Can Nonresponse Followup Increase Total Bias? Some Insights from Total Error Modeling

Paul Biemer
RTI International
University of North Carolina

April 17, 2009

Motivation

- By some theories, nonresponse error is related to measurement error
  - Example: Units with low response propensities (e.g., reluctant or elusive persons) may be more prone to misclassification
- If so, pursuing nonrespondents may increase measurement bias.
- Total bias (i.e., nonresponse + measurement bias) may increase as nonresponse bias is decreased.

Our Approach

- Postulate models for combining nonresponse and misclassification error as a function of followup level of effort (LOE)
- Simulate various plausible scenarios for the relationships among these parameters
- Draw insights into the nature of the total bias as a function of the LOE

An Expression for Total Bias (B_T)

$$B_T = \bar{y}_r - \pi =$$

$$= (\bar{y}_r - P) + (P - \pi)$$

$$= E\left(\frac{\sum_{i=1}^{n} R_i y_i}{\sum_{i=1}^{n} R_i} - P \right) + (P - \pi)$$

$$= \frac{Cov(y_i, \rho_i)}{\rho} + (P - \pi)$$

Covariance Term

NR bias can be expressed as

$$B_{NR} = \frac{Cov(\mu_i, \rho_i)}{\rho}$$

If instead of observing the true values, μ_i , we observe y_i subject to measurement errors, then

$$B_{NR} = \frac{Cov(y_i, \rho_i)}{\rho}$$

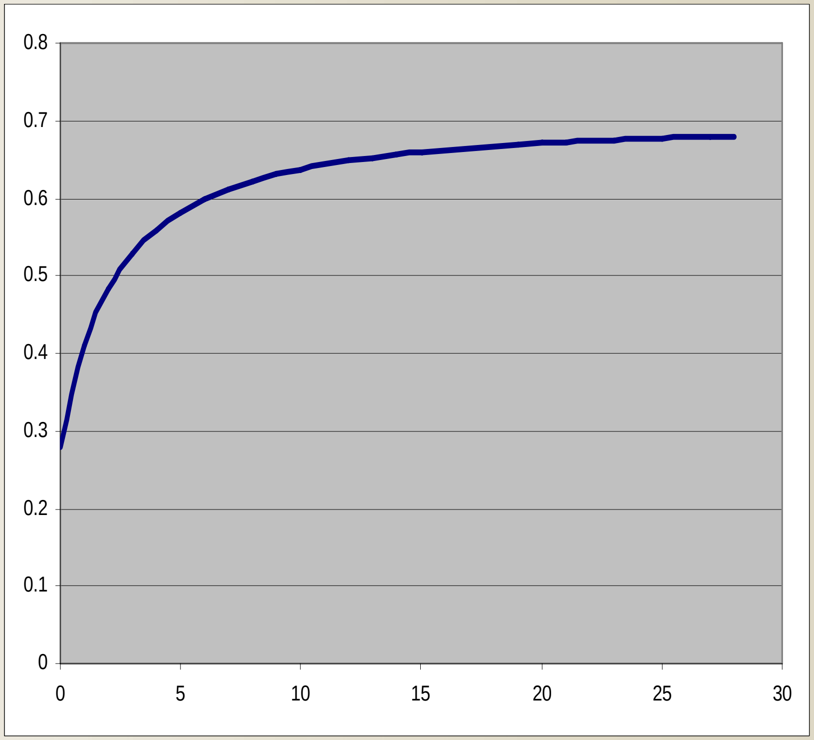
is a combination of NR bias and ME bias components

Essential Ideas (cont'd)

ρ_i is a function of the LOE

As LOE ↑ so does response propensity

It follows that ↑ LOE will tend to attenuate $\frac{Cov(y_i, \rho_i)}{\rho}$ regardless of the level of measurement error



Conjecture: Under this model, it is unlikely that ↑ LOE will appreciably increase ME bias

The remaining slides essentially demonstrate this mathematically for various “worse case” scenarios for the correlation between response propensity and misclassification error parameters.

Model for Dichotomous Variables

Notation

$$\mu_i = \begin{cases} 1 & \text{if unit } i \text{ truly possesses the characteristic} \\ 0 & \text{if unit does not} \end{cases}$$

$y_i =$ Observed value of μ_i

$\pi = E(\mu_i)$ True population prevalence

Classification Error Parameters

$$\phi_i = \Pr(y_i = 1 \mid \mu_i = 0) \qquad \phi = E(\phi_i \mid \mu_i = 0)$$

$$\theta_i = \Pr(y_i = 0 \mid \mu_i = 1) \qquad \theta = E(\theta_i \mid \mu_i = 1)$$

$$y_i = \mu_i(1-\theta_i) + (1-\mu_i)\phi_i$$

$$E(y_i) = P = \pi(1-\theta) + (1-\pi)\phi$$

Nonresponse Error Parameters

$R_i =$ Response/nonresponse indicator

$$\rho_i = \Pr(R_i = 1 \mid i)$$

$$\rho_1 = E(\rho_i \mid \mu_i = 1) \qquad \rho_0 = E(\rho_i \mid \mu_i = 0)$$

$$\rho = E(\rho_i) = \pi\rho_1 + (1-\pi)\rho_0$$

Partitioning the Total Bias

$$B_T = \frac{Cov(y_i, R_i)}{\rho} + (P - \pi)$$

$$= \frac{Cov[\mu_i(1-\theta_i) + (1-\mu_i)\phi_i, \rho_i]}{\rho} + (P - \pi)$$

$$= \underbrace{\left\{ \frac{Cov[-\mu_i\theta_i + (1-\mu_i)\phi_i, \rho_i]}{\rho} + P - \pi \right\}}_{\text{Measurement bias, } B_{ME}} + \underbrace{\frac{Cov(\mu_i, \rho_i)}{\rho}}_{\text{Nonresponse bias, } B_{NR}}$$

Simple Expression for the Total Bias

$$\begin{aligned} B_T &= B_{ME} + B_{NR} \\ &= B_{NR}(1-\theta-\phi) \\ &+\frac{1}{\rho}\left[(1-\pi)Cov_0(\rho_i,\phi_i)-\pi Cov_1(\rho_i,\theta_i)\right] \\ &+(P-\pi) \end{aligned}$$

where $B_{NR} = \pi(1-\pi)\dfrac{\rho_1-\rho_2}{\rho}$

Modeling Total Bias as a Function of LOE

$$K = \text{ Number of attempts to contact}$$

$$\alpha_i = \Pr(\text{unit } i \text{ is contacted at a call attempt})$$

$$\beta_i = \Pr(\text{unit } i \text{ is interviewed}|\text{contacted})$$

$$\rho_{iK|\mu} = \Pr(\text{respondent } i \text{ responds after } K \text{ call-backs}|\mu_i)$$

$$= \sum_{k=0}^{K} (1-\alpha_{i|\mu})^k \alpha_{i|\mu} \beta_{i|\mu}$$

Application

Using this model we can explore the magnitude of B_T for:

- Alternate assumptions for the distribution of $\rho_i, \alpha_i,$ and β_i (employ beta-distributions)
- Number of callbacks, K
- Varying assumptions regarding the magnitudes of the covariance terms

$$Cov_0(\rho_{iK|0}, \phi_i) \text{ and } Cov_1(\rho_{iK|1}, \theta_i)$$

Illustration

Consider

$$Cov_1(\rho_{iK|1}, \theta_i) = \sum_{k=0}^{K} Cov_1[(1-\alpha_{i|1})^k \alpha_{i|1}\beta_{i|1}, \theta_i]$$

Do the error probabilities depend upon

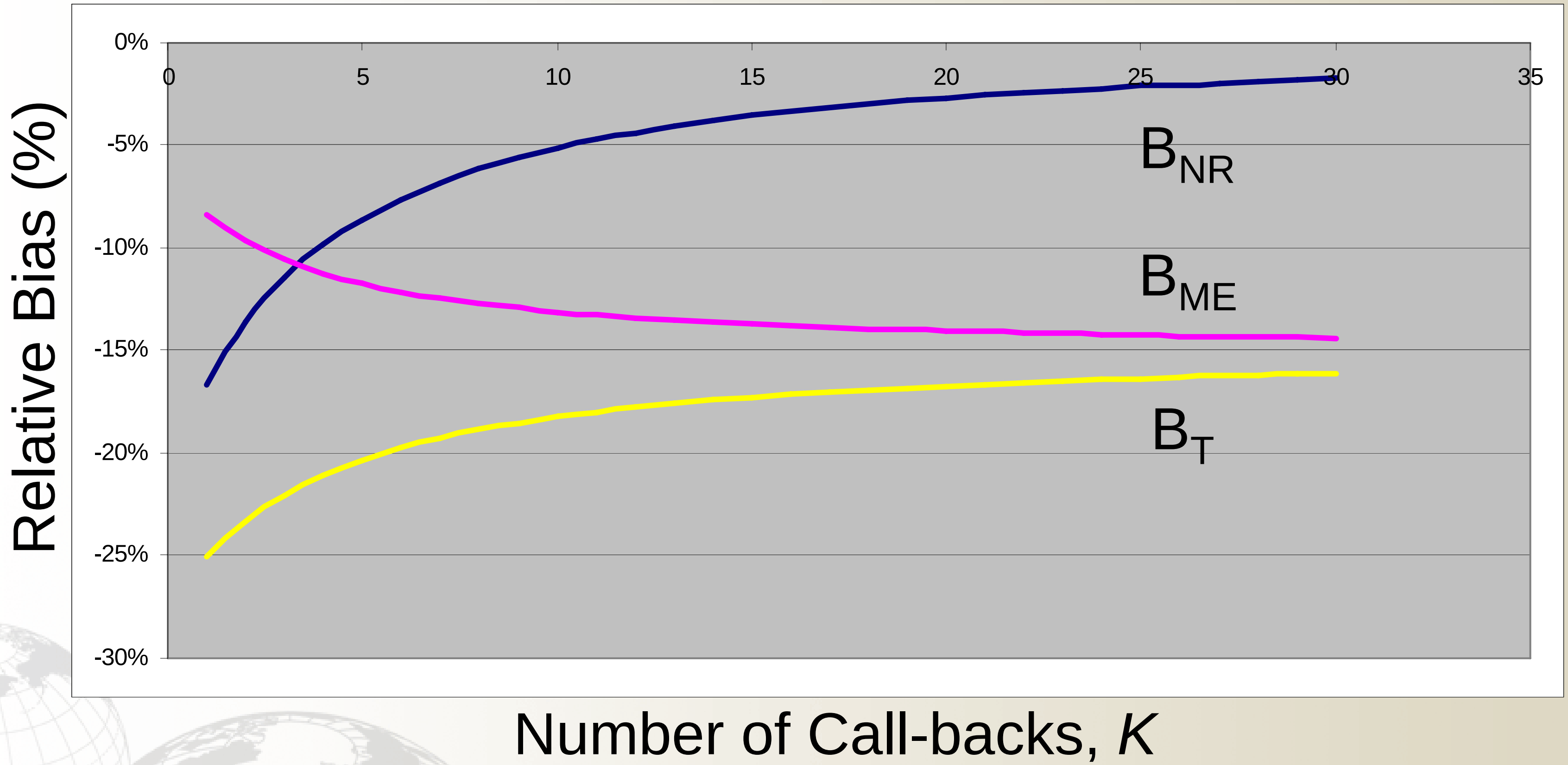
- Contact probabilities, α
- Interview probabilities, β
- Contact attempts, K
- Or, some combination

Suppose Error Probabilities are Related to Contact or Interview Probabilities

- Assume that contact probs (α's) or interview probs(β's) are strongly correlated with the misclassification probs (θ's and ϕ's)
- Worse case scenario:
  - correlation is 1 (e.g., $\theta_i = constant \times \alpha_i$)
  - errors are one-sided (i.e., either θ=1 or ϕ=1)
  - Small nonresponse bias (i.e., $\rho_1 \approx \rho_2$)

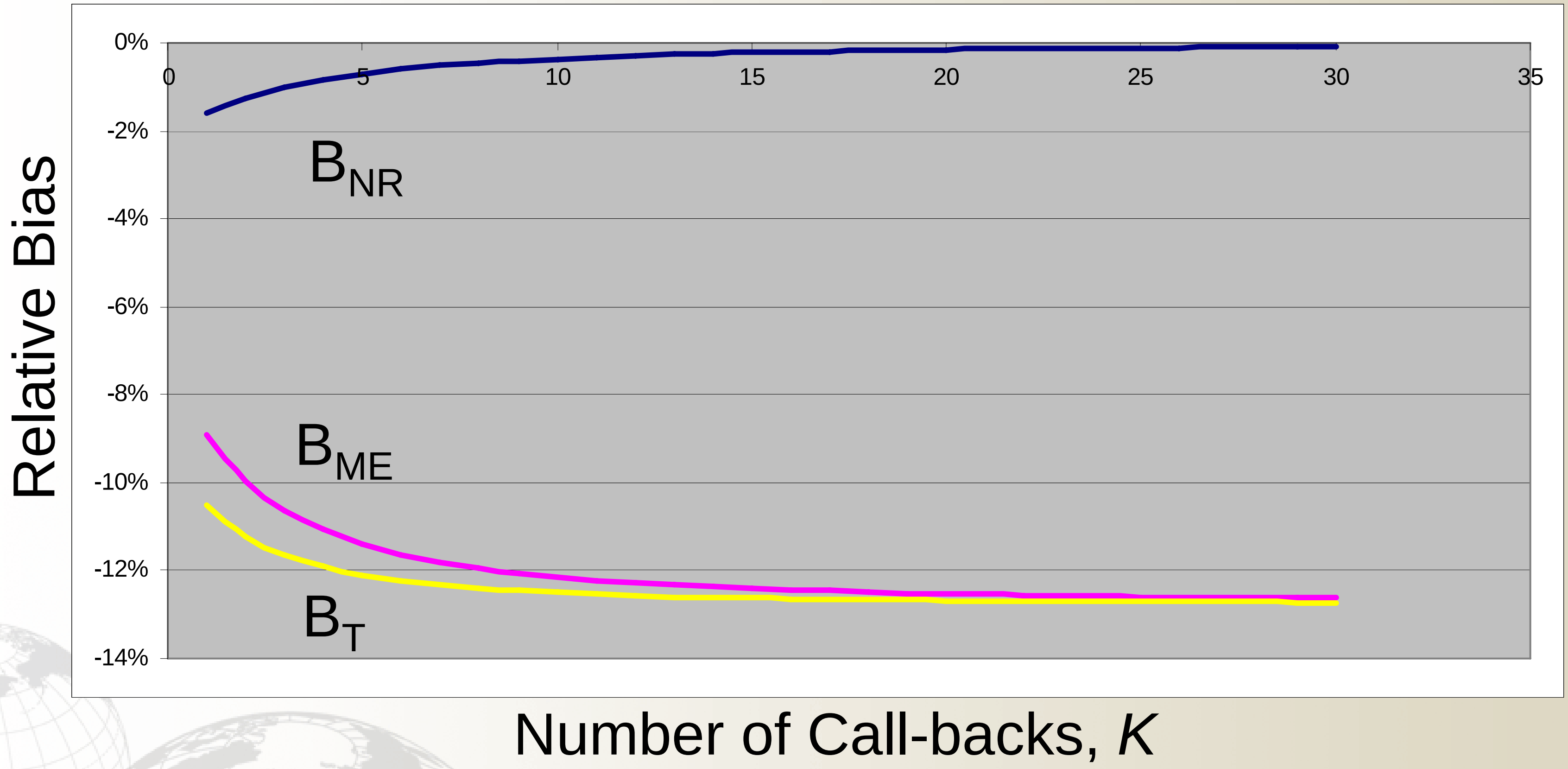
Classification Probability is Perfect Correlated with Response Propensity

B_{NR} is Small



Worst Case Scenario ($|B_{NR}| \approx 0$)

B_{NR} is Negligible



Summary of Findings

- Many scenarios were considered under various model assumptions
- When B_{NR} is very small, total bias can increase as as LOE increases due to increasing measurement errors.
- However, the increase was small, even under the most extreme conditions
- This work questions concerns that followup LOE can increase total bias through greater measurement bias.

Conclusions

- The proposed model combines the effects of measurement error and nonresponse bias as a function of LOE
- It is useful for gaining insights into the relationships among these parameters
- Model validation work is underway

For More Information

Paul Biemer

ppb@rti.org

919-541-6056