell frame sample about the telephone status of their household.

Sample Design for Screening

- The screening design is essentially a stratified sample (some technical differences), so stratified sampling procedures can be applied.
- Allocation methods for stratified sample designs if screening could be done prior to sampling.
- Same rules apply if screening is after sampling but need to estimate overlap that is expected to be removed to get N_b.

Cost Function for Screening Design

$$E(C) = c_A n_A + c_b n_b$$

c_A is the cost of a landline interview

c_B is the cost of a cellphone interview

$$c_b = c_B + (N_B N_b^{-1} - 1)c_s$$

c_s is the cost of screening

$$c_b \geq c_B$$

Screener Estimator

- For the screener estimator, we assume we know the total number of persons in cell-only households (N_b) and the total number of persons from which we get the total number of persons not in cell-only households (N_A). We post-stratify to these control totals.

$$\hat{y}_{sc} = N_A \hat{N}_A^{-1} \hat{y}_A + N_b \hat{N}_b^{-1} \hat{y}_b$$

$$\hat{N}_A = \sum_{i \in s_A} \pi_i^{-1} \delta_i; \; \hat{N}_b = \sum_{i \in s_b} \pi_i^{-1} \delta_i$$

where $\delta_i = 1$ if eligible and 0 otherwise

Variance with Screener

- To compute the optimal allocation we assume the element variances across domains is constant (σ^2). In this case, the variance of an estimated total is

$$\begin{aligned} v_{sc}^2 &= \sigma^2 \left(N_A^2 n_A^{-1} + N_b^2 n_b^{-1} \right) \\ &= \sigma^2 \left(N_A^2 n_A^{-1} + N_B N_b n_B^{-1} \right) \end{aligned}$$

Optimal Screener Allocation

- Allocation to frames that minimizes the variance of the simple unbiased estimator assuming perfect conditions is

$$n_{s,A} = \frac{CN_A}{N_A c_A + N_b \sqrt{c_A c_b}} \qquad n_{s,B} = \frac{CN_b}{N_A \sqrt{c_A c_b} + c_b N_b}$$

yielding $n_{s,b} = \dfrac{CN_b^2}{N_B(N_A\sqrt{c_A c_b} + c_b N_B)}$

Example- Assumptions

- Let N_A=0.8; N_B=0.8; N_{ab}=0.6.
  - so N_a=0.2; N_b=0.2
- Let c_A=100, c_B=300, c_s=200
- To find one cell-only we have to sample 4 on average

$$N_B N_b^{-1} = .8 / .2 = 4$$

- So c_b=900 since

$$\begin{aligned} c_b &= c_B + (N_B N_b^{-1} - 1)c_s \\ &= 300 + (4 - 1)200 \\ &= 900 \end{aligned}$$

Example- Screener Allocation

- Given the assumptions (N_A=N_B=0.8; N_{ab}=0.6; c_A=100, c_B=300, c_s=200), suppose we have a total cost of C=100,000 available for the survey.
- Then applying the optimal allocation formulae we get $n_{s,A}$=571, n_B=190, n_b=48
- As a result, we have an expected total of 620 interviews with 143 landline-only, 429 dual (from land frame), and 48 cell-only.

Sample Design with Overlap

- Assume independent samples selected from the two overlapping frames with:
  - Probability of selection of units from A = $\pi_{A,\,i}$ (n_A/N_A)
  - Probability of selection of units from B = $\pi_{B,i}$ (n_B/N_B)
- Sample allocation depends on the estimator.

Decomposition

The population total can be partitioned into three components:

$$Y = Y_a + Y_b + Y_{ab}$$

The first two components can be estimated by $\hat{y}_a$ and $\hat{y}_b$.

The overlap can be estimated by either $\hat{y}_{ab}$ or $\hat{y}_{ba}$.

With screening we made $\hat{y}_{ba}$ equal zero.

$$\left[e.g., \hat{y}_a = \sum_{i \in A\bar{B}} \pi_{Ai}^{-1} y_i \right]$$

Overlap Estimators

- Single frame estimator
- Composite estimators
  - **Average**
  - Optimal
  - Compromise
- Psuedo-maximum-likelihood estimator (PMLE)
- See Lohr (2009)

Composite – Average Estimator

An unbiased estimate is given by:

$$\hat{y}_{ave} = \hat{y}_a + \hat{y}_b + \left(\lambda \hat{y}_{ab} + (1-\lambda)\,\hat{y}_{ba}\right)$$

where $0 \le \lambda \le 1$

For example,

$$\hat{y}_{ave} = \hat{y}_a + \hat{y}_b + \left(0.5\hat{y}_{ab} + 0.5\hat{y}_{ba}\right)$$

Poststratified Average Estimator

- Estimators that take advantage of known domain totals tend to have better statistical properties, especially for estimators of totals, than those that do not use this information. The PMLE is similar to a poststratified estimator in some senses.
- In dual frame telephone surveys, we generally know N_A and N_B, and for national surveys we can use NHIS data for N_a, N_b, and N_{ab}. This makes a poststratified estimator possible:

$$\hat{y}_{ps} = N_a \hat{N}_a^{-1} \hat{y}_a + N_b \hat{N}_b^{-1} \hat{y}_b + \lambda g' \hat{y}_{ab} + (1-\lambda) g'' \hat{y}_{ba}$$

$$g' = N_{ab} / \hat{N}_{ab}; \quad g'' = N_{ab} / \hat{N}_{ba}$$

Poststratified Estimator- Overlap

- Since we are generally very concerned about the overlap, we write the poststratified estimator for the overlap as

$$\hat{y}_{ps,ab} = \lambda g' \hat{y}_{ab} + (1-\lambda) g'' \hat{y}_{ba}$$

Overlap Sample Designs

- With overlap, the sample size is determined for the two frames and the number of interviews in the overlap is determined by the allocation since all units are sampled.
- Sample allocation methods using standard sample survey methods can be applied with the poststratified average estimator .

Cost Function for Overlap Design

$$E(C) = c_A n_A + c_B n_B$$

c_A is the cost of a landline interview

c_B is the cost of a cellphone interview

Variance with Overlap

- To compute the optimal allocation, we assume the variance is constant across domains and is

$$v_{ov}^2 = \sigma^2 \left(\frac{N_A \left(N_A + \lambda^2 N_{ab} \right)}{n_A} + \frac{N_B \left(N_b + (1-\lambda)^2 N_{ab} \right)}{n_B} \right)$$

Optimal Overlap Allocation

- Allocation that minimizes the variance of the poststratified average estimator assuming perfect conditions is

$$n_{o,A} = \frac{C}{\tau\sqrt{N_A(N_a + \lambda^2 N_{ab}) / c_A}}$$

$$n_{o,B} = \frac{C}{\tau\sqrt{N_B(N_b + (1-\lambda)^2 N_{ab}) / c_B}}$$

where $\tau = \sqrt{c_A N_A(N_a + \lambda^2 N_{ab})} + \sqrt{c_B N_B(N_b + (1-\lambda)^2 N_{ab})}$

Example - Overlap Allocation

- Use same cost and population parameters as before.
- The optimal overlap allocation with λ=0.5 is

$$n_{o,A} = 366,\ n_{o,B} = 211$$

- For this example we have a total of 92 landline-only, 53 cell-only, 275 dual users from the land frame and 158 dual users from the cell frame.

Comparing Designs

Screener (n=620)	Overlap (n=577)
$n_a = 143$	$n_a = 92$
$n_b = 48$	$n_b = 53$
$n_{ab} = 429$	$n_{ab} = 275$
$n_{ba} = 0$	$n_{ba} = 158$
$v_{sc}^2 = 1.96x10^{-3}$	$v_{ov}^2 = 2.09x10^{-3}$

When is Overlap Better?

Optimally allocated overlap design has a smaller variance (assuming only sampling errors) when the cost of screening is larger enough so that

$$c_b > \frac{\left(\tau - N_A\sqrt{c_A}\right)^2}{N_b^2} = \Delta$$

In the example $c_b = 900$ and $\Delta = 1042$

NONRESPONSE AND OTHER NONSAMPLING ERROR SOURCES

Error Sources

- All surveys suffer from nonsampling errors such as nonresponse, noncoverage, and measurement errors.
- These errors may result in biases in estimates, and increased variability in estimates.
- Our main concern is bias because it has the most serious effects on inferences.

Nonsampling Errors

- Effect of nonsampling error in dual frame surveys is qualitatively different from single frame surveys.
- Errors often make it difficult to compute the probability of selection when domain membership is ascertained in data collection.
- Nonresponse may be related directly to sample frame causing differential effects.
- Complexity adds potential for more differential effects.

Nonresponse

- Response propensity may be related to the sample frame in dual frame telephone surveys.
  - Adults who rarely use cell phones may respond at a lower rate than those who regularly use them when the cell is sampled.
  - Adults who use their cell phone almost always may respond at a lower level than those who use them less frequently when sampled on a landline.

JPSM Dual Frame Telephone Study

- In 2004, the JPSM Practicum Survey sampled landline and cell phone frames. See Brick et al. (2007) for details.
- An average estimator was used with $\lambda=0.5$
- Completed extended interviews
  -- 787 from cell sample
  -- 571 from land sample
- Indicators from this survey lead authors to suggest differential nonresponse due to accessibility bias.

2007 California Health Interview Survey

- In 2007, the procedures used in a smaller pilot (see Brick, Edwards and Lee 2007) were applied to a larger survey in California using a screening design.
- Sampled one adult per household using a screening design.
- Landline sample of 49,000 adults, cell sample interviewed 825 cell-only adults.

CHIS 2007 Bias Indicators

Telephone usage	NHIS West in landline households	CHIS landline sample	Ratio	NHIS West in cell households	CHIS cell sample	Ratio
Landline–only	24%	34%	1.5	---	---	
Land-mainly	57%	53%	0.9	61%	19%	0.3
Cell-mainly	20%	13%	0.6	21%	31%	1.5
Cell-only	---	---		18%	50%	2.8

NHIS West is the West region of the 2007 National Health Interview Survey, and California is over 50% of this region.
The definitions of usage for dual users (land-mainly and cell-mainly are discussed shortly).

Pew Research Center for the People & the Press

- In 2008 and 2009, conducted series of 8 dual frame surveys on a range of topic. See Keeter et al. (2009).
- Sampled one adult per household using an overlap design.
- Aggregated over the 8 surveys, the landline sample was 11,300 adults, and cell sample interviewed 3,800 adults.

Pew Surveys Bias Indicators

Telephone usage	NHIS in landline households	Pew landline sample	Ratio	NHIS in cell households	Pew cell sample	Ratio
Landline–only	19%	23%	1.2	---	---	
Land-mainly	59%	63%	1.0	59%	42%	0.7
Cell-mainly	19%	14%	0.7	18%	24%	1.3
Cell-only	---	---		23%	34%	1.5

The definitions of usage for dual users (land-mainly and cell-mainly are discussed shortly).

Summary of Findings

- The CHIS and Pew studies both show differential reporting by frame, with more landline-only and landline-mainly reported from landline frames and cell-only and cell-mainly reported from cell frames.
- There is a very large difference in the extent of differential results, especially in the cell phone frame, for the two surveys.
- We have assumed for now that these results may be due to nonresponse, but there may be more error sources at work in these that need a fuller examination (see Brick et al, submitted).

Nonresponse Bias in Dual Frames

- Assume nonresponse bias is in estimates from the overlap (estimates poststratified to telephone status).

- Bias due to differential response rates of dual users by
  - ➢ cell-mainly – all or almost all calls on cell phone
  - ➢ land-mainly – dual users who are not cell mainly

- We explore the nature and magnitude of the bias due to this differential response and explore estimators to reduce the bias.

Overlap Nonresponse Bias

$$b\left(\hat{y}_{ps,ab}\right) \mathrm{B}\ WN_{ab}(\bar{Y}_{ml} - \bar{Y}_{mc})(\lambda r_{l1} r_l^{-1} + (1-\lambda) r_{c1} r_c^{-1} - 1)$$

W = proportion land-mainly

$\bar{Y}_{ml}$ = mean for land-mainly

$\bar{Y}_{mc}$ = mean for cell-mainly

r_l = landline response rate

r_{l1} = landline land-mainly response rate

r_c = cell response rate

r_{c1} = cell land-mainly response rate

When Overlap Nonresponse Bias = 0

- The bias equals zero for the overlap when either of the following two conditions hold:

$$\bar{Y}_{ml} = \bar{Y}_{mc}$$

$$\lambda r_{l1} r_l^{-1} + (1-\lambda) r_{c1} r_c^{-1} = 1$$

- The two condition are similar to single frame expressions, but not identical.

Condition 1 for Bias=0

- The bias equals zero for the overlap when

$$\bar{Y}_{ml} = \bar{Y}_{mc}$$

- Similar to single frame expression

$$\bar{Y}_r = \bar{Y}_m$$

Condition 2 for Bias=0

- The bias equals zero for the overlap when

$$\lambda r_{l1} r_l^{-1} + (1-\lambda) r_{c1} r_c^{-1} = 1$$

- This expression holds and is independent of λ when

$$r_{l1} r_{l2}^{-1} = r_{c1} r_{c2}^{-1}$$

- Otherwise, the bias depends on λ

Overlap Nonresponse Bias in CHIS

$$W = 0.74;\ r_{l1}r_l^{-1} = 1.09;\ r_{c1}r_c^{-1} = 0.50$$

$$b\left(\bar{y}_{ps,ab}\right) \mathrm{B}\ W(\bar{Y}_{ml} - \bar{Y}_{mc})(\lambda r_{l1}r_l^{-1} + (1-\lambda)r_{c1}r_c^{-1} - 1)$$

$$b\left(\bar{y}_{ps,ab}\right) \mathrm{B}\ .74(\bar{Y}_{ml} - \bar{Y}_{mc})(1.09\lambda + (1-\lambda)0.5 - 1)$$

$$b\left(\bar{y}_{ps,ab} | \lambda = .5\right) = .74(\bar{Y}_{ml} - \bar{Y}_{mc})(-.2)$$

$$if\ \bar{Y}_{ml} - \bar{Y}_{mc} = .2 \rightarrow b\left(\bar{y}_{ps,ab} | \lambda = .5\right) = -3.3\%$$

Overlap Nonresponse Bias in Pew Surveys

$$W = 0.81 : r_{l1} r_l^{-1} = 1.07 : r_{c1} r_c^{-1} = 0.84$$

$$b\left(\bar{y}_{ps,ab}\right) \mathrm{B} .81(\bar{Y}_{ml} - \bar{Y}_{mc})(1.07\lambda + (1 - \lambda)0.84 - 1)$$

$$b\left(\bar{y}_{ps,ab} \middle| \lambda = .5\right) = .81(\bar{Y}_{ml} - \bar{Y}_{mc})(-0.045)$$

if $\bar{Y}_{ml} - \bar{Y}_{mc} = .2 \rightarrow b\left(\bar{y}_{ps,ab} \middle| \lambda = .5\right) = -0.7\%$

Screener Nonresponse Bias

$$b\left(\hat{y}_{sc,ab}\right) \mathrm{B}\, WN_{ab}(\bar{Y}_{ml} - \bar{Y}_{mc})(r_{l1}r_l^{-1} - 1)$$

$W =$ proportion land-mainly

$\bar{Y}_{ml} =$ mean for land-mainly

$\bar{Y}_{mc} =$ mean for cell-mainly

$r_l =$ landline response rate

$r_{l1} =$ landline land-mainly response rate

When Screener Nonresponse Bias = 0

- The bias equals zero when either of the following two conditions hold:

$$\bar{Y}_{ml} = \bar{Y}_{mc}$$

$$r_{l1} r_l^{-1} = 1 \Rightarrow r_{l1} = r_{l2}$$

- These are single frame expressions.

Overlap and Screener Nonresponse Bias

Bias in Screener Design

- CHIS
  - 1.3%
- Pew surveys
  - 1.1%

Overlap Design

- CHIS
  - -3.3%
- Pew surveys
  - -0.7%

Summary of Nonresponse Bias

- The extent of nonresponse bias in the overlap depends on the sample design, the estimator, and the response rates.
- With large variation in response rates within the cell frame for the overlap (e.g., CHIS), the screener approach has a lower bias than the overlap.
- With smaller variation in response rates (e.g., Pew), the overlap approach has lower nonresponse bias.
- With the overlap approach, the bias depends on the compositing factor and can be reduced by changing this factor. No such factor is available within the screener approach.

SAMPLING AND WEIGHTING STRATEGIES FOR SCREENING AND OVERLAP DESIGNS

Bias and Yield Adjustments

- Now we examine what can be done to address the bias of the estimates, most specifically for the overlap designs.
  - Revise optimal allocation to deal with bias rather than minimizing variance.
- The final examination deals with the differential yield in the completed interviews by telephone usage and this affects both the screener and overlap designs.
  - Revise optimal allocation to deal with yield differences

Controlling Bias in Overlap Designs
By Choice of Compositing Factor

$$b\left(\hat{y}_{ps,ab}\right) \mathrm{B}\, WN_{ab}(\overline{Y}_{ml} - \overline{Y}_{mc})(\lambda r_{l1} r_l^{-1} + (1-\lambda) r_{c1} r_c^{-1} - 1)$$

Bias is zero when:

$$\lambda r_{l1} r_l^{-1} + (1-\lambda) r_{c1} r_c^{-1} = 1$$

Thus, to have zero bias set:

$$\lambda_0 = \frac{r_l (r_c - r_{c1})}{r_c r_{l1} - r_l r_{c1}}$$

Comparing Overlap Designs

λ_0=0.84	λ=0.5
$n_a = 124$	$n_a = 92$
$n_b = 42$	$n_b = 53$
$n_{ab} = 372$	$n_{ab} = 274$
$n_{ba} = 126$	$n_{ba} = 158$
$v_{ov}^2 = 2.03 x 10^{-3}$	$v_{ov}^2 = 2.09 x 10^{-3}$

Controlling Bias in Overlap By Using a 'Separate' Estimator

- If control totals by telephone usage are available, we can use them to poststratify within the overlap.
- The separate estimator is

$$\begin{aligned}\hat{y}_{sep} = {} & N_a \hat{N}_a^{-1} \hat{y}_a + N_b \hat{N}_b^{-1} \hat{y}_b + \\ & \lambda_1 g'_{ml} \hat{y}'_{ab}(ml) + (1-\lambda_1) g''_{ml} \hat{y}''_{ab}(ml) + \\ & \lambda_2 g'_{mc} \hat{y}'_{ab}(mc) + (1-\lambda_2) g''_{mc} \hat{y}''_{ab}(mc)\end{aligned}$$

$$g'_{ml} = N_{ml} \hat{N}'^{-1}_{ml}; \quad g'_{mc} = N_{mc} \hat{N}'^{-1}_{mc}$$

$$g''_{ml} = N_{ml} \hat{N}''^{-1}_{ml}; \quad g''_{mc} = N_{mc} \hat{N}''^{-1}_{mc}$$

Implications of Choice

- Revising the compositing factor is simple and can be effective for reducing the bias in the overlap design. This may result in deviations from optimal variance reduction.
- Using the separate estimator reduces the bias provided additional poststratification totals are available and accurate.
- Next, we examine the effect differential nonresponse has on the yield by telephone usage and its implications for optimal allocation.

Accounting for Differential Yield

- We noted earlier that the response rate varied by frame and telephone usage – more cell-only when sampling from the cell frame, for example.
- The allocation formulae do not account for the yield differences since they are based on population expectations.
- Assume the yield of land-only interviews from the landline frame is t_1 times more than expected (t_1^* less for the dual users); similarly for the cell frame, cell-only yield is t_2 times more than expected (t_2^* less for dual users)

Optimal Screener Allocation Accounting for Yield Differences

$$n^*_{s,A} = \frac{C\sqrt{N_A c_A^{-1}\left(N_a t_1^{-1} + t_1^{*\,-1} N_{ab}\right)}}{\sqrt{N_A c_A\left(N_a t_1^{-1} + t_1^{*\,-1} N_{ab}\right)} + \sqrt{N_B c_X N_b t_2^{-1}}}$$

$$n^*_{s,B} = \frac{C\sqrt{N_B c_X^{-1} N_b t_2^{-1}}}{\sqrt{N_A c_A\left(N_a t_1^{-1} + t_1^{*\,-1} N_{ab}\right)} + \sqrt{N_B c_X N_b t_2^{-1}}}$$

yielding $n^*_{s,b} = \frac{N_b}{N_B} n_B t_2$

where $c_x = (c_B t_2 \frac{N_b}{N_B} + c_s t_2^* \frac{N_{ab}}{N_B})$

Comparing Screener Designs

Screener (no yield)	Screener (with yield)
$n_a = 143$	$n_a = 247$
$n_b = 48$	$n_b = 85$
$n_{ab} = 429$	$n_{ab} = 431$
$n_{ba} = 0$	$n_{ba} = 0$
$v_{sc}^2 = 1.96 x 10^{-3}$	$v_{sc}^{*2} = 1.47 x 10^{-3}$

Optimal Overlap Allocation Accounting for Yield Differences

$$n_{o,A}^{*} = C\tau^{*-1} n_A = \frac{1}{\sqrt{\phi}} \sqrt{N_A c_A^{-1} \left(N_a t_1^{-1} + \lambda^2 t_1^{*-1} N_{ab} \right)} = \frac{\tau_1}{\sqrt{\phi}}$$

$$n_{o,B}^{*} = C\tau^{*-1} \sqrt{N_B c_B^{-1} \left(N_b t_2^{-1} + (1-\lambda)^2 t_2^{*-1} N_{ab} \right)}$$

$$\tau^{*} = \sqrt{N_A c_A \left(N_a t_1^{-1} + \lambda^2 t_1^{*-1} N_{ab} \right)} + \sqrt{N_B c_B \left(N_b t_2^{-1} + (1-\lambda)^2 t_2^{*-1} N_{ab} \right)}$$

Comparing Overlap Designs λ=0.84

Overlap (no yield)	Overlap (yield)
$n_a = 124$	$n_a = 211$
$n_b = 42$	$n_b = 100$
$n_{ab} = 372$	$n_{ab} = 370$
$n_{ba} = 126$	$n_{ba} = 40$
$v_{ov}^{2} = 2.03x10^{-3}$	$v_{ov}^{*2} = 1.51x10^{-3}$

Effect of Yield

- For both the screener and the overlap designs, the efficiency is increased when we account for yield differences.
- Why? The yields are increased for the domains that are smallest (cell-only for the cell frame and land-only for the landline frame). These are the domains that are controlling the variance of the estimates.

Comparing Designs with Yield

Screener

$$n_a = 247$$

$$n_b = 85$$

$$n_{ab} = 431$$

$$n_{ba} = 0$$

$$v_{sc}^{*2} = 1.47x10^{-3}$$

Overlap (λ=0.84)

$$n_a = 211$$

$$n_b = 100$$

$$n_{ab} = 370$$

$$n_{ba} = 40$$

$$v_{ov}^{*2} = 1.51x10^{-3}$$

Optimizing Accounting for Yield

- The screener approach still has a bias that depends on the relative response rates that cannot be controlled by a compositing factor. If the bias is large, then the screener approach should be avoided.
- The overlap factor cannot be optimized for the variance since we want to control the bias. Generally, bias dominates the variance in large (and even relatively small) sample sizes.
- The separate estimator with the overlap design avoids these issues, but requires more data for control totals.

APPLICATIONS IN NATIONAL SURVEYS

National Surveys – Overlap Designs

- Conduct a full dual frame survey with overlap
- Allocate the sample taking yield into account
- Estimator options
  - Use $\hat{y}_{ps}$ with λ_0 as the compositing factor
  - Use $\hat{y}_{sep}$
    - Choice of estimator depends on how well telephone usage estimates from NHIS correspond to the same characteristic reported in survey

National Surveys – Screener Designs

- Screen for cell-only households from the cell frame
- Allocate the sample taking yield into account
  - Use $\hat{y}_{sc}$
- Option of screen for cell-only &cell-mainly households from the cell frame (not covered today and not recommended as general solution).

National Survey
Advantage – Overlap Design

- Screener advantages
  - Least requirements for control totals
  - The weighting is simple
  - Assumptions about bias from nonresponse are reasonable
  - The variance is low, often lower than with overlap

- Overlap advantages
  - *The bias can be reduced or eliminated by choice of* λ.
  - The data collection is simpler (no screening).
  - Assumptions about bias from nonresponse are reasonable.

APPLICATIONS IN
STATE AND LOCAL SURVEYS

Two Key Issues for State and Local Surveys

- Sampling
  - No good way to identify local areas from the sampling frame of cell phones.
- Estimation
  - The lack of reliable control totals for telephone status and telephone usage severely limits the type of estimators that can be used and the ability to reduce nonresponse bias.

State Surveys

- NHIS state level cell-only estimates could be used as controls in a cell-only screening design, but these have large errors and have to be projected forward.
- With an overlap design, estimates of the telephone status (landline-only, dual, and cell-only) for the state must be developed for the poststratified estimator. Separate estimator requires usage control totals.
- For state surveys, the cell-only screening approach using the NHIS or other model-based estimates of cell-only population totals may be best for all but the largest states.

Local Area Surveys

- Below the state level, there are essentially no control totals that are reliable.
- An overlap design uses controls by telephone status and that make it suspect; screening design may be a safer alternative (need cell-only control).
- The cost and accuracy of the estimates of conducting a dual frame survey should be compared to a standard landline only survey.
  - The landline-only survey might still be the best alternative.