How Errors Cumulate: Two Examples

Roger Tourangeau, Westat

Hansen Lecture
October 11, 2018
Washington, DC

Hansen's Contributions

- The Total Survey Error model has served as the paradigm for most methodological work
- Hansen's contributions to the paradigm were huge
- I hope my talk today extends that paradigm, even just a little

Outline for Today's Talk

- Introduction: Are there systematic relationships between different sources of error?
- Example 1: Selection, coverage, and nonresponse error
- Example 2: Nonresponse and data quality
- Conclusions

Introduction

Introduction

- Basic distinction between observation and non-observation error
- Is there any reason to suspect any sort of general connection?
  - No, propensity to respond, join a panel, etc., mostly about motivation; observation errors largely driven by cognitive variables: "Nonresponse typically is seen as a function of motivational variables (e.g., interest in the survey topic …), whereas measurement error is considered primarily a function of cognitive factors." (Fricker and Tourangeau, 2010, p. 935; see also Yan, Tourangeau, and Arens, 2004)
  - Yes, other alternative: "This assumption of independent causal factors may be untenable, however, because the same motivations that affect participation decisions may also affect performance during the interview." (Also Fricker and Tourangeau, 2010)

Introduction—2

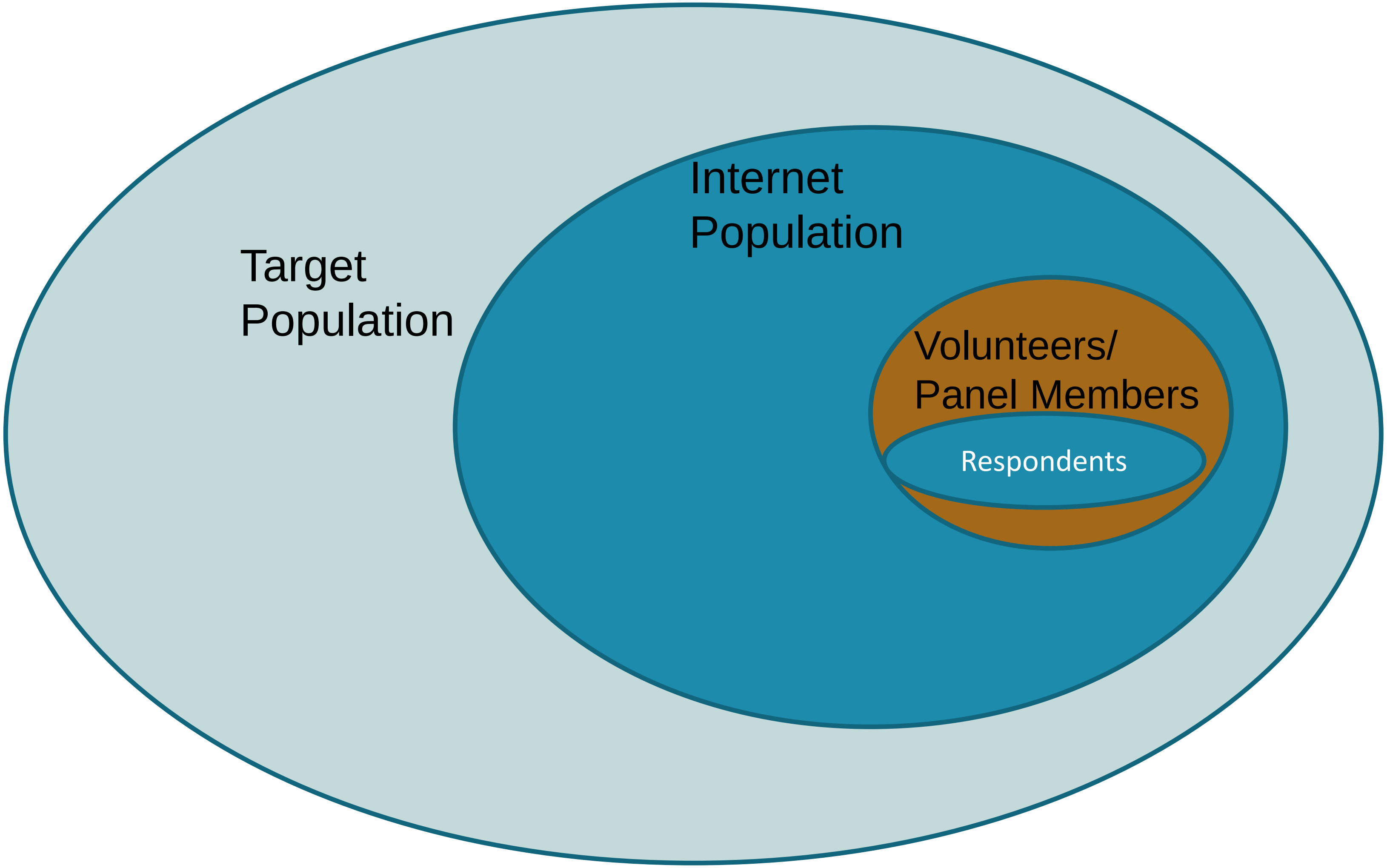
- At a very general level, there is the issue of various possible ways errors of different types can cumulate
  - Cancellation: The miracle of 2010, in the decennial census, errors of omission and duplications cancel, leaving near-zero net error
  - Accentuation: Errors from different sources—coverage problems, nonresponse—are correlated making estimates worse and worse
  - Independence: Errors from different sources unrelated (coding errors and coverage)

Introduction—3

- I want to look at two cases where there is at least some evidence about the joint impact of different errors sources
- Case 1: Non-prob web panels versus probability samples (web and telephone)—in an era of single-digit response rates, do selection biases and coverage errors make things worse?
- Case 2: Are reluctant respondents worse reporters? Is it worth it to bring them in?
- Both issues important in an era of declining response rates, higher levels of effort, and rising costs

Case 1: Coverage, Selection, and Nonresponse

Three Forms of Non-Observation Error



Westat

Three Forms of Non-Observation Error

- Coverage
- Sampling/Selection Bias
- Nonresponse
- Not clear how effects cumulate; some evidence suggests they do *not* cancel out (see Krosnick and Chang, summarized below on slide 23)

Three Forms of Non-Observation Error—2

$$B(\hat{\theta})_{tot} = B_{Int} + B_{sel} + B_{nr} \ ,$$

$$B_{cv} = \theta_{Int} - \theta_{full} \ ,$$

$$B_{sel} = \theta_{sam} - \theta_{Int} \ ,$$

$$B_{nr} = \theta_{res} - \theta_{sam} \ ,$$

$$B(\hat{\theta})_{tot} = (\theta_{Int} - \theta_{full}) + (\theta_{sam} - \theta_{Int}) + (\theta_{res} - \theta_{sam}) \, .$$

Coverage　　Selection　　Non-Participation

Biases from Non-Probability Sampling

- The key statistical consequence of nonprob sampling is bias
- Unadjusted estimates (means or proportions) from non-probability samples are likely to be biased estimates of the population means or proportions
- The size and direction of the bias depend on two factors:
  - one reflecting the proportion of the population with no chance of inclusion in the sample (for example, people without web access or people who would never join a web panel)
  - one reflecting differences in the inclusion probabilities among the different members of the sample who could in principle complete the survey

Bias Expression

- $Bias = E(\bar{y} - \bar{Y})$

$$= P_0(\bar{Y}_1 - \bar{Y}_0) + \frac{Cov(p, y)}{\bar{p}}$$

$\bar{y}$: a sample mean (or a sample proportion) based on those who complete the web survey

$\bar{Y}$: the corresponding population mean or proportion

P_0: the proportion of the population of interest with no chance at all of participating in the survey (e.g., those without web access);

$\bar{Y}_1$: the mean among those with a non-zero chance of taking part;

$\bar{Y}_0$: the mean among those with zero probability of taking part;

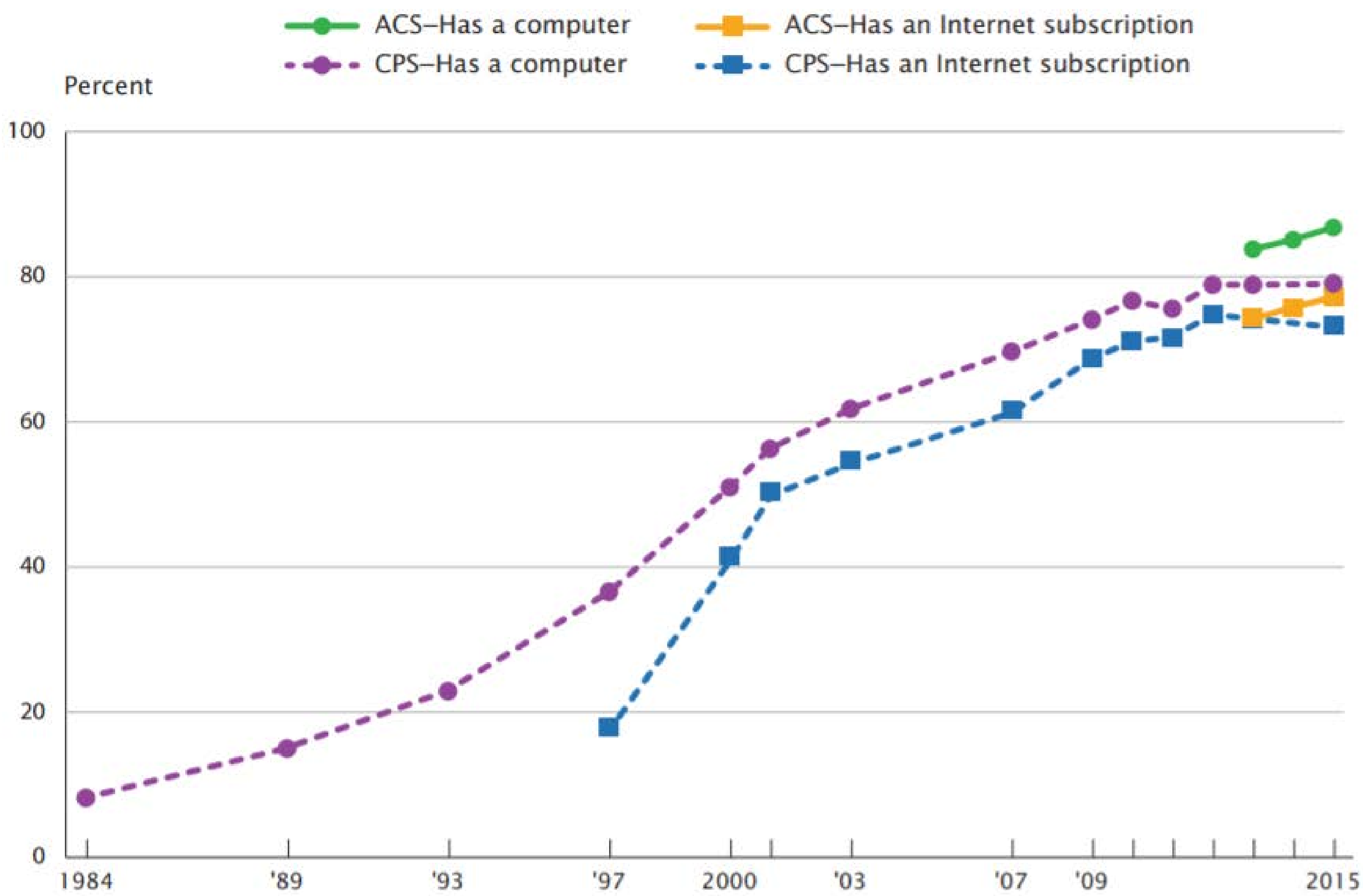
Cov (p,y): the covariance between the probabilities of participation (*p*) and the survey variable of interest (*y*) among those with some chance of taking part

$\bar{p}$: mean probability of participation among those with a non-zero probability of taking part

Trends in Internet Coverage

- Two key factors—proportion who aren't covered and differences between those who are covered and those who are $P_0(\bar{Y}_1 - \bar{Y}_0)$
- Percent with access and characteristics of those with and without access

Trends in Access in US



The Digital Divide

- Who has access and who doesn't?

Percentage of U.S. Population with Internet Access
by Demographic Subgroup

Subgroup	HINTS 2007	HINTS 2005
Male	66.4%	**61.3%**
Female	70.6	**60.9**
18-34 years old	80.3	**74.4**
25-49 years old	76.0	**67.4**
50-64 years old	68.4	**59.3**
65-74 years old	45.1	**32.7**
75 and older	21.6	**17.6**
Less than high school	27.0	**22.9**
High school graduate	56.8	**49.1**
Some college	80.4	**74.1**
College graduate	91.0	**87.2**
Hispanic	**49.3**	**36.2**
Non-Hispanic Black	**56.8**	**52.5**
Non-Hispanic White	**75.0**	**68.4**
All Other	**74.2**	**60.9**

Westat

The Impact of Smartphones

- Couper et al. (in press): “Overall, we estimate that 82.9% of the target NSFG population (age 15-44) has Internet access, and 81.6% has a smartphone. Combined, this means that about 90.7% of U.S. residents age 15–44 have Internet access, via either traditional devices or a smartphone.”
- Still, find differences in coverage by age, race, income, and education

The Impact of Smartphones—2

	NHIS	Pew
18-29	92%	99%
30-49	85	96
50-64	76	87
65 +	51	64
< HS	--	68
HS	--	81
College Grad	--	98

- Even controlling for demographics, find biases due to noncoverage on some NSFG variables

Westat

Selection Bias

Who joins web panels?

- Information is sketchy
- We do have some information on *why* they join (Poynter and Comley, 2003; Baker, Blumberg, Brick, Couper, Courtright et al. 2010); the most common reasons are
  - the incentives (which are paltry)
  - curiosity
  - enjoyment of surveys (!)
- Must have lots of time (part-time workers, people not in the labor force)
- They are mostly recruited online; must be highly active on the Internet
- They are peculiar in one way—they do lots and lots of surveys
- Miller (2006): 30 percent of all online surveys done by highly active participants, who made up just 0.25 percent of the population, belong to seven panels on average, and complete a survey nearly every day(!)

Types of Nonresponse

Bosnjak distiguishes

- **Complete responders**: answer all questions in the survey.
- **Unit nonresponders**: do not respond to the request for participation. This may include those who visit the survey's welcome page (thereby providing some evidence that they received the invitation) but did not proceed to the survey itself
- **Drop-outs**: answer some of the questions, but break off before reaching the end of the survey (aka called abandonments, breakoffs, or partial responses)
- **Lurkers**: these are a special category who viewed the survey questions without answering them
- **Item nonresponders**: those who reach the end of the survey without answering all items

Westat

What are Web Response Rates Like?

- Two recent meta-analyses compare response rates for web surveys with other modes; probability samples
- Lozar Manfreda and colleagues (2008) examine 45 experimental comparisons between web and other surveys modes (mostly mail)
- On average, response rates to the web surveys were 11 percentage points lower than those in the alternative mode; when they just looked at mail-web comparisons, the average difference was 12 percentage points.
- Shih and Fan (2008) restricted their analysis to 39 studies directly comparing web to mail. An average unweighted response rate of 34% for web surveys and 45% for mail surveys, for a weighted difference of 11 percentage points

What are Participation Rates Like?

- Tourangeau, Conrad, and Couper (2013, pp. 42-43) report:

  We have seen participation rates decline from a high near 20 percent in 2002 to the low single digits since 2006, with a survey done in June-July 2010 yielding a participation rate of just 1 percent. Similarly, in 2008, one of our surveys required invitations to almost 62,000 members to yield 1,200 completes, for a 1.9 percent participation rate. With one panel claiming about 1.2 million U.S. members at the time, this meant that about one in twenty of all panel members were invited to that survey

- In addition to low participation rates, web surveys are subject to higher rates of breakoff than most other modes of data collection
- In the last five web surveys by Tourangeau, Conrad, and Couper (2013), all involving non-probability panels, the average breakoff rate was 22 percent

Westat

Impact on Web Estimates

	Unweighted Estimates			
	RDD Sample	Knowledge Networks	Harris Interactive	CPS (March 2000)
Education				
Some high school	7.0%	6.7%	2.0%	16.9%
High school graduate	31.3	24.4	11.8	32.8
Some College	19.6	32.3	36.6	19.8
College +	42.1	36.6	49.5	30.5
Sample size	1504	4925	2306	--
Age				
18-24	10.0%	7.8%	8.0%	13.2%
25-34	17.9	19.1	21.2	18.7
35-44	24.5	25.8	21.5	22.1
45-54	20.7	23.0	27.9	18.3
55-64	12.1	12.4	15.5	11.6
65 and older	14.9	11.9	5.8	16.1
Sample Size	1496	4923	2306	--
	Weighted Estimates			
	RDD Sample	Knowledge Networks	Harris Interactive	CPS (March 2000)
Education				
Some high school	17.1%	12.3%	7.9%	16.9%
High school graduate	32.7	33.5	36.5	32.8
Some College	19.8	28.5	26.9	19.8
College graduate +	30.3	25.6	28.8	30.5
Sample size	1504	4925	2250	--

Westat

Yeager et al. (2011)

- Compared 7 non-prob panels to an RDD survey and prob web panel; examined 13 benchmarks (mostly from other FTF surveys)

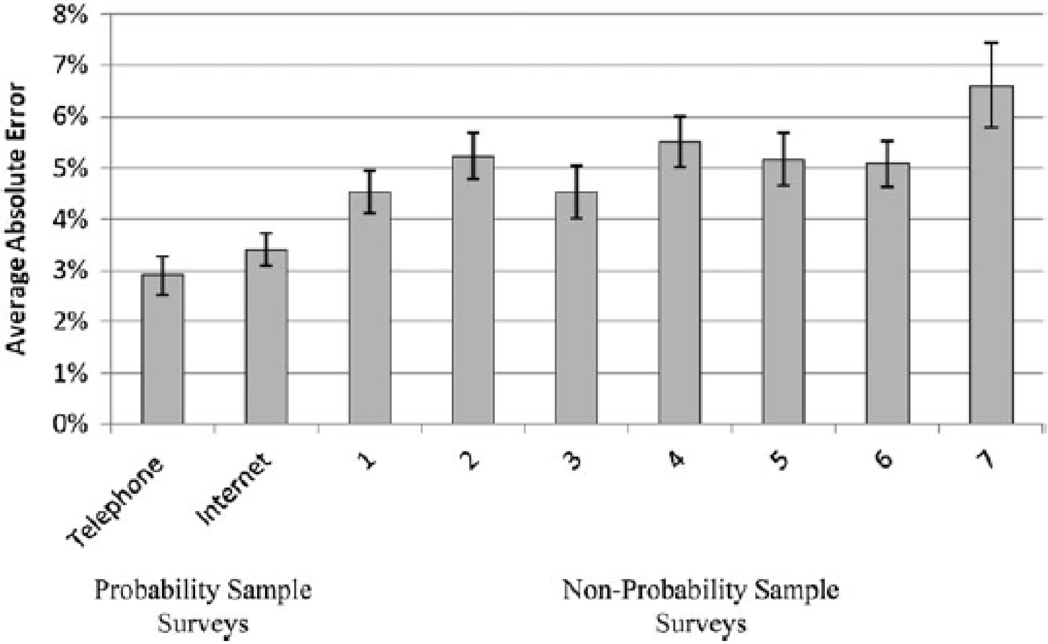


Figure 1. Average Percentage Point Absolute Errors for Commissioned Probability and Non-Probability Sample Surveys across Thirteen Secondary Demographics and Non-Demographics, with Post-Stratification.

Kennedy et al. (2018)

- Error in 2016 election polls done in final 13 days before the election

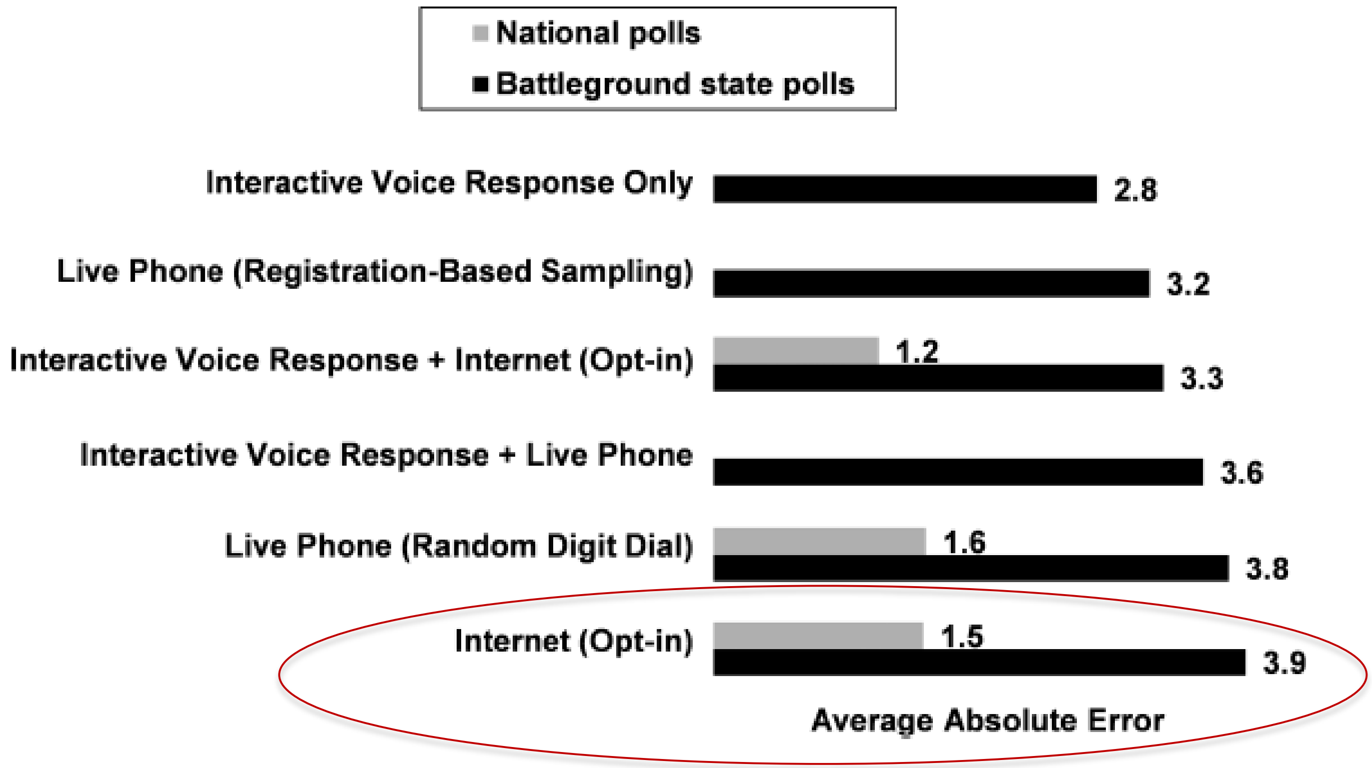


Figure 4. Average absolute error for 2016 general election polls, by design. Figures based on polls conducted during the final 13 days. Sample

Westat

Accentuation

- Coverage, enrollment in panels, and nonresponse are correlated phenomenon
- The propensity to be covered related to the propensity to join a panel and both related to the propensity to respond
- All three driven by similar variables?
  - Demographic variables (age, education, race, income, urbanicity)
  - Attitude—engagement with politics, the world at large
- The product of the three produces large biases

Can Weighting Fix Things?

- Calibration weighting
  – Post-stratification, or ratio adjustment
  – Raking
  – GREG weighting
- Propensity weighting

Application to Web Panels

- Use propensity weighting to calibrate web panels estimate to another survey (typically RDD estimate)
- Calculate probability of being in RDD sample or web sample
- Use estimated probability to adjust weight
- Schonlau et al. (2007)
  - Use attitude and other items to adjust web sample (Harris Internet panel) to RDD sample
  - Strong ignorability:

$$Y_{Web}, Y_{RDD} \perp Z \mid X$$

Westat

Weighting and Web Panels

Studies Evaluating Statistical Adjustments for Web Surveys

Study	Calibration Survey/ Web Survey(s)	Adjustment Method	Results		
			n of Estimates (Outliers)	Mean (Median) Reduction in Bias	Mean (Median) Relbias after Adjustment
Berrens, Bohara, Jenkins-Smith, Silva, & Weimer (2003)	RDD Survey/				
	Harris Interactive (January)	Raking	13 (0)	10.8 (19.4)	26.6 (8.3)
	Harris Interactive (July)	Propensity scoring	13 (2)	31.8 (36.7)	17.1 (4.7)
	Knowledge Network	Raking	13 (0)	-3.0 (-2.3)	20.6 (15.9)
Dever, Rafferty, & Valliant (2008)	Full Michigan BRFSS/ BRFSS Internet Users	GREG estimator (7 covariates)	25 (0)	23.9 (70.0)	4.3 (2.3)
Lee (2006)	Full General Social Survey/ GSS Internet users	Propensity scoring	2 (0)	31.0 (31.0)	5.4 (5.4)
Lee & Valliant (2009)	Full Michigan BRFSS/ BRFSS Internet Users	Propensity scoring (30 covariates)	5 (0)	62.8 (60.8)	5.8 (6.9)
		Propensity scoring plus GREG estimator		73.3 (80.8)	4.3 (3.9)
Schonlau, van Soest, & Kapteyn (2007)	RDD Survey/ Rand web panel	Propensity scoring (demographic variables)	24 (5)	24.2 (24.6)	21.1 (14.4)
		Propensity scoring (all variables)	24 (3)	62.7 (72.6)	10.3 (3.7)

Note: Reduction in biases and relative biases (Relbias) are expressed as percentages. Means in the last two columns are computed after deletion of outliers; the medians include all observations.

Summary: Case 1

- Coverage, selection, and nonresponse errors in non-prob web panels tend to push estimates in the same direction; the combination is worse than any single source
- Accentuation seems to occur
- Two other examples
  - Antoun et al.: PC vs. smartphone samples of LISS panel members; noncoverage (not having a smartphone) and nonresponse push estimates in same direction
  - Lundquist and Särndal (2013): Continuing the same fieldwork strategy (repeated callbacks by telephone) reduced the representativeness of the sample
- Weighting helps but doesn't solve the problems of correlated coverage, selection, and nonresponse errors
- Weighting is good, but balanced data are better

Westat

Case 2: Nonresponse and Measurement Error

Introduction

- Early paper by Cannell and Fowler (1963) raised the possibility that low-propensity respondents (in their case, R's who took longer to recruit) provide worse data
- The notion that similar motivations involved in responding at all and giving accurate answers developed more fully by Bollinger and David (2001):
  - "We hypothesize that a latent variable—propensity to cooperate—determines both response error and missed interviews" in SIPP (p. 129)
  - They found "Error-prone households [misreporting about FS recipiency] are more likely to miss interviews [later rounds of SIPP] than correct reporters"

Olson’s Review

- The “latent co-operation continuum” is probably still the most popular explanation for the potential link between nonresponse error and response error (see also Yan and Curtin on the “response continuum”)
- Still, Olson’s excellent review distinguishes seven possible hypotheses regarding a link
  - Cooperativeness (agreeing to participate and willingness to make an effort to provide good data)
  - Reactance
  - Interest in the topic, positive views about the sponsor
  - R characteristics
  - Research importance
  - Self-perception
  - Changes in survey protocols (Mode switches later in the field period)

Westat

Two other hypotheses

- Altruism/social desirability: People most at risk of large measurement errors are less likely to cooperate (may be harder to contact as well); Tourangeau, Groves, and Redline (2010)
- Commitment: Deciding to take part changes your motivation
- The ratio of theories to findings may be a little off!!

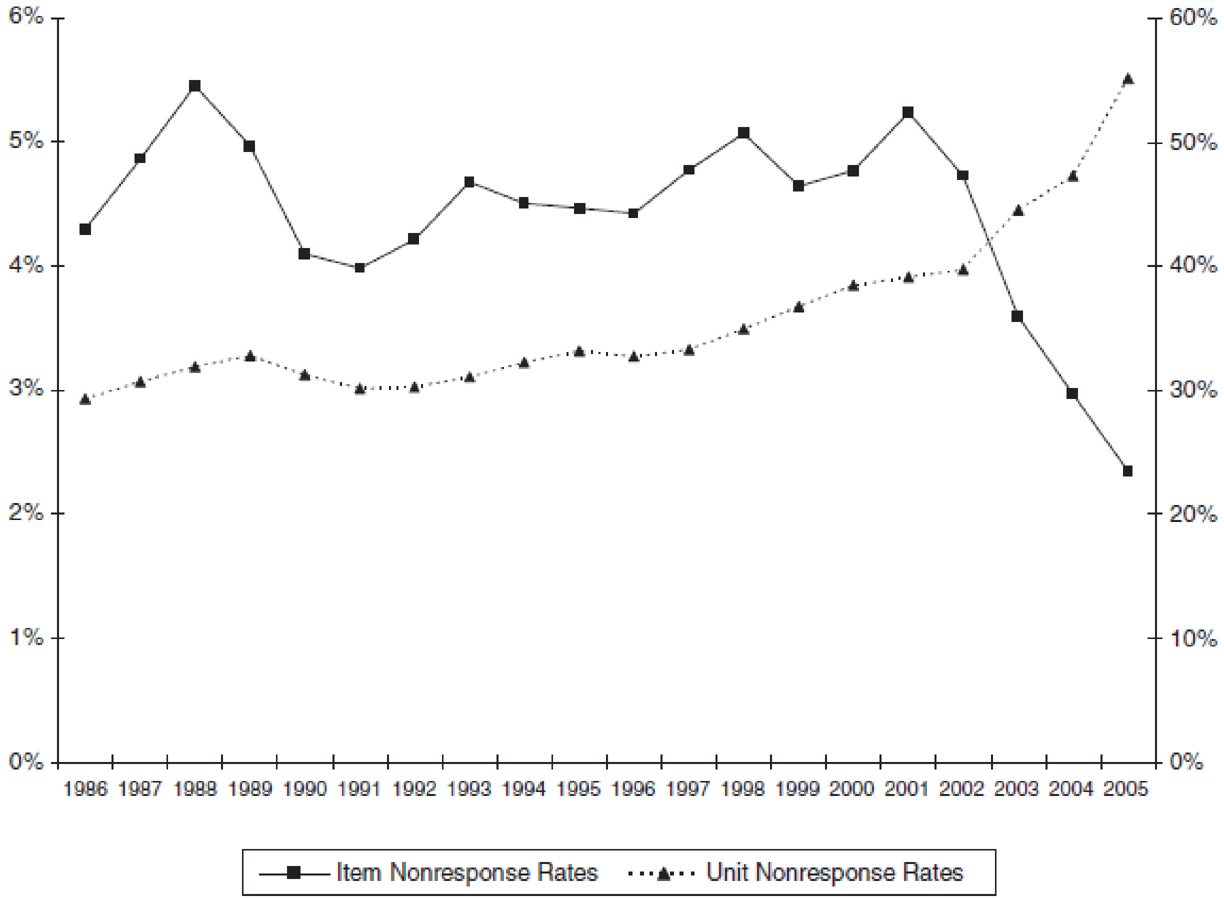
Some Findings

- Olson's review
  - More follow-up attempts (calls or mailings) associated with higher levels of question-specific item nonresponse; weaker evidence when overall item nr rates examined
  - Similarly, converted refusers prone to high item nr both for specific items and overall

Yan and Curtin (2010)

- Examined level of effort and item non-response in SCA
- At the aggregate level, both changing over time

Item nonresponse rates and unit nonresponse rates, 1986–2005



Yan and Curtin (cont'd)

- Both initial refusal and more call attempts associated with higher levels of item-nr (initial refusers had .6 percent higher item nr)
- Across months of the survey, the higher the unit nr rate the lower the item nr rate
- Rs with high levels of item nr less likely to do second interview
- Consistent with Olson's review

Fricker and Tourangeau

- Examined response propensities and response quality in the CPS and ATUS
- Here's a representative finding

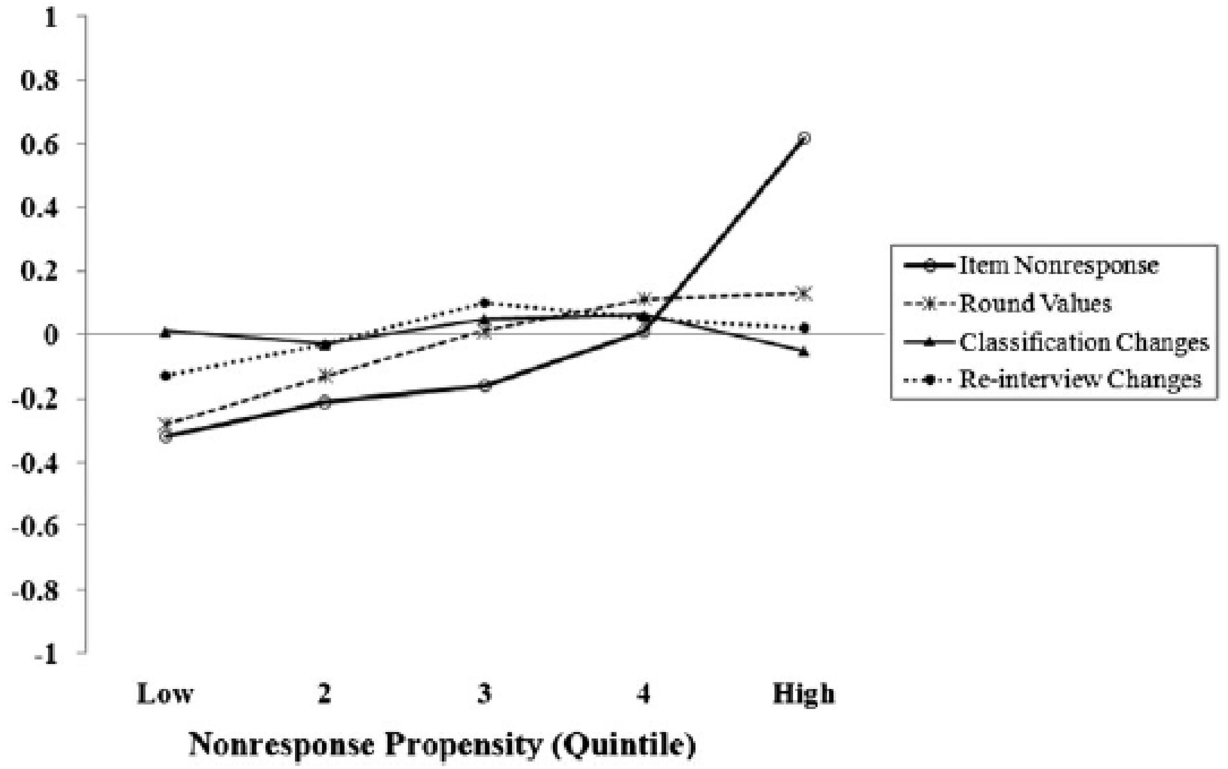


Figure 1. Relationship of CPS Data-quality Indicators (in Standard Deviation Units) to CPS Nonresponse Propensity.

Westat

Fricker and Tourangeau (cont'd)

- CPS R's with high item nr less likely to respond in ATUS
- ATUS propensity related to number of activities reported

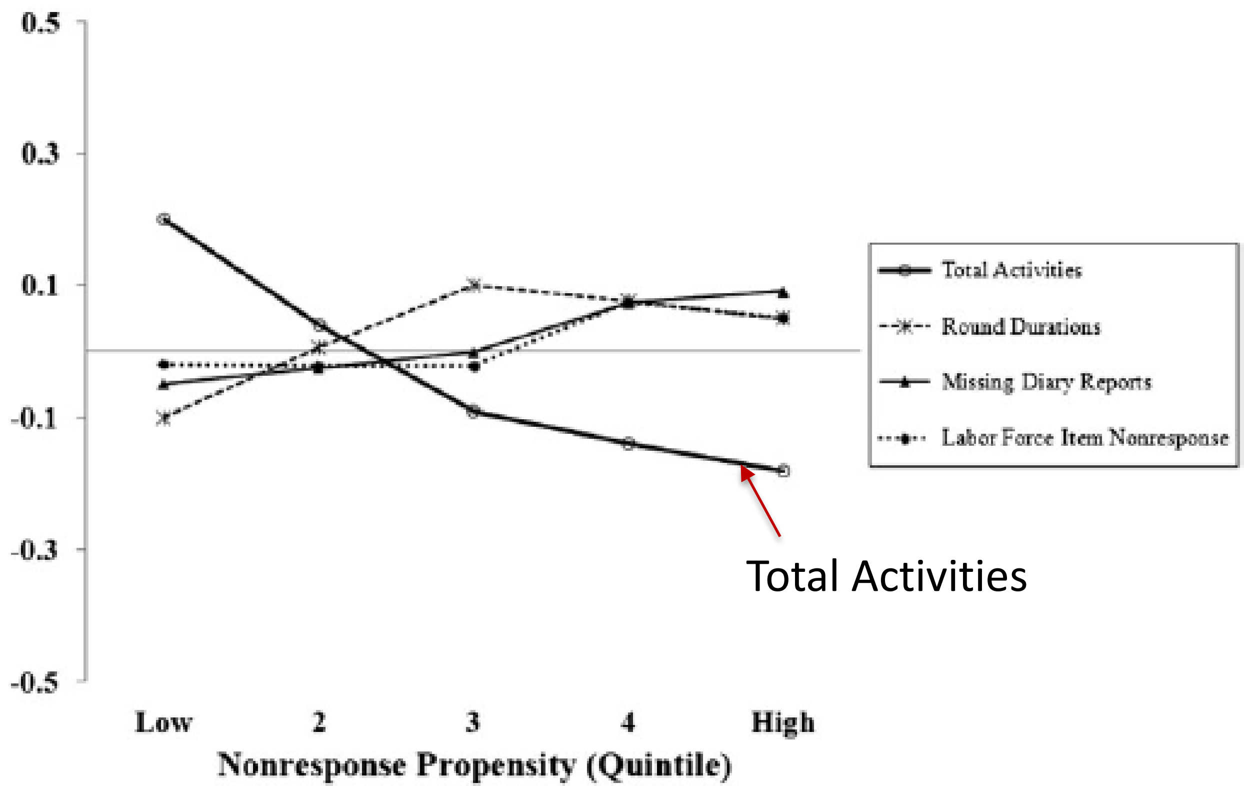


Figure 3. Relationship of ATUS Data-quality Indicators (in Standard Deviation Units) to ATUS Nonresponse Propensity.

Tourangeau, Groves, and Redline (2010)

- Direct assessment of measurement error via comparison to gold standard
- Sample of Maryland residents who are registered to vote; sample stratified by voter status
- Experimentally varied mode; have true scores (from frame on key variables)
- Response rates (overall 34%) reflect incentive (44% vs. 23%) and voter status (41% vs. 26%)

Bias Estimates

Estimated Percentage of Voters in 2006

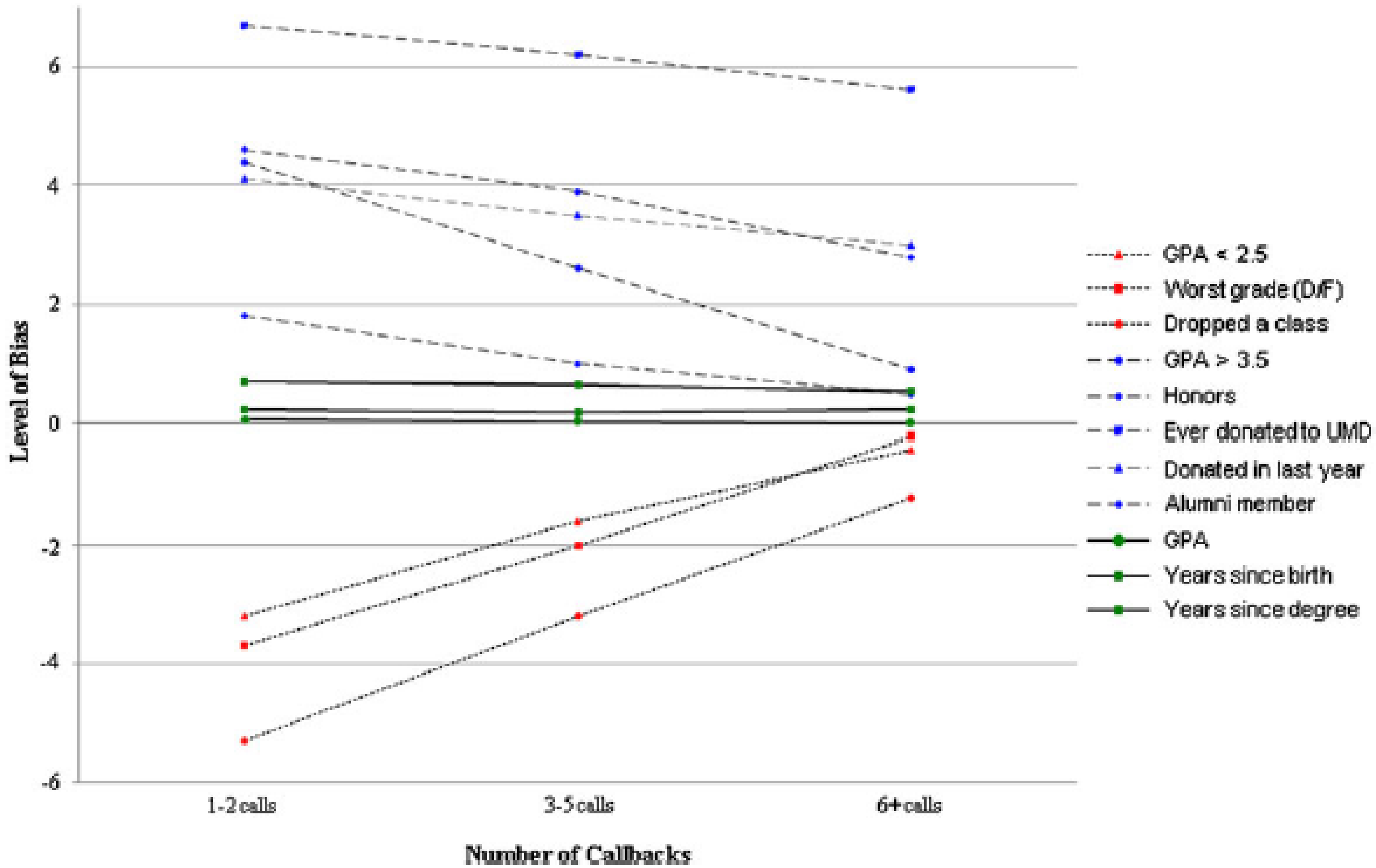
Subgroup	Entire Sample (Frame Data)	Respondents (Frame Data)	Respondents (Survey Reports)	Bias	
				Nonresponse	Measurement
Overall	43.7 (2689)	57.0 (904)	76.0 (895)	13.3	19.0
Topic					
Politics	42.6 (1346)	58.5 (441)	77.4 (438)	15.9	19.4
Health	44.7 (1343)	55.5 (463)	74.6 (457)	10.8	18.9
Incentive					
$5	43.4 (1349)	54.8 (591)	75.9 (586)	11.4	21.1
$0	44.0 (1340)	61.0 (313)	76.0 (309)	17.0	15.0
Mode					
Telephone	43.2 (1020)	57.4 (350)	79.4 (345)	17.0	15.2
Mail	43.9 (1669)	56.7 (554)	73.8 (550)	14.2	22.0

Tourangeau, Groves, and Redline (cont'd)

- Those at highest risk of reporting error (non-voters) less likely to respond; nr and measurement errors in same direction
- Similar findings in Kreuter, Presser, and Tourangeau (2008)—alumni with academic problems less likely to respond, more likely to misreport if they do respond
- A further wrinkle: Switch from CATI to lower propensity/higher accuracy modes (IVR and web) in Kreuter et al. study produced partially offsetting biases—higher nonresponse bias but lower measurement error (Sakshaug, Yan, and Tourangeau, 2010)

Sakshaug, Yan, and Tourangeau (2010)

- Cases reached more easily had more socially desirable outcomes



Westat

Summary: Case 2

- Propensity for unit and item nonresponse seem related
- Both converted refusals and cases requiring more calls more likely to provide missing data
- Case for increased measurement error is unclear; Fricker found some evidence in ATUS
- Overall, the evidence doesn't seem to fit the idea that reluctant Rs more likely to satisfice
  - Medway and Tourangeau (2015) examined prepaid cash incentive $5 (versus $) in a phone survey
  - They found significant differences between the incentive and control groups only for two of these 11 indicators
  - Respondents who got the incentive had less item nr but spent less time per question

Westat

Summary: Case 2 (cont'd)

- Strong case for a causal path involving SD bias—those most at risk for large reporting errors also more likely not to respond

Conclusions

Conclusions

- The world would be a lot simpler if the different forms of error were unrelated!
- With nonprob web surveys, it appears that there are common demographic correlates of coverage, enrollment, and response propensities; there may be common basic dispositional correlates as well (engagement in the modern world)
- As a result, the different sources of error accentuate each other
- Weighting helps; data are better

Conclusions—2

- It seems clear item and unit nr are related; those at risk of not responding at all are also at risk of providing missing data
- People in the socially undesirable categories less inclined to respond, but more inclined to misreport

- TSE lives!! Thank you, Morris!
-

Thank you very much!!

Westat