Nonresponse Bias: Three Paradoxes

Presented at the GESIS-Leibniz Institute for Social Sciences
Mannheim, Germany
Roger Tourangeau, Westat

Tuesday, November 17, 2015
Three Paradoxes

• Why is the nonresponse rate so poorly related to nonresponse bias?
• Are alternative indicators any better?
• Can we reduce bias or accomplish other objectives through responsive design?
The Relationship between NR Rate and NR Bias

• The theory

\[
NR\ Bias(\bar{y}_r) = P_0 (\mu_1 - \mu_0) + \frac{Cov(p_{1i}, \mu_i)}{P_1}
\]

Portion due to zero propensity cases  
Portion due to nonzero propensity cases

• How can there not be a relationship?
The Findings

• Six widely cited studies suggest little or no relationship
  — Groves (2006)
  — Groves and Peytcheva (2008)
  — Keeter, Miller, Kohut, Groves, and Presser (2000)
  — Keeter, Kennedy, Dimock, Best, and Craighill (2006)
Merkle and Edelman

• Looked at relationship between response rates and errors in exit polls

Table 1. Correlation between Precinct Response Rate and Signed and Unsigned Errors in the Estimated Vote Share, by Election

<table>
<thead>
<tr>
<th>Election</th>
<th>Signed errors</th>
<th>Unsigned errors</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>.10</td>
<td>−.13</td>
<td>1,005</td>
</tr>
<tr>
<td>1994</td>
<td>.00</td>
<td>−.07</td>
<td>885</td>
</tr>
<tr>
<td>1996</td>
<td>−.01</td>
<td>−.04</td>
<td>1,205</td>
</tr>
<tr>
<td>1998</td>
<td>.01</td>
<td>−.07</td>
<td>894</td>
</tr>
</tbody>
</table>
Groves and Peytcheva (2008)

• Looked at 59 studies with bias estimates (959 estimates)
Groves and Petcheva (cont’d)

• Most comprehensive look

• Two conclusions
  — Little or no relation between n-r bias and n-r rate
  — Tremendous with study variability

• Second conclusion is very important because it means that no study-level indicator can tell us much about the n-r bias
Reanalysis

- Absolute relbias can be a misleading statistic: \((2\% - 1\%)/1\% = \text{absolute relbias of 100\%}\)

- I looked at absolute error in 422 proportions (out of 556 estimates Peytcheva provided)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Absolute Diff</th>
<th>Absolute Relbias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportions only (individual stats)</td>
<td>-.339 (422)</td>
<td>-.286 (422)</td>
</tr>
<tr>
<td>Proportions only (study-level means)</td>
<td>-.025 (41)</td>
<td>-.333 (41)</td>
</tr>
<tr>
<td>ANOVA (R squared)</td>
<td>.576</td>
<td>.343</td>
</tr>
</tbody>
</table>
y = -0.1518x + 20.655
\( R^2 = 0.1108 \)

y = -0.1124x + 16.378
\( R^2 = 0.0512 \)
Reanalysis Conclusions

• There is a relation between n-r rate and n-r bias
• Some study level characteristics beside n-r rate are important (e.g., method of estimating bias)
• Big differences in bias by study; study accounts for most of the variance
Some Theories accounting for Low Relationship

• No bias in general
  — Propensities determined by study-level features (e.g., distance of interviewer from exit from precinct)
  — Propensities essentially random, product of many highly variable characteristics of R

• Low bias
  — Propensities stable across many design features; design features determine overall RR
  — Propensities stable but determined by R characteristics unrelated to survey variables

• Hidden bias
  — Observed propensities truncated; bias produced by zero propensity cases
No Bias

1) Response propensities largely a function of design characteristics (e.g., advance letter, incentives, mode of data collection, interviewer skill, length of field period, number and timing of contact attempts; survey sponsor, whether survey is mandatory); these are completely unrelated to respondent characteristics, including survey variables

2) Response propensities essentially random: a product of a large number of R characteristics, many of them transient (temporary illness, how busy at work, call screening policies, travel schedule, mood, whether R reads advance letter, etc.)
Low Bias

3) Contrary to leverage-salience theory, same people are more likely to respond regardless of topic and sponsor; more formally, base response propensity plus main effects for various design characteristics, such as mode and incentives.

4) For example, regardless of mode, surveys overrepresent older, more educated, more prosperous people, people living outside central cities, those who own their homes, have landlines, etc.

• Bias similar whatever overall RR
Hidden Bias

5) The problem is not the people with some propensity of responding

• Instead, there is a large bias component that reflects growing segment of the population with essentially no chance of taking part (see Slide 3)

• For example, with a telephone survey, 50 percent are in zero propensity stratum; rest (15-20 percent of the other stratum who do respond) essentially a product of chance factors
Relation Nonlinear

6) When propensities very low (telephone survey) or very high (mandatory FTF survey), little variation in response propensities and therefore little bias

- Overall, the RR-bias relationship is non-monotonic
Paradox 2: Alternatives to the Response Rate

• Schouten and his colleague have proposed the R-indicator and Särndal has proposed the balance indicator as alternatives to the response rate of the threat of bias

• Both have a lot of appeal
  — R Indicator
    ✦ If propensities are constant (and non-zero) then threat of bias is eliminated (for all variables)
    ✦ Thus, variance (or SD) of propensities should be a good indicator of threat
  — Balance indicator is also good proxy for bias
    ✦ Distance of Rs from a vector of population targets $Balance = D'\Sigma^{-1}D$
    ✦ $D$ is a vector of differences between the means for the respondents and for the full sample or population on the auxiliary variables; $\Sigma$ is a cross-products matrix for the auxiliaries
Problems with the Alternatives

• Both share an issue: If within-study variation in bias swamps between-study variance, no one number will ever be an adequate indicator; this is a key problem

• R-indicator has some additional problems
  — Only as good as the propensity model (bad model suggests propensities don’t vary much)
  — Doesn’t seem to vary much
  — Simulations suggest it is not a good indicator of bias (Beaumont, Bocci, and Haziza, 2014)

• Balance indicator—overall distance from frame variables or other auxiliaries may not be good proxy for distance between Rs and non-respondents on survey variables
Paradox 3: Responsive/Adaptive Design

• Responsive designs: Designs with multiple phases, with aim of achieving less biased, more representative sample (Groves and Heeringa, 2006); American Community Survey has done this all along (mail, telephone, FTF for a subsample)

• Adaptive designs: Designs tailored from the outset (Luiten and Schouten, 2013) or adapted continuously throughout the field period Peytchev, Riley, Rosen, Murphy, and Lindblad, 2010)
Do These Designs Work? Luiten and Schouten (2013)

• The control group for the experiment was the regular SCS (CATI-only)

• Experimental group: two phases of data collection.

• Phase 1
  — Cases with low cooperation propensities were sent a mail questionnaire;
  — Those with the highest propensities invited to complete a web survey;
  — Those with intermediate propensities were given a choice
Luiten and Schouten (2013)—(Cont’d)

• Phase 2: NR follow-up by telephone
  — Those in the highest contact propensity quartile were fielded later and were called during the day;
  — Those in the second highest contact propensity quartile were called twice at night and then switched to a schedule that alternated daytime and nighttime calls;
  — Those in the lowest two contact propensity quartiles were called on every shift of every day.
  — The best telephone interviewers were assigned to cases with the lowest cooperation propensities and the worst telephone interviewers were assigned to those with the highest cooperation propensities.
## Luiten and Schouten (2013)—Results

<table>
<thead>
<tr>
<th>Contact Propensity Quartile</th>
<th>Contact Rates</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Lowest Contact Propensity</td>
<td>87.1</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td>Second Lowest Contact Propensity</td>
<td>96.6</td>
<td>94.5</td>
<td></td>
</tr>
<tr>
<td>Second Highest Contact Propensity</td>
<td>93.7</td>
<td>95.7</td>
<td></td>
</tr>
<tr>
<td>Highest Contact Propensity</td>
<td>95.3</td>
<td>96.9</td>
<td></td>
</tr>
</tbody>
</table>

| Cooperation Propensity Quartile                  | Cooperation Rates |          |          |
|                                                  | Experimental      | Control  |
| Lowest Cooperation Propensity                     | 65.1             | 62.7     |
| Second Lowest Cooperation Propensity              | 71.4             | 68.4     |
| Second Highest Cooperation Propensity             | 72.8             | 75.3     |
| Highest Cooperation Propensity                    | 74.7             | 79.2     |

References: Luiten and Schouten (2013)
Luiten and Schouten (2013)—Discussion

• Overall, the experimental, adaptive field work group had a slightly higher response rate than the regular SCS cases (63.8 percent versus 62.8 percent, a non-significant difference).

• The representativeness of the experimental sample was significantly higher than that of the control sample (R-indicators of .85 and .77, respectively)

• Less variation in rates across quartiles
Ideal Intervention

- Two phase design: At end of Phase 2, want all propensities to be as close to equal as possible.

- If $p_{1i} > k$ then stop; overall propensity is $p_{1i}$; otherwise, move on to second phase.

- Phase 2 propensity is, in the ideal, $k - p_{1i}$: In Phase 2 work low propensity cases harder than high propensity cases or change protocol somehow.

- If that can be achieved (it cannot), then bias is reduced:

$$NR\ Bias(\bar{y}_r) = \frac{\bar{p}_1 \cdot Cov(p_{1i}, \mu_i)}{\bar{p}_1 + \bar{p}_2 \cdot \bar{p}_1}.$$
Some Practical Limits to Responsive/Adaptive Design

• Often, don’t know how to raise propensities; three moves common in U.S. surveys
  — Change mode (e.g., move to face-to-face)
  — Increase incentive
  — Shorter questionnaire

• Obstacles to all three

• If we knew what to do, would have done it from the beginning!

• Often easier to lower propensities of high propensity cases by limiting effort on them (Lundquist and Särndal, 2013; Särndal and Lundquist, 2014)
Conclusions

• NR rate *not* a worthless indicator; R indicator and balance indicator may be better, but there is a significant relation between NR rate and NR bias

• Not clear that NR bias isn’t strongly related to study characteristics; if it isn’t, any study-level indicator will not be much help

• Not clear why relation between RR and NR bias not stronger or more linear—I outlined six possibilities

• We don’t know how to raise response propensities in current environment; optimal strategy is to adjust field work to reduce variation in overall propensities