Adaptive and Responsive Designs: A Review and Assessment

Roger Tourangeau, Mike Brick, Sharon Lohr, and Jane Li

December 3, 2015

Federal Committee on Statistical Methodology Research Conference

Background

- Trouble with the NSFG Cycle 6

- Lower than expected response rate

- How to cope with it?

  – Two-phase design to concentrate resources and address non-response bias

  – Similar solution in prior round of NSFG (Judkins, Mosher, and Botman, 1991)

  – Hansen and Hurwitz discussed the use of two-phase sampling for nonresponse in 1946

NSFG-6 (Groves and Heeringa, 2006)

- Three phase design: Phase 1 looked at best number of callbacks

- After main phase (phase 2), a final phase
  - Retained about a third of the cases from Phase 2
  - Used the most productive interviewers from the prior phases
  - Relaxed the rules for collecting screening data from proxy informants
  - Used a prepaid incentive of $5 for the screener,
  - increased the incentive amount for the main interview (offering an additional $40 over the Phase 2 amount)

Adaptive and Responsive Survey Design

Phases in NSFG 6

- Phase 1: decided to cap call attempts at 14 during the second phase; estimates unlikely to change enough to justify the added costs of further callbacks

- Phase 2: two propensity (discrete hazard) models developed, based on frame variables and variables derived from paradata
  - likelihood a household would be screened on the next call
  - the likelihood that a sample case would complete a main interview on the next call

- Phase 3: SSUs grouped into 16 strata, based on the number of active cases in the SSU at the end of Phase 2 and the total estimated propensities for those cases; segments in the highest quartiles on each of these variables were sampled at higher rates than those in the lower quartiles

Goals of Final Phase

- Focus effort on SSUs with many possible cases

- Focus effort where cases have high (not low) propensities

- Implied goals: Maximize RR and sample size, while controlling costs

Set Off a Flood of Studies

- Why? Not really much that was new
  - Real-time use of propensity models
  - New rhetoric (ACS-type design):

    Responsive designs are organized about design phases. A design phase is a time period of a data collection during which the same set *of sampling frame, mode of data collection, sample design, recruitment protocols, and measurement conditions* are extant. For example, a survey may start with a mail questionnaire attempt in the first phase, follow it with a telephone interview phase on non-respondents to the first phase and then have a final third phase of face-to-face interviewing…Note that this use of 'phase' includes more design features than merely the sample design, which are common to the term 'multi-phase sampling.' (Groves and Heeringa, 2006, pp. 440-441)

Adaptive and Responsive Survey Design

Follow-Up Studies

- Since then at least five experimental studies and four simulation studies

- Twin problems of rising costs, falling response rates

- In Canada, additional problem of cap on number of calls

- Minimize the impact of these problems on survey quality (that is, variance and non-response bias)

Follow-Up Studies: 2-Phase Design

- NISVS (Peytchev, Baxter, and Carley-Baxter, 2009)

- Telephone study with a two-phase design

- Phase 1 RDD sample with supplemental sample of listed telephone numbers

- Numbers that could be matched to an address were sent an advance letter with a $1 incentive; all cases were promised either $10 or $20. Phase 1 Also varied description of topic

- Most cases received up to twenty calls during the first phase, but some got even more. In addition, refusal conversion was attempted during Phase 1. Overall, this phase produced a response rate 28.5 percent.

- Phase 2: Subsample cases, questionnaire went from 30 to 14 minutes, offered a prepaid incentive of $5 and a conditional incentive of $20. This phase produced a response rate of 9.8 percent (or 35.5 percent overall).

Peytchev et al. (2009) Results

- Look at 6 key estimates for males and females

- Compare phase 1 early versus late (6+ attempts): No significant change in estimates

- Phase 1 versus Phase 2: Male respondents were more likely to report victimizations in Phase 1 than in the Phase 2, with significant differences on four of six estimated victimization rates

- Impact of Phase 1 refusal conversion: male cases who never refused and those who were converted after refusing; like the Phase 2 male respondents, the converted Phase 1 male refusals also showed significantly lower victimization rates on four of six key estimates

Basic Strategies

- Two (or more) phase designs

- Case prioritization—vary level of effort for different types of cases

- Tailored data collection (e.g., different modes for cases with different anticipated propensities)

Options for Phase 2 (or Adaptive Designs)

- Stop working some cases (as in subsampling for Phase 2 or stop rules explored by Särndal and Lundquist)

- Increase incentives/tailor incentives

- Shorten questionnaire

- Switch modes/tailor modes

- Reassign to best interviewers (Luiten and Schouten, 2013)
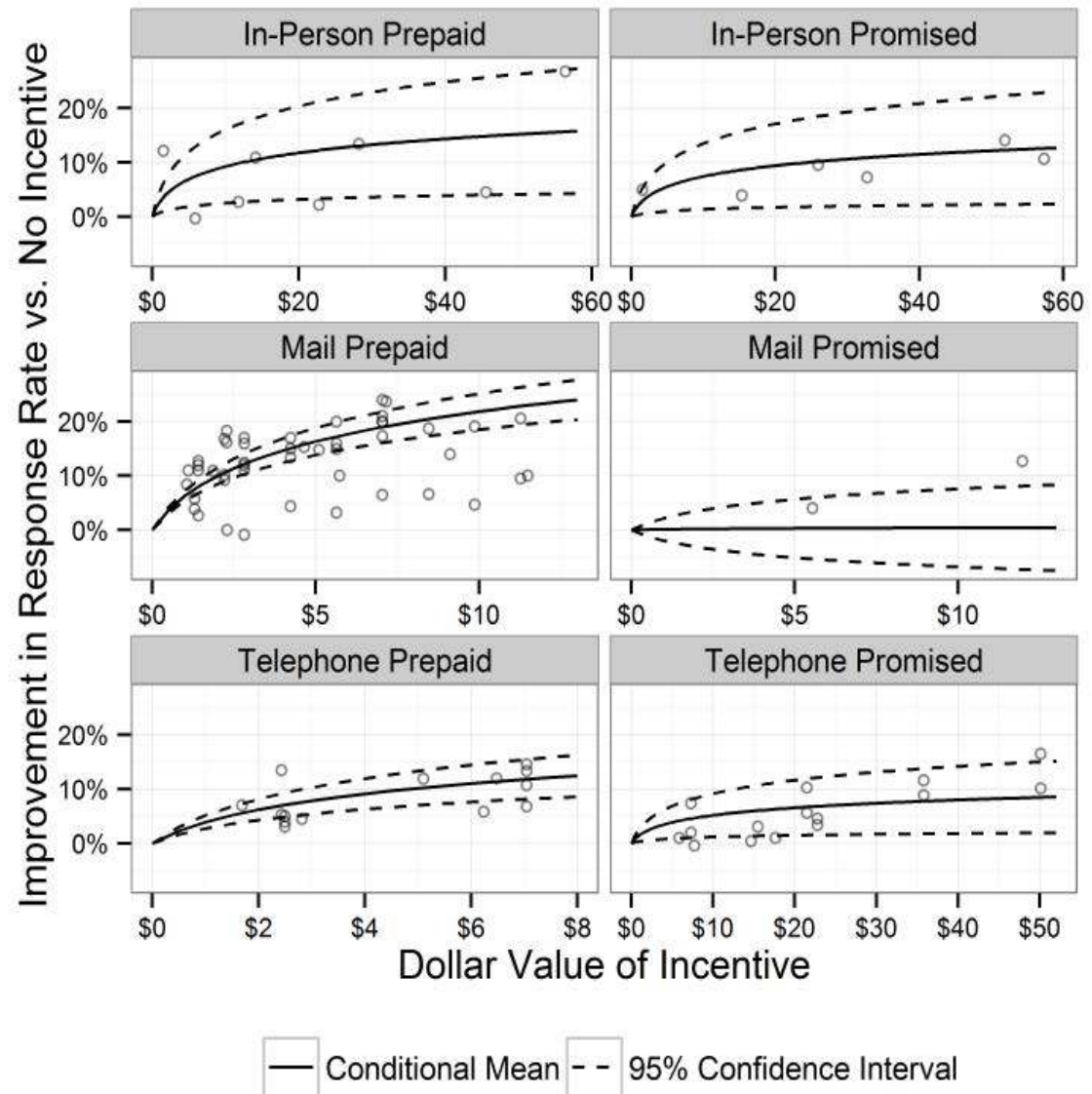
- Prioritize cases

Pros and Cons: Dropping Cases

- Dropping cases definitely is easy to implement, especially in phone surveys

- Generally easier cases are dropped

  - Reduced response rate

  - Smaller sample size (Less impact on effective sample size)

  - Response rate maximization=Sample size maximization

Incentives

Diminishing returns (Mercer et al., 2015)

Long-term effect on survey climate?

Shortening the Questionnaire

- Loss of precision
- Is there an optimal point where there is a reasonable chance to impute remaining items?

Switching Modes

- Start with cheap mode for everyone

- Tailor mode choice to subgroup—assign high propensity cases to cheaper, low propensity mode (as in Luiten and Schouten, 2013)—increase balance, decrease costs

- Two potential issues

  - Mode effects on measurement: An exaggerated concern?

  - Kolenikov and Kennedy (2014) examine 297 variables,19 showed significant differences between the telephone only and web groups but only four of these remained significant after adjusting for demographic differences

  - Starting with low propensity mode may lower ultimate response rate—having refused in one mode, cases may be more likely to refuse in a second mode

Switching Modes—2

Study	Population	Mode Sequence	Overall Response Rate
Dillman et al. (2009)	Volunteers	Mail then Telephone Telephone then Mail Web then Telephone	82.8 (2000) 80.4 (2999) 47.7 (2000)
Holmberg, et al. (2010)	Stockholm residents (18-65)	Mail then Web Web then Mail	74.8 (1993) 72.2 (5991)
Messer and Dillman (2011)	Residents of Washington state	Experiment 1a Mail then Web ($5 prepaid) Web then Mail ($5 prepaid) Mail then Web (No incentive) Web then Mail (No incentive) Experiment 1b Mail then Web ($5 prepaid) Web then Mail ($5 prepaid)	 53.6 (474) 44.4 (1018) 40.1 (648) 25.7 (643) 55.0 (476) 43.8 (1332)
Millar and Dillman (2011)	WSU undergrads	Experiment 1 Mail then Web Web then Mail Web then Mail—with two more email invitations	 53.2 (681) 50.2 (676) 64.5 (678)
Smyth et al., 2010	Residents two WA towns	Mail then Web Web then Mail	71.1 (367) 55.1 (566)

Reassigning/Prioritizing Cases

- Easier to do with telephone

- Potential cost tradeoffs in FTF (travel in better interviewers)

- Interviewer compliance also an issue: Wagner et al. (2012)—16 experiments in NSFG

  - Significantly increased number of calls to priority cases in seven of 16 experiments

  - Significantly increased response rate in two experiments

  - Maybe be optimizing the wrong thing (trip to the field vs. individual case)

Who to Prioritize?

- To date, "inverse" case prioritization—equalize propensities after intervention

$$k = p_1 + (1 - p_1)p_2$$

$$p_2 = \frac{k - p_1}{1 - p_1}$$

- Target cases that reduce bias, improve balance the most

$$V_i = \hat{\rho}_i W_i D_i \,,$$

$$\text{where } D_i = \sqrt{(\mathbf{x}_i - \overline{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}})} \,.$$

Some Issues

- Unfortunately, we understand how to lower propensities better than to raise them
- Still, stopping rules may be best method for equalizing propensities and lowering costs; it would be useful to know how costs vary as function of time in the field (everyone thinks the last cases are much more expensive than early, easy cases)
- Balance the sample/equalizing the propensities:
  - During data collection vs. after the fact
  - It would be nice to have further empirical demonstrations (Särndal and Lundquist, 2014) as well as theory showing that balance during data collection helps with both bias and variance
- Better understanding of practical limits of weighting (cf. Tourangeau, Conrad, and Couper, 2013)

Adaptive and Responsive Survey Design

Some Issues—2

- Most important: It would be useful to understand better what has happened to the survey climate (world-wide) and what if anything can be done to improve it

- We rushed to cure the problem without stopping to diagnose it

Adaptive and Responsive Survey Design

Conclusions

- To return to my earlier question: What is new here?

- Three big shifts
  - Statistical models (propensity models) supplementing if not supplanting FI and supervisor judgment
    - Are we missing anything?
    - Expert system approach?
  - Difference between field and telephone narrowing (CARI, GPS)
  - Design focused less on sampling error and more on bias (especially non-response bias)

Adaptive and Responsive Survey Design

Thank You!!