Responsive Designs: Some Recent Research Findings

6th International Workshop For Advances In Adaptive And Responsive Survey Design

Michael Elliott[1,2,3], James Wagner[2,3], Brady West[2,3], Stephanie Coffey[3,4]

[1]Department of Biostatistics, University of Michigan
[2]Survey Methodology Program, Institute for Social Research
[3]Joint Program in Survey Methodology, University of Maryland
[4]US Census Bureau

Adaptive, Responsive, Tailored Design

- The “long crisis” in probability sampling – increasing costs and declining response rates – have lead to the formalization of ideas around Adaptive, Responsive, and Tailored design (Murphy et al. 2018).
- Overlapping concepts.

Adaptive, Responsive, Tailored Design

- The “long crisis” in probability sampling – increasing costs and declining response rates – have lead to the formalization of ideas around Adaptive, Responsive, and Tailored design (Murphy et al. 2018).
- Overlapping concepts.
- Adaptive design: change protocol based on known information about a potential respondent.
  - Often set in place for a given round of data collection based on information gained from pilot studies or previous waves.

Adaptive, Responsive, Tailored Design

- The "long crisis" in probability sampling – increasing costs and declining response rates – have lead to the formalization of ideas around Adaptive, Responsive, and Tailored design (Murphy et al. 2018).
- Overlapping concepts.
- Adaptive design: change protocol based on known information about a potential respondent.
  - Often set in place for a given round of data collection based on information gained from pilot studies or previous waves.
- Responsive design: change protocol based on preliminary information obtained during data collection.
  - Updated in real time based on data collected as part of the main study.

Adaptive, Responsive, Tailored Design

- The “long crisis” in probability sampling – increasing costs and declining response rates – have lead to the formalization of ideas around Adaptive, Responsive, and Tailored design (Murphy et al. 2018).
- Overlapping concepts.
- Adaptive design: change protocol based on known information about a potential respondent.
    - Often set in place for a given round of data collection based on information gained from pilot studies or previous waves.
- Responsive design: change protocol based on preliminary information obtained during data collection.
    - Updated in real time based on data collected as part of the main study.
- Tailored design: protocol varied to individual preferences or to an overall design.

Adaptive, Responsive, Tailored Design

- The "long crisis" in probability sampling – increasing costs and declining response rates – have lead to the formalization of ideas around Adaptive, Responsive, and Tailored design (Murphy et al. 2018).
- Overlapping concepts.
- Adaptive design: change protocol based on known information about a potential respondent.
  - Often set in place for a given round of data collection based on information gained from pilot studies or previous waves.
- Responsive design: change protocol based on preliminary information obtained during data collection.
  - Updated in real time based on data collected as part of the main study.
- Tailored design: protocol varied to individual preferences or to an overall design.
- Goal: good question! "Minimize cost and maximize information". How to formalize?

Adaptive Design: Non-respondent Sampling

- Concept embedded almost from the beginning of probability sampling.

Adaptive Design: Non-respondent Sampling

- Concept embedded almost from the beginning of probability sampling.
- Releasing replicates to control sample size in the presence of non-response.

Adaptive Design: Non-respondent Sampling

- Concept embedded almost from the beginning of probability sampling.
- Releasing replicates to control sample size in the presence of non-response.
- Subsampling non-respondents (Hansen and Hurwitz 1946)

Adaptive Design: Non-respondent Sampling

- Concept embedded almost from the beginning of probability sampling.
- Releasing replicates to control sample size in the presence of non-response.
- Subsampling non-respondents (Hansen and Hurwitz 1946)
  - Begin with low-cost mode with high non-response (mail) and switch to high cost mode with low nonresponse (face to face) for a random subsample of non-respondents.
  - Under SRS and equal variance across the first-stage respondent and nonrespondent strata, subsample fraction of non-respondents $r = \sqrt{\frac{C_1 + RC_2}{RC_3}}$, where $C_1 + RC_2$ is the total cost associated with conducting the low-cost mode and RC_3 is the total cost of the high cost mode.
  - Weight by $1/r$ to account for underrepresentation (assuming perfect response).

Adaptive Design: Subsampling callbacks

- Elliott et al. (2000) extended this to multicallback settings, with or without change of mode.
- Consider dropping a fraction of respondents α_m at the mth callback, $m = 1, ..., K$.

Adaptive Design: Subsampling callbacks

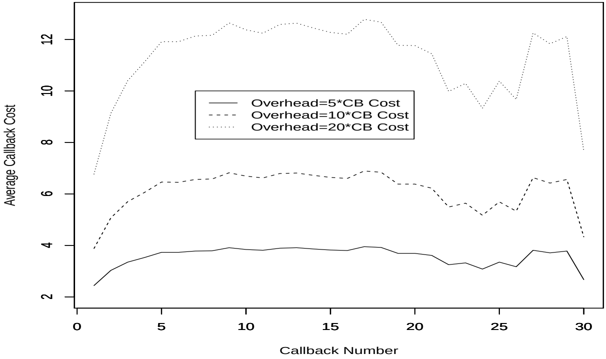
- Elliott et al. (2000) extended this to multicallback settings, with or without change of mode.
- Consider dropping a fraction of respondents α_m at the mth callback, $m = 1, ..., K$.
- Optimized by determining $\min_m \alpha_m$, $\alpha_m = min\left(\sqrt{\frac{(1-v_0^m)(1-v_1^m)}{v_0^m v_1^m}}, 1\right)$, where under certain conditions, v_1^m is the proportion of cost of remaining calls without subsampling and $1 - v_0$ is the proportion of interviews to be attempted without subsampling.

Adaptive Design: Subsampling callbacks

- Elliott et al. (2000) extended this to multicallback settings, with or without change of mode.
- Consider dropping a fraction of respondents α_m at the mth callback, $m = 1, ..., K$.
- Optimized by determining $\min_m \alpha_m$, $\alpha_m = min\left(\sqrt{\frac{(1-v_0^m)(1-v_1^m)}{v_0^m v_1^m}}, 1\right)$, where under certain conditions, v_1^m is the proportion of cost of remaining calls without subsampling and $1 - v_0$ is the proportion of interviews to be attempted without subsampling.
- Weight up cases by a factor of $1/\alpha_m$ for cases in the mth callback or later, and increase sample size to $n^{\alpha_m} = n^1[v_0^m + (1/\alpha)(1 - v_0^m)]$ to retain equal variance.
  - $\alpha_m < 1$ when $v_0^m + v_1^m > 1 \rightarrow$ when proportion of costs in the callbacks to be subsampled is less that the proportion of interviews remaining.

Adaptive Design: Subsampling callbacks

- Applied to 1992 National Comorbidity Survey: face-to-face survey with up to 29 contact attempts.
- Although costs increased over time and savings could be obtained by subsampling, it was relatively minor: subsampling at the 7th or 8th call, dropping between 23% and 30% of non-respondents reduced costs by 0.4-1.5%.



Adaptive Design: Subsampling callbacks

- Applied to 1996 American Communities survey: 2 mail attempts, 3 phone call attempts (if listed), 3 face-to-face attempts.
- Depending on assumptions about cost, major savings could be achieved: from 18% assuming telephone contact averages 2 times and face-to-face 10 times the cost of mail, to 40% assuming telephone averaged 5 times and face-to-face 50 times the cost of mail.

Adaptive Design: Mode Switching

- Calinescu et al. (2013) undertook a fairly comprehensive approach to optimizing a survey by allowing for switching among $m = 1, ..., M - 1$ modes or stopping ($m = M$) at each $t = 1, ..., T$ contact attempt.
- Formulate as non-linear optimization function. Subdivide the population into $g = 1, ..., G$ homogenous groups and maximize

$$\sum_g \sum_t \sum_m P_g f_g(t-1) x_g(t,m) p_g(t,m) r_g(t,m)$$

  - P_g=proportion of the sample in the group,
  - $f_g(t) = \prod_m [1 - x_g(t,m) p_g(t,m)] f_g(t-1)$=marginal probability that a contact attempt fails at time t for model contact probability $p_g(t,m)$ and mode indicator $x_g(t,m)$
  - $r_g(t,m)$=participation probability given contact.

Adaptive Design: Mode Switching

- Optimize with respect to $x_g(t,m)$.
- Constraints
  - $\sum_m x_g(t,m) \leq 1$ (only choose one mode at a given contact attempt).
  - $\sum_t x_g(t,m) \leq \overline{k}_g(m)$ (budget constraint for a given mode.)
- Too complex to solve directly, but can be reformulated as a Markov decision problem starting from time T and solving recursively.

Adaptive Design: Mode Switching

- Can implement by considering a variety of budget levels and the associated response rate under the optimal design.

Input data for group g.

Mode	Probability	t_1	t_2	t_3	t_4	t_5	t_6
Face-to-face	$p_g(t,m)$	0.3	0.4	0.8	0.2	0.3	0.7
	$r_g(t,m)$	0.9	0.7	0.3	0.8	0.8	0.6
Phone	$p_g(t,m)$	0.4	0.5	0.9	0.4	0.4	0.8
	$r_g(t,m)$	0.8	0.5	0.7	0.6	0.4	0.6

Input data for group g_2.

Mode	Probability	t_1	t_2	t_3	t_4	t_5	t_6
Face-to-face	$p_{g_2}(t,m)$	0.8	0.6	0.4	0.6	0.4	0.2
	$r_{g_2}(t,m)$	0.9	0.7	0.6	0.8	0.5	0.3
Phone	$p_{g_2}(t,m)$	0.7	0.6	0.5	0.6	0.5	0.4
	$r_{g_2}(t,m)$	0.8	0.6	0.5	0.6	0.4	0.2

B	Group	Time slot						Group response rate	Response rate
		t_1	t_2	t_3	t_4	t_5	t_6		
1000	g_1	0	0	0	0	0	0	0	0.322
	g_2	F2F	F2F	0	F2F	Ph	0	0.849	
1250	g_1	0	0	0	0	0	0	0	0.323
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	
1750	g_1	0	0	Ph	F2F	F2F	Ph	0.692	0.429
	g_2	0	0	0	0	0	0	0	
2250	g_1	0	0	Ph	0	0	0	0.63	0.701
	g_2	F2F	0	0	F2F	0	0	0.816	
2500	g_1	0	0	Ph	0	F2F	Ph	0.688	0.749
	g_2	F2F	F2F	0	F2F	Ph	0	0.849	
2750	g_1	0	0	Ph	F2F	F2F	Ph	0.692	0.752
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	
3250	g_1	F2F	0	Ph	0	0	Ph	0.745	0.782
	g_2	F2F	F2F	0	F2F	0	0	0.842	
3500	g_1	F2F	0	Ph	F2F	F2F	Ph	0.754	0.791
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	
4000	g_1	F2F	F2F	Ph	F2F	F2F	Ph	0.757	0.793
	g_2	F2F	F2F	F2F	F2F	F2F	Ph	0.851	

- Does this design introduce potential biases?

Responsive design

- In the examples above measures such as response probabilities and costs treated as known.
- Groves and Heeringa (2006) laid out an initial view of responsive design.
- Design phase: stable periods of data collection: mode, recruitment, non-contact and non-response follow-up, etc.
  - Embed experiments in early design phases: multiple vs. single adult selection.
- Phase capacity: stability of estimation within a given design phase.
  - Point estimate stabilized after a certain number of callbacks.

Responsive design: Stopping rules

- In practice phase capacity has often been implemented through stopping rules.
- Stop once data suggest changes in estimators are sufficiently small.

Responsive design: Stopping rules

- In practice phase capacity has often been implemented through stopping rules.
- Stop once data suggest changes in estimators are sufficiently small.
  - Rao et al. 2008: stop when wave and interaction between key predictors of nonresponse and wave are not significant, or when changes in key outcomes are sufficiently small.

Responsive design: Stopping rules

- In practice phase capacity has often been implemented through stopping rules.
- Stop once data suggest changes in estimators are sufficiently small.
  - Rao et al. 2008: stop when wave and interaction between key predictors of nonresponse and wave are not significant, or when changes in key outcomes are sufficiently small.
  - Wagner and Raghunathan (2010): predict future outcomes based on available data, then subsample a set of cases to collect actual values, and cease if the probability that the predicted mean under the imputed values and using the actual data is sufficiently small.

Responsive design: Stopping rules

- In practice phase capacity has often been implemented through stopping rules.
- Stop once data suggest changes in estimators are sufficiently small.
  - Rao et al. 2008: stop when wave and interaction between key predictors of nonresponse and wave are not significant, or when changes in key outcomes are sufficiently small.
  - Wagner and Raghunathan (2010): predict future outcomes based on available data, then subsample a set of cases to collect actual values, and cease if the probability that the predicted mean under the imputed values and using the actual data is sufficiently small.
  - Lewis (2017, 2019) replaces the idea of imputation in the Rao et al. and Wagner and Raghunathan approach with non-response weighting.
    - Cease data collection once differences in weighted estimators are no longer statistically significant.
    - Use power calculations to determine significance level (post-hoc?)

Responsive design: Case prioritization

- Attempt to reduce non-response bias by changing priorities for contact in process.
- Key insight comes from nonresponse bias of mean expressed as covariance between the survey variable y and the response propensity ρ divided by response rate (Bethlehem 2002):

$$\text{Bias}(\overline{y}_r) = \frac{\sigma_{y,\rho}}{\overline{\rho}}$$

- Focus in on reducing $\sigma_{y,p}$ rather than increasing $\overline{p}$ (Peytchev et al. 2010).
- R-indicator (Schouten et al. 2009) provides a sort of "surrogate" for response bias by focusing on the variability in the response propensities: $R = 1 - 2S_{\rho}$, $S_p^2 = (N-1)^1 \sum_{i=1}^N (\rho_i - \overline{\rho})^2$
    - $\hat{R}$ replaces S_ρ^2 with s_p, the sample variances of the estimated propensities; bias adjusted version available (Shlomo et al. 2012).

Responsive design: Case prioritization

- Schouten and Shlomo (2017) use *partial* R-indicators: $S^2(\rho \mid Z) = \sum_k P_k(\overline{\rho}_k - \overline{\rho})^2$ for a single categorical variable Z with K levels used in the overall prediction ρ.
- Find set of variables with statistically significant partial R indicators and re-stratify based on cross classification.
- Recompute partial R indicator and order by significance.
- Select the maximum number of cases available for followup based on cost constraints prioritizing strata with most significant partial R indicators.

Responsive design: Evaluation

- As pointed out by Tourangeau et al. (2017), the target of responsive designs has been somewhat inconsistent.
- Targeting high propensity respondents can increase bias while simultaneously increasing sample size for a given cost.
- Increasing the R-indicator by recruiting fewer high propensity subjects. can reduce response rates (while perhaps reducing response bias) and decrease sample sizes.
  - Bias-variance tradeoff
- Reducing R-indicator increased estimated non-response bias in Peytchev et al. (2010).

Responsive design: Evaluation

- R indicator uses the Cauchy-Schwartz inequality to note that response bias in bounded by the variance of the propensities.

$$\text{Bias}(\overline{y}_r) = \frac{\sigma_{y,\rho}}{\overline{\rho}} \leq \frac{S(\rho)S(y)}{\overline{\rho}} = \frac{(1-2R(\rho))_S(y)}{2\overline{\rho}}$$

- Non-linear relationships between probability of selection and outcome. means that bias is not a direct linear function of a given value of R however.
- If

$$Y_i = \alpha + \beta\rho_i + \varepsilon_i,\ \varepsilon_i \sim (0,\sigma^2),\ \varepsilon_i \perp \rho_i$$

  then we can show $Cov(Y,\rho) \approx \beta S^2(\rho)$ and $B(\overline{y}_r) \approx \frac{\beta}{\overline{\rho}}\left(\frac{1-R(\hat{\rho})}{2}\right)$.
- But if there is an additional non-linear $f(\rho_i)$ term in the mean function, then $B(\overline{y}_r) \approx \frac{\beta}{\overline{\rho}}\left(\frac{1-R(\hat{\rho})}{2} + Cov(f(\rho),\rho)\right)$.

Responsive design: Evaluation

- Compare $Y_i \sim N(\rho_i, 1)$ with $Y_i \sim N(\rho_i + 10\rho_i^4, 1)$, $i = 1, ..., 10000$.
  - Suppose $\rho_i \sim UNI(0, .5)$, with sample size of $n = 500$.
  - Standardized bias increases from .08 to .13 while R-indicator remains .76

Responsive design: Evaluation

- Two major issues in responsive design research currently open.
- What should we be targeting in our responsive designs?

- How can we best incorporate prior information in our design decisions?

Responsive design: Target

- Tourangeau et al. (2017) suggest focusing on cases that are most likely to affect final estimates, as a product of response propensity, sample weight, and distance from sample mean:

$$V_i = \rho_i w_i D_i,\, D_i = \sqrt{(\hat{y}_i - \bar(y))^T S^-1(\hat{y}_i - \bar(y))}$$

  or, accounting for cost, $V_i/\hat{c}_i$.
- That cost term shows up rather innocuously.
- How do we estimate that??? BIG issue going forward.

Responsive design: Prior

- Schouten et al. (2018) suggest using a Bayesian framework to incorporate prior information.
- Response propensities could be estimated from prior waves of a survey and/or from a literature review to construct an informative prior distribution that is then combined with initial estimates via Bayes rule to provide a posterior distribution to be used for targeting and design changes.
  - Prior wave of the survey gives logistic regression estimates of response propensity based on a set of covariates x_i as $\hat{\beta}_0$ with variance estimate $v(\hat{\beta}_0)$.
  - Current wave of survey gives logistic regression estimates of response propensity based on same set of covariates as $\hat{\beta}$ with variance estimate $v(\hat{\beta})$.
  - Point estimate of β is then obtained (assuming an approximate normal distribution for both) as

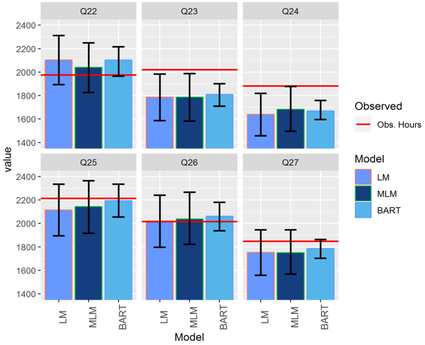
$$\left(v(\hat{\beta}_0)^{-1} + v(\hat{\beta})^{-1}\right)^{-1} \left(v(\hat{\beta}_0)^{-1}\hat{\beta}_0 + v(\hat{\beta})^{-1}\hat{\beta}\right)$$

Predicting Costs in Surveys

- No existing literature on cost prediction in surveys(!)
- Wagner et al. (2019) use a linear mixed model (LMM) and a Bayesian additive regression tree (BART) model (Chipman et al. 2010) to predict interviewer hours in the National Survey for Family Growth.
- Predict Phase 2 (two week non-response followup) costs based on Phase I data.
- Data from the first 22 quarters were used to make predictions that were assessed using data from the last 6 quarters.

Predicting Costs in Surveys

- Predictors included geography, local Census tract information, commercial estimates of age and income as a target household, and paradata based on interviewer observations, level of effort, and interviewer IDs.
- LMM suggested individual interviewers explain around 20-25% of overall variance.
- Both models produced reasonably good predictions:



Predicting Response using Bayesian Methods

- National Survey of Family Growth is a repeated cross-sectional survey collected in quarterly replicates.
  - Restricted to persons 15-49, and begins with a household screening interview for the presence of such persons.
- West et al. (2019) considering the use of informative priors to improve estimation of probability of responding to a screening interview.
- Elicited priors
  1. Using eight prior quarterly data collection periods.
  2. Using a literature review of propensity models that included similar covariates.

Eliciting Priors using Historical Data

- Fit discrete time-to-event models of the form

$$\log\left(\frac{P(Y_{it}) = 1 \mid X_{it}}{1 - P(Y_{it}) = 1 \mid X_{it}}\right) = \beta^T X_{it}$$

- Consider two approaches
  - Use the most recent period to obtain MLE $\hat{\beta}$ and associated variance estimate $v(\hat{\beta})$, and use prior $N(\hat{\beta}, v(\hat{\beta}))$.
  - Combine previous eight quarters by obtaining MLEs and associated variance estimates $\hat{\beta}_q$ and $v(\hat{\beta}_q)$ for $q = 1, ..., 8$, and use precision-weighted prior $N(\sum_q v(\hat{\beta}_q)^{-1}\hat{\beta}_q, (\sum_q v(\hat{\beta}_q)^{-1}/8)^{-1})$.
- Predictors included geography, local Census tract information, commercial estimates of age and income as a target household, and paradata based on interviewer observations and level of effort.

Eliciting Priors using Literature Review

- Found eight studies where response propensity was predicted using one or more of estimators used in the historical data: $\hat{\beta}_{pi}$ for $i = 1, ..., K_p$, where K_p is the number of studies where the pth predictor was used.
- Prior for the p*th* components given by $N(\hat{\beta}_p, v(\hat{\beta}_p))$, where $\hat{\beta}_p = K_p^{-1} \sum_{i=1}^{K_p} \hat{\beta}_{pi}$, $v(\hat{\beta}_p) = (K_p - 1)^{-1} \sum_{i=1}^{K_p} (\hat{\beta}_{pi} - \hat{\beta}_p)^2$.

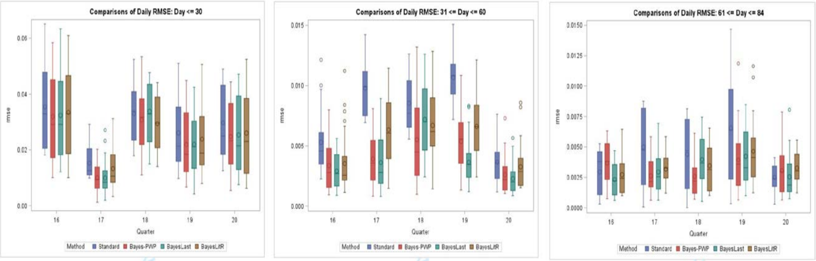
Eliciting Priors using Literature Review

- Found eight studies where response propensity was predicted using one or more of estimators used in the historical data: $\hat{\beta}_{pi}$ for $i = 1, ..., K_p$, where K_p is the number of studies where the pth predictor was used.
- Prior for the pth components given by $N(\hat{\beta}_p, v(\hat{\beta}_p))$, where $\hat{\beta}_p = K_p^{-1} \sum_{i=1}^{K_p} \hat{\beta}_{pi}$, $v(\hat{\beta}_p) = (K_p - 1)^{-1} \sum_{i=1}^{K_p} (\hat{\beta}_{pi} - \hat{\beta}_p)^2$.
    - Note that the prior variance obtained using the literature review assumes independence between each predictor regression estimator, whereas the data-based prior allows for empirically estimated correlations.

Evaluating the Quality of the Bayesian Prediction

1. Compute predicted probability of response $\frac{\exp(\beta_{ps}^T X_{it})}{1+\exp(\beta_{ps}^T X_{it})}$ for each subject, where β_{ps} is the posterior mean of β given by $(v(\hat{\beta})^{-1} + V(\beta_p)^{-1})^{-1}(v(\hat{\beta})^{-1}\hat{\beta}_t + V(\beta_p)^{-1}\beta_p)$ where β_p and $V(\beta_p)$ are the prior means and variance, and t is the day of the interview for that quarter ($t = 7, ..., 84$).
2. Compute difference between this predicted probability and the benchmark predicted probability that uses all the data for the quarter, for each day.

Evaluating the Quality of the Bayesian Prediction



- Major gains for Bayesian approach, especially in the middle period of data collection.
- Effect of prior information diminishes in later period
- Precision weighted prior does best, although last quarter does nearly as well.
- Use of literature-based prior still an improvement over ignoring prior data.

Fully Integrated Response Design Methodology

Fully Integrated Response Design Methodology

- At each phase
  1. Use the results of West et al. 2019 to obtain posterior estimates of response propensity for all unresolved cases.
  2. Use available paradata and sampling frame information to impute missing values for all unresolved cases for a key survey variable.
  3. Use the results of Wagner et al. 2019 to predict future costs of for all unresolved cases.
  4. Repeat 2) and 3) after applying responsive design decision rules to each observation, and estimate the product of predicted cost and the root mean squared error (RMSE) for key survey variable using the results from 2) obtained using the full data as true value. For multiple phases, 2) and 3) should incorporate future decision rules based on predicted outcomes
  5. Repeat 4) for a variety of decisions. Find the decision rule that minimizes the product of cost and RMSE and implement that rule
- Repeat for each phase based on observed outcomes and remaining unresolved cases.

Fully Integrated Response Design Methodology

- Currently implementing in National Survey of College Graduates.
- Data collection schedule:
  - Two web invites (Week 1 and 5)
  - Mailing (Week 8)

Fully Integrated Response Design Methodology

- Currently implementing in National Survey of College Graduates.
- Data collection schedule:
  - Two web invites (Week 1 and 5)
  - Mailing (Week 8)
    - Replace with web invite
  - Telephone NRFU (Week 12)

Fully Integrated Response Design Methodology

- Currently implementing in National Survey of College Graduates.
- Data collection schedule:
  - Two web invites (Week 1 and 5)
  - Mailing (Week 8)
    - Replace with web invite
  - Telephone NRFU (Week 12)
    - No NRFU
  - Second mailing (Week 18)

Fully Integrated Response Design Methodology

- Currently implementing in National Survey of College Graduates.
- Data collection schedule:
  - Two web invites (Week 1 and 5)
  - Mailing (Week 8)
    - Replace with web invite
  - Telephone NRFU (Week 12)
    - No NRFU
  - Second mailing (Week 18)
    - Replace with web invite
  - Third web invite (Week 23)

Fully Integrated Response Design Methodology

- Currently implementing in National Survey of College Graduates.
- Data collection schedule:
  - Two web invites (Week 1 and 5)
  - Mailing (Week 8)
    - Replace with web invite
  - Telephone NRFU (Week 12)
    - No NRFU
  - Second mailing (Week 18)
    - Replace with web invite
  - Third web invite (Week 23)
    - No invite
- Using cutpoints at 5th, 10th,...,50th percentiles of response propensity.

- Alternatively could order by the estimated effect of a given observation on product of cost and variance:

$$\psi_j = \left(\hat{C} - \hat{C}_j\right)\left(\hat{B}_j^2 + \left(\frac{n}{n-1}\right)\hat{V}\right)$$

where $\hat{C} = \sum_{i=1}^{n_0} C_i + \sum_{i=n_0+1}^{n} \hat{C}_i$, $\hat{V} = \frac{1}{n(n-1)}\left[\hat{Y}_2 - \frac{1}{n}\hat{Y}^2\right]$ for $\hat{Y}_2 = \sum_{i=1}^{n_0} y_i^2 + \sum_{i=n_0+1}^{n} \hat{Y}_i^2$, $\hat{Y} = \sum_{i=1}^{n_0} y_i + \sum_{i=n_0+1}^{n} \hat{Y}_i$, and $\hat{B}_j^2 = \left(\frac{1}{n-1}\left(\hat{Y} - \hat{Y}_j\right) - \frac{1}{n}\hat{Y}\right)^2$

- Compute ψ_{s_j} where the elements in s_j consist of the $s-1$ previously selected elements that minimized ψ_{s-1}; find the jth element that minimizes ψ_{s_j} to obtain ψ_s.
- Compute $\delta_s = \psi_0 - \psi_s$ for $s = 1, ..., n - n_0$ and $\psi_0 = \hat{C}\hat{V}$, and choose the value of s that maximizes δ_s
  - Smooth estimator of δ_s to stabilize.
- Might need to limit increase in MSE to avoid getting too large.

Link to Adaptive Clinical Trials

- Often other fields of statistics (e.g. causal inference) "reinvent the wheel" for methods that have already been established in survey statistics.
- Here we might have the opposite situation: adaptive designs in clinical trials have long grappled with issues such as when to stop trials once benefit has been established, or how to maximize benefit in a population with heterogenous responses to treatment (Tourangeau et al. 2017).

Link to Adaptive Clinical Trials

- Often other fields of statistics (e.g. causal inference) "reinvent the wheel" for methods that have already been established in survey statistics.
- Here we might have the opposite situation: adaptive designs in clinical trials have long grappled with issues such as when to stop trials once benefit has been established, or how to maximize benefit in a population with heterogenous responses to treatment (Tourangeau et al. 2017).
    - Different problems but similar solutions – update data collection based on information obtained in the study.

Link to Adaptive Clinical Trials

- Group sequential trials: terminate early if drug shows effect (Jennison and Turnbull 1999).
- Sample size modification: adjust sample size based on early measures of variability in outcome (Cox 1952).
- Outcome adaptive trial: use initial results to “bias” randomization toward successful treatments (Zelen 1969).
- Enrichment designs: inclusion/exclusion criteria designed to concentrate treatment on subpopulation most likely to benefit (Wang et a. 2009).
  - Stopping rules can play havoc with statistical inference (Korn 2001); adaptive trials can mislead (Hey et al. 2015).

Link to Adaptive Clinical Trials

- Group sequential trials: terminate early if drug shows effect (Jennison and Turnbull 1999).
- Sample size modification: adjust sample size based on early measures of variability in outcome (Cox 1952).
- Outcome adaptive trial: use initial results to “bias” randomization toward successful treatments (Zelen 1969).
- Enrichment designs: inclusion/exclusion criteria designed to concentrate treatment on subpopulation most likely to benefit (Wang et a. 2009).
  - Stopping rules can play havoc with statistical inference (Korn 2001); adaptive trials can mislead (Hey et al. 2015).
- What can we do to borrow from this literature while minimizing problems with it?
- “The survey field is missing a link with the adaptive clinical trial field, which has become very sophisticated on the modeling side. If I were 30 years old and smart enough, that would be the link I would try to make.” (Bob Groves, from Habermann, Kennedy, and Lahiri 2017).

Impact on Estimation

- Often the adaptive and especially the responsive design literature ignores the potential impact of design decisions on estimation.
- Not always: Elliott et al. (2000) weights underrepresented later callbacks to provide design-consistent estimation.
- Balancing on response propensity also implicitly at least considers a sort of design-sensitive estimation.
- But other recent examples may, at least implicitly, rely on model assumptions to drop cases based on predicted values.
    - Could drop stochastically and introduce weights for design based estimation or probability of retention in model estimation (Zhang and Little 2009).

Where Do We Go From Here?

"[The Census Bureau wasn't] going to abandon traditional sample surveys, but maybe they could patch through with responsive design. We won't know for ten years whether that is successful." (Bob Groves, from Habermann, Kennedy, and Lahiri 2017).

- We are two years into Bob's cloudy crystal ball.
- A great deal of work is being done in a myriad of direction on response design.
- Major questions going forward:
  - What are we targeting? More focus on optimization.
  - How do changes affect inference? More focus on estimation.
  - What data can we collect and how can we use it efficiently? More focus on paradata and its use in optimization and estimation.

Acknowledgements

- NIH grant 1R01AG058599-01.

References

Bethlehem, J. G. (2002). Weighting adjustments for ignorable nonresponse. In *Survey Nonresponse*, Editors Groves R.M., Dilman D.A., Rltinge J.L., and Little R.J.A., Wiley, NY, p. 275-287.

Calinescu, M., Bhulai, S., and Schouten, B. (2013). Optimal resource allocation in survey designs. *European Journal of Operational Research*, 226:115-121.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4:266-298.

Cox, D. R. (1952). Some recent work on systematic experimental designs. *Journal of the Royal Statistical Society*, B14:211-219.

Elliott, M. R., Little, R. J., and Lewitzky, S. (2000). Subsampling callbacks to improve survey efficiency. *Journal of the American Statistical Association*, 95:730-738.

Groves, R. M., and Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society*, A169:439-457.

Habermann, H., Kennedy, C., and Lahiri, P. (2017). A conversation with Robert Groves.

References

Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41:517-529.

Hey, S. P., and Kimmelman, J. (2015). Are outcome-adaptive allocation trials ethical?. *Clinical Trials*, 12:102-106.

Jennison, C., and Turnbull, B. W. (1999). *Group Sequential Methods with Applications to Clinical Trials.* Chapman and Hall/CRC, Boca Raton,FL.

Korn, E. L., Arbuck, S. G., Pluda, J. M., Simon, R., Kaplan, R. S., and Christian, M. C. (2001). Clinical trial designs for cytostatic agents: are new approaches needed?. *Journal of Clinical Oncology*, 19:265-272.

Lewis, T.H. (2017). Univariate tests for phase capacity: tools for identifying when to modify a survey's data collection protocol. *Journal of Official Statistics*, 33:601-624.

Lewis, T. H. (2019). Multivariate tests for phase capacity. *Survey Research Methods*, 13:153-165.

References

Murphy, J., Biemer, P., and Berry, C. (2018). Transitioning a survey to self-administration using adaptive, responsive, and tailored (ART) design principles and data visualization. *Journal of Official Statistics*, 34:625-648.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., and Lindblad, M. (2010). Reduction of nonresponse bias through case prioritization. *Survey Research Methods*, 4:21-29.

Rao, R. S., Glickman, M. E., and Glynn, R. J. (2008). Stopping rules for surveys with multiple waves of nonrespondent follow-up. *Statistics in Medicine*, 27:2196-2213.

Schouten, B., Cobben, F., and Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, 35:101-113.

Schouten, B., Mushkudiani, N., Shlomo, N., Durrant, G., Lundquist, P., and Wagner, J. (2018). A Bayesian analysis of design parameters in survey data collection. *Journal of Survey Statistics and Methodology*, 6:431-464.

Schouten, B., and Shlomo, N. (2017). Selecting adaptive survey design strata with partial R-indicators. *International Statistical Review*, 85:143-163.

References

Shlomo, N., Skinner, C., and Schouten, B. (2012). Estimation of an indicator of the representativeness of survey response. *Journal of Statistical Planning and Inference*, 142:201-211.

Tourangeau, R., Michael Brick, J., Lohr, S., and Li, J. (2017). Adaptive and responsive survey designs: A review and assessment. *Journal of the Royal Statistical Society*, A180:203-223.

Wagner, J., and Raghunathan, T. E. (2010). A new stopping rule for surveys. *Statistics in Medicine*, 29:1014-1024.

Wagner, J., West, B., Elliott, M.R., and Coffey, S. (2019). Comparing regression modeling and Bayesian additive regression trees to predict costs in a responsive survey design context. Submitted.

Wang, S. J., James Hung, H. M., and O'Neill, R. T. (2009). Adaptive patient enrichment designs in therapeutic trials. *Biometrical Journal*, 51:358-374.

References

West, B., Wagner, J., Coffey, S., and Elliott, M.R. (2019). The elicitation of prior distributions for bayesian responsive survey design: historical data analysis vs. literature review. Submitted.

Zhang, G., and Little, R. (2009). Extensions of the penalized spline of propensity prediction method of imputation. *Biometrics*, 65:911-918.

Zelen, M. (1969). Play the winner rule and the controlled clinical trial. *Journal of the American Statistical Association*, 64:131-146.