What are Optimal Response Rates?

• Optimal unit response rates suggest a maximization problem
  — Optimal response rates have been requested in federal government RFPs
  — Requires some some context

• Two conceptualizations – both fixed cost
  — Maximize overall response rates
  — Rates to minimize nonresponse bias
Maximize Response Rates With Fixed Data Collection Cost

• Goal: maximum overall response rate

• Design features of survey to achieve goal within fixed cost to conduct the survey

• A host of factors are related to both response rates and cost
Key Factors

- Sponsorship
- Content/salience
- Mode
- Material design/length
- Field period
- Incentives
- Number of attempts
- Training of staff
- Allocation of resources
Relative Importance of Factors Varies

• Household surveys
  — Cross-sectional
  — Longitudinal
  — Burden

• Provider Surveys
  — Medical
  — Teachers

• Business Surveys
  — Size of business
Biggest Impact for Cross-sectional Household Surveys per Dollar?

• Token monetary incentive
  — more than token if considerable burden e.g., medical specimen

• Material design investment

• Number of contact/conversion attempts
Maximize Overall Response Rates

• Choice of factors are related as suggested by leverage/saliency theory

• Maximize response rates by
Maximize Overall Response Rates

• Choice of factors are related as suggested by leverage/saliency theory

• Maximize response rates by

• “Cherry picking” – target those with highest propensity to respond
  — Older, female, home owners, not central city, English-speakers, upper-middle income
Maximizing Response Rates?

• As an industry, we have done this for years so that we can tout our survey’s response rates.

• Hope we have changed our ways?
  — We don’t need more “old, female, home owners, not central city, English-speakers, upper-middle income” respondents

• Why? Because of potential nonresponse bias
Response Rates to Minimize Nonresponse Bias

• Keeter et al. (2000), Curtin et al. (2000) and subsequent studies show response rates do not predict nonresponse bias well

• Meta-analysis by Groves (2006) struck a final blow – getting a high response rate does not eliminate nonresponse bias or even minimize it in all cases
Response Rates to Minimize Nonresponse Bias

• Keeter et al. (2000), Curtin et al. (2000) and subsequent studies show response rates do not predict nonresponse bias well

• Meta-analysis by Groves (2006) struck a final blow – getting a high response rate does not eliminate nonresponse bias or even minimize it in all cases

• We still have biased estimates – sometimes large biases
Models of Nonresponse Bias

- Deterministic – bias is a function of difference between respondent and nonrespondent characteristics and nonresponse rate

- Stochastic – bias depends on correlation between characteristic and response propensity
Deterministic View

• Assume the population contains a stratum of respondents and a stratum of nonrespondents, with population means of $\bar{Y}_r$ and $\bar{Y}_m$, respectively. The respondent stratum is $R$ percent of $N$.

• The bias of the unadjusted estimator is

$$bias(\hat{Y}_0) \approx (1 - R)(\bar{Y}_r - \bar{Y}_n)$$
Stochastic View

• Assume every unit in the population has some nonzero probability of responding, $\phi_i$.

• The bias of the unadjusted estimator is

$$bias(\bar{Y}_0) \approx \frac{1}{N\bar{\phi}} \sum^N_{i=1} (Y_i - \bar{Y})(\phi_i - \bar{\phi})$$
## Other Estimates

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Unadjusted estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>$\sum_i (\phi_i - 1) Y_i$</td>
</tr>
<tr>
<td>Domain total</td>
<td>$NP(\phi' - 1)$</td>
</tr>
<tr>
<td>Domain mean</td>
<td>$\frac{\sum_i (\phi'_i - 1) Y_i}{NP(\phi' - 1)} - \bar{Y}_A$</td>
</tr>
<tr>
<td>Difference of domain totals</td>
<td>$-N{(P_A - P_B) - (P_A \phi'_A - P_B \phi'_B)}$</td>
</tr>
<tr>
<td>Ratio</td>
<td>$\frac{\sum \phi_i Y_i}{\sum \phi_i W_i} - \frac{Y}{W}$</td>
</tr>
</tbody>
</table>
Other Estimates

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Poststratified estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>$$\sum_h \bar{\phi}<em>h^{-1} \sum_i Y</em>{hi} (\phi_{hi} - \bar{\phi}_h)$$</td>
</tr>
<tr>
<td>Domain total</td>
<td>$$\sum_h N_h P_h \left( \bar{\phi}_h^{-1} \phi_h' - 1 \right)$$</td>
</tr>
<tr>
<td>Domain mean</td>
<td>$$\frac{\sum_h \sum_i \bar{\phi}<em>h^{-1} \phi</em>{hi} Y_{hi}}{\sum_h N_h P_h \left( \bar{\phi}_h^{-1} \phi_h' - 1 \right)} - \bar{Y}_A$$</td>
</tr>
<tr>
<td>Difference of domain totals</td>
<td>$$- \sum_h N_h \left{ \left( P_{A_h} - P_{B_h} \right) - \bar{\phi}<em>h^{-1} \left( P</em>{A_h} \phi_A' - P_{B_h} \phi_B' \right) \right}$$</td>
</tr>
<tr>
<td>Ratio</td>
<td>$$\frac{\sum_h \bar{\phi}<em>h^{-1} \sum_i \phi</em>{hi} Y_{hi}}{\sum_h \bar{\phi}<em>h^{-1} \sum_i \phi</em>{hi} W_{hi}} - \frac{Y}{W}$$</td>
</tr>
</tbody>
</table>
Message

• Nonresponse bias is not simple

• It is a function of
  — Response propensities of units in the domain being estimated
  — The type of estimate (total, mean, ratio, odds ratio)
  — The specific estimator and auxiliary data
Implications

• How do we ‘optimize’ for nonresponse bias when most surveys are multi-purpose and produce many types of estimates?

• How does ‘response’ theory relate to findings of Keeter et al and others that survey estimates are relatively robust in terms of bias to response rates?
Understanding Reasons for Bias

• Not every statistic is biased because there must be a ‘mechanism’ to cause nonresponse bias

• The mechanism may be a function of
  — what the survey does (e.g., sponsorship, content, mode, questionnaire, allocation of resources)
  — characteristics of respondents (e.g., altruistic, connected)
Estimating a Proportion ($P$)

- Suppose response propensities for units with the characteristic $= \phi_1$, while for units without the characteristic $= \phi_2$. Let $\lambda = \phi_2 \phi_1^{-1}$.

- The bias of the unadjusted estimate is

$$bias(p_{HT}^*) = P(1 - P)(\phi_1 - \phi_2)\{P \phi_1 + (1 - P) \phi_2\}^{-1}$$

$$= P(1 - P)(1 - \lambda)\{P + (1 - P) \lambda\}^{-1}$$
Volunteer Example

• Estimate the percent of adults who volunteer. Response rate of volunteers is 45% and non-volunteers is 30%

\[ \lambda = \frac{0.30}{0.45} = 0.67 \quad (1/\lambda = 1.5) \]

• If P= 25% (volunteers), then unadjusted estimate is about 32% (rel-bias 27%)
Rel-bias, bias($p)/p$, for various $\lambda = \phi_2 \phi_1^{-1}$

<table>
<thead>
<tr>
<th>$P$</th>
<th>1.1</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>-9%</td>
<td>-33%</td>
<td>-50%</td>
<td>-60%</td>
</tr>
<tr>
<td>10%</td>
<td>-8%</td>
<td>-31%</td>
<td>-47%</td>
<td>-57%</td>
</tr>
<tr>
<td>25%</td>
<td>-7%</td>
<td>-27%</td>
<td>-43%</td>
<td>-53%</td>
</tr>
<tr>
<td>50%</td>
<td>-5%</td>
<td>-20%</td>
<td>-33%</td>
<td>-43%</td>
</tr>
<tr>
<td>75%</td>
<td>-2%</td>
<td>-11%</td>
<td>-20%</td>
<td>-27%</td>
</tr>
<tr>
<td>90%</td>
<td>-1%</td>
<td>-5%</td>
<td>-9%</td>
<td>-13%</td>
</tr>
<tr>
<td>99%</td>
<td>0%</td>
<td>0%</td>
<td>-1%</td>
<td>-1%</td>
</tr>
</tbody>
</table>
Message

• Relative ratios of response rates for domains must be large to incur a large bias

• Relative effects are different depending on the size of the estimate.
Message

• Relative ratios of response rates for domains must be large to incur a large bias

• Relative effects are different depending on the size of the estimate.

• This tells us the implications of different rates, but does not tell us when to expect different rates.
Bounding Nonresponse Bias

Use of this theory requires understanding the causes of nonresponse (accessibility and amenability) that induce response propensities to differ by domain.

• Direct causes or variables highly correlated with direct causes
Theories of Response

Key aspects have been addressed but it is not comprehensive

• Sociological (e.g., Goyder 1987)

• Psychological (e.g., Tourangeau et al. 2000)

• Operational (e.g., Dillman 1978; Groves and Couper 1998; Stoop et al. 2010)
Predicting Nonresponse Bias?

• Some theory is mainly descriptive.

• Some is prescriptive, but not quantitative.

• Groves et al. (2004) manipulated design factors to try to produce nonresponse bias.
  — Large nonresponse biases not produced in many cases when expected
  — If we can’t produce it, we do not understand it.
Needed Research - 1

Comparative analysis of respondents and nonrespondents.

• Focus on features/interactions that can be manipulated

• Basic research need (perhaps in-depth studies like ethnographic studies in spirit)?
Needed Research - 2

Procedures to test theories in practice

• Programs that aim to produce differential response rates for domains as predicted

• Programs that aim to change domain response rates based on sequences of activities

• Focus on low cost factors such as material design
Needed Research - 3

Evaluations linking statistical adjustments and data collection procedures

• If we can adjust for it afterwards, does it help to deal with it in data collection?

• Identifying methods that improve the quality of different types of statistics thus being of general purpose.
Optimization – Final Thoughts

• We need to be clear about what we are optimizing

• We need a program, not a study, to move in the directions needed

• Progress is likely to be limited without more comprehensive understanding of response