



MICHIGAN



JPSM

The Joint Program  
in Survey Methodology

# Response Rates and Nonresponse Bias - What We Know and What We Don't Know

**Robert M. Groves**

University of Michigan and the Joint Program in Survey Methodology

**September, 2006**

Funding from the Methodology, Measurement, and Statistics Program  
of the US National Science Foundation Grant 0297435.

# Four Mutually-Problematic Observations

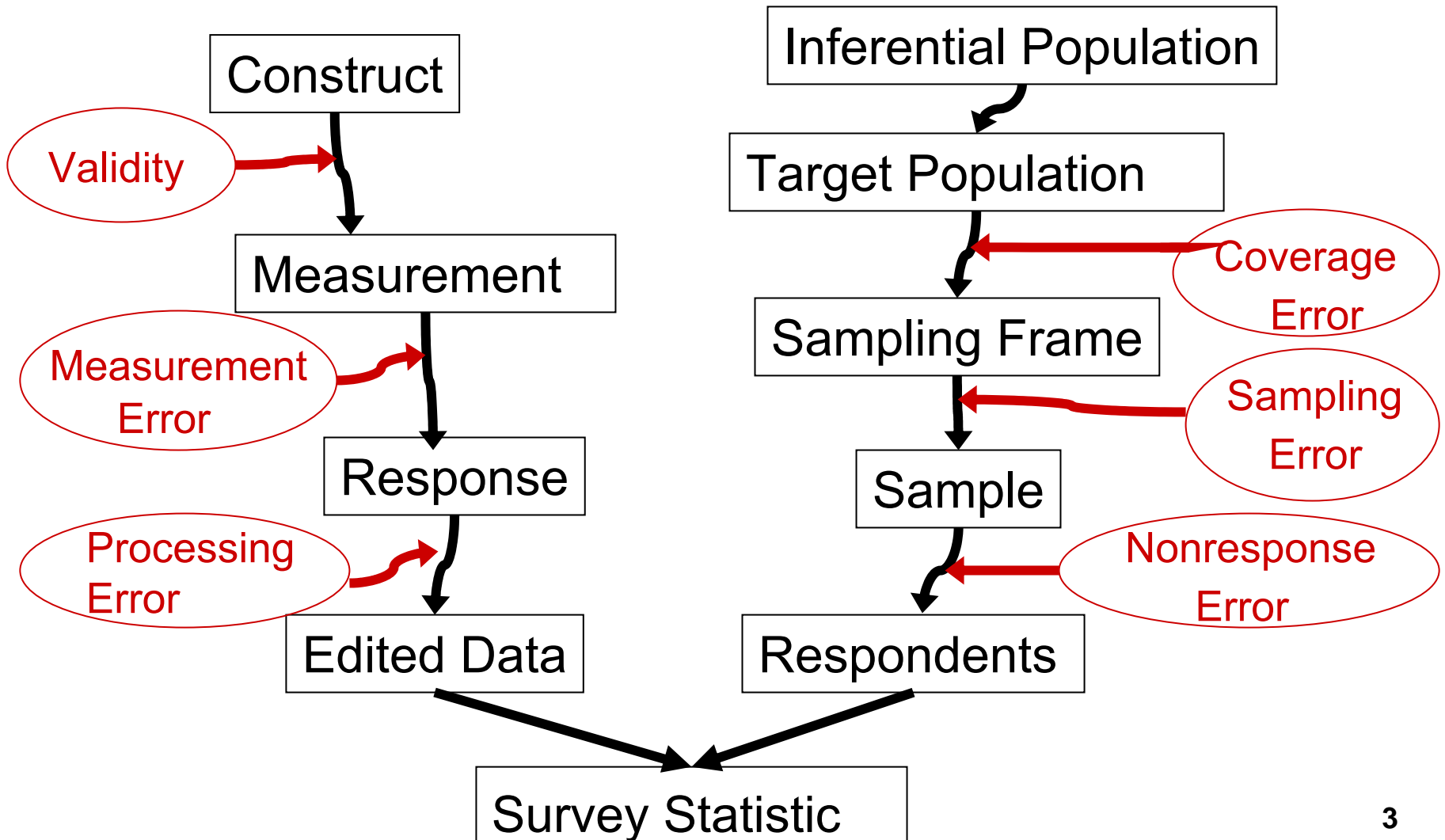
1. **With 100% response rates probability samples provide measurable uncertainties for unbiased estimates**
2. **Response rates are declining**
3. **Keeter *et al.* (2000), Curtin *et al.* (2000), Merkle and Edelman (2002) show no nonresponse bias associated with varying nonresponse rates**
4. **We continue to urge practitioners to achieve high response rates**

**Result: Confusion among practitioners**

# Nonresponse Viewed in the Total Survey Error Perspective

Measurement

Representation



# Assembly of Prior Studies of Nonresponse Bias

- **Search of peer-reviewed and other publications**
- **47 articles reporting 59 studies**
- **About 959 separate estimates (566 percentages)**
  - mean nonresponse rate is 36%
  - mean bias is 8% of the full sample estimate
- **We treat this as 959 observations, weighted by sample sizes, multiply-imputed for item missing data, standard errors reflecting clustering into 59 studies and imputation variance**

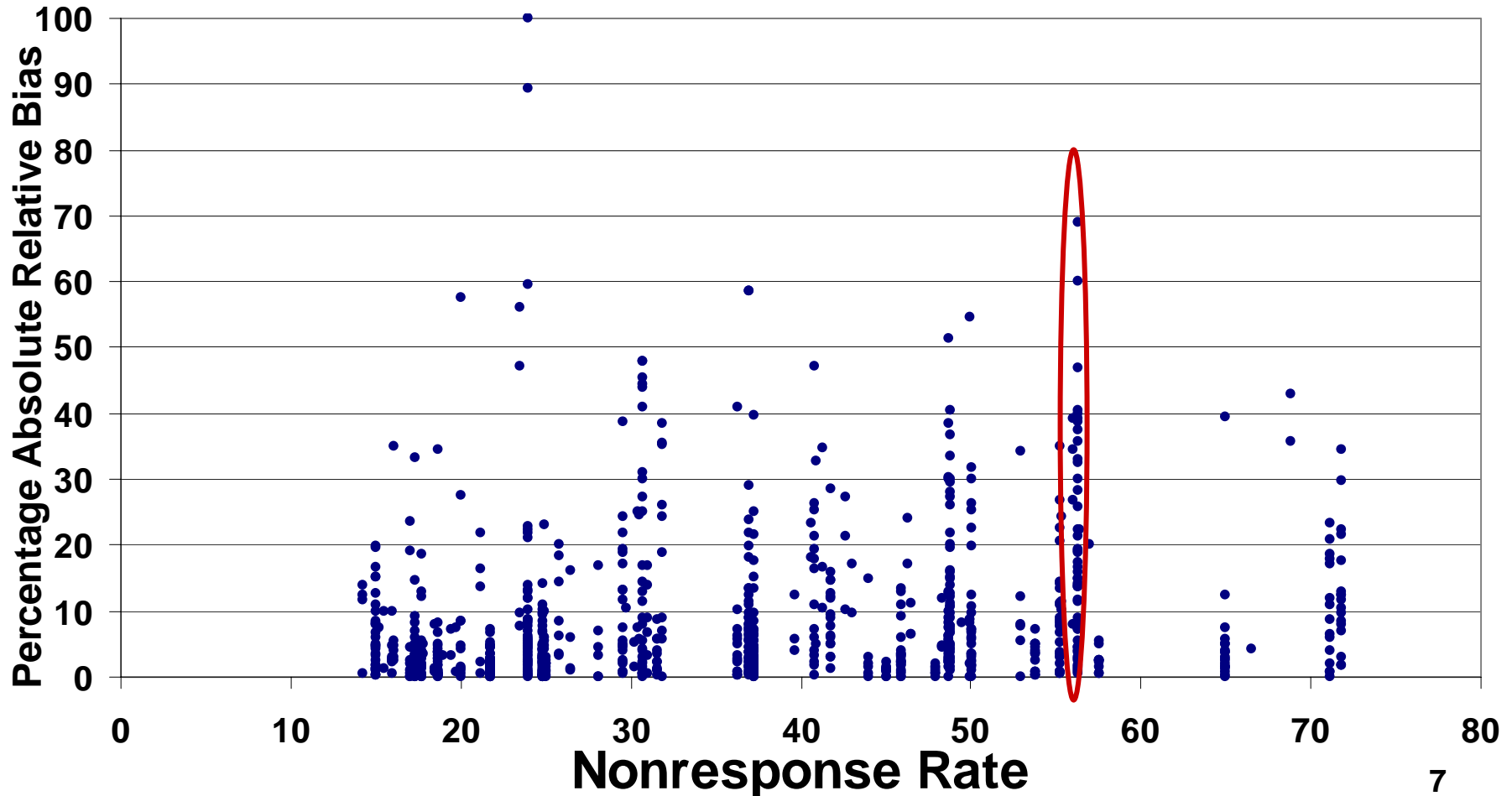
# Meta-Analysis of Nonresponse Error Studies

- Each has data available to compute a relative bias due to nonresponse. The absolute value of the relative bias is

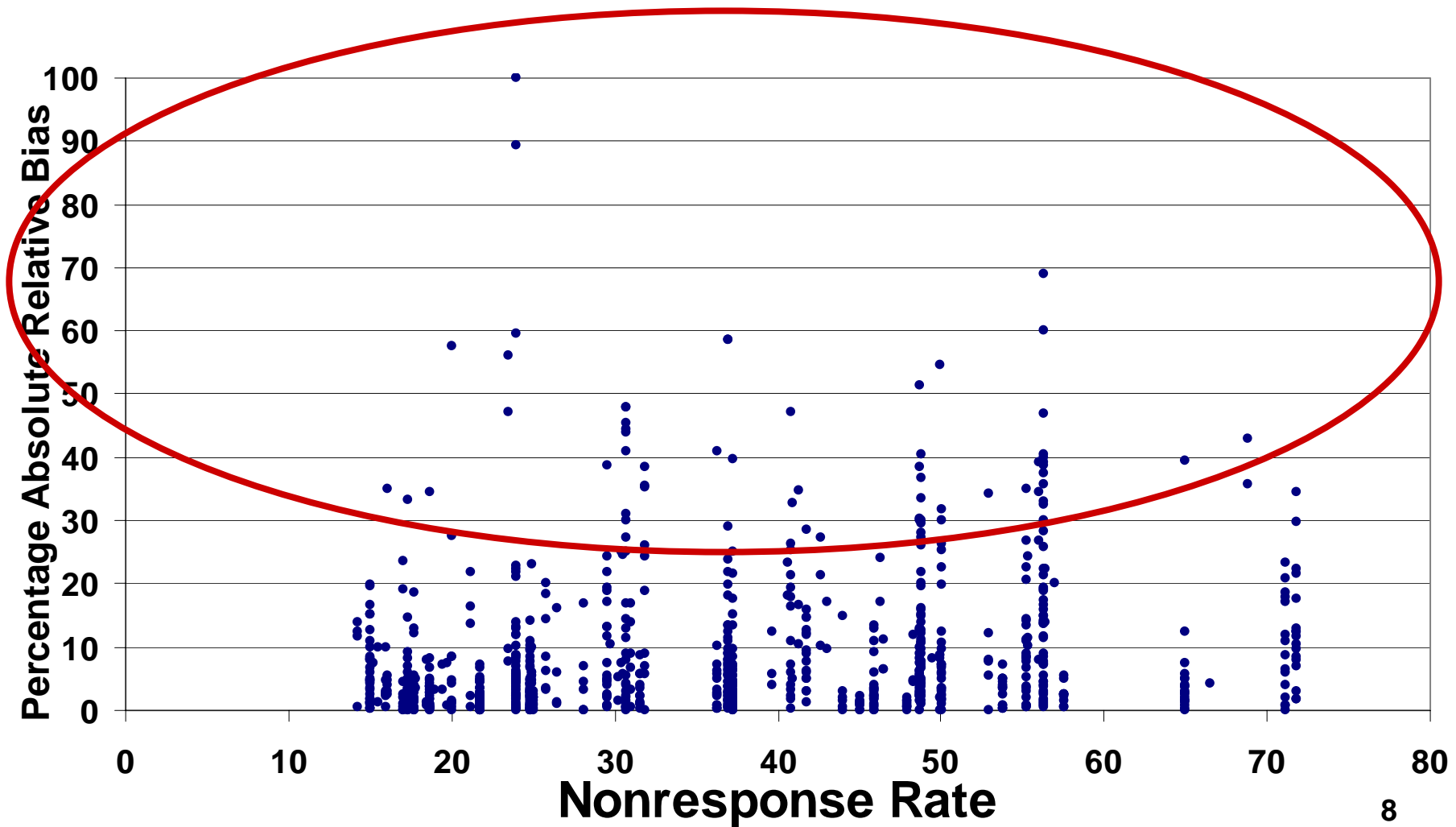
$$\left| \frac{(\bar{y}_r - \bar{y}_n)}{\bar{y}_n} \right|$$



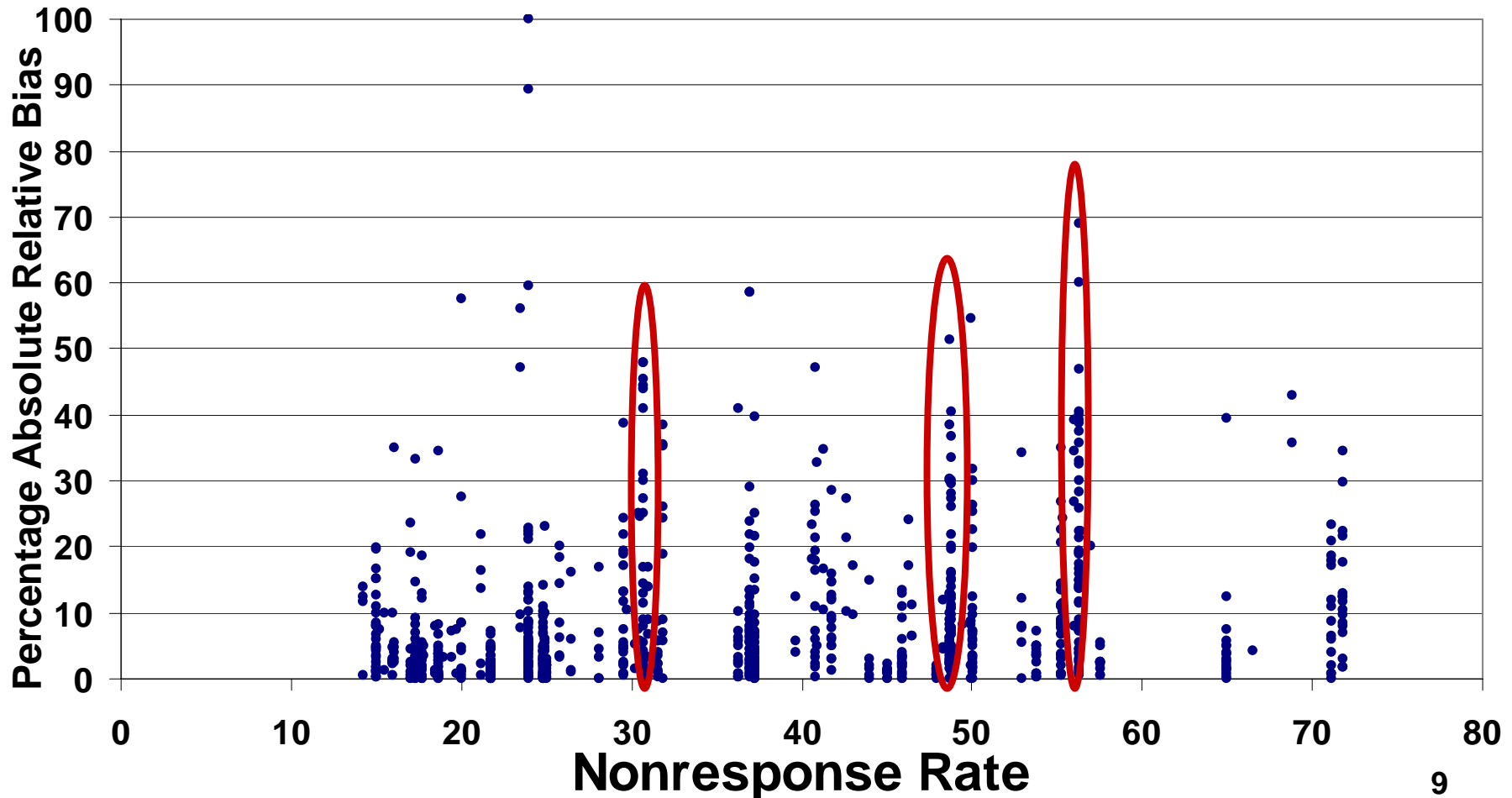
# Percentage Absolute Relative Nonresponse Bias by Nonresponse Rate for 959 Estimates from 59 Studies



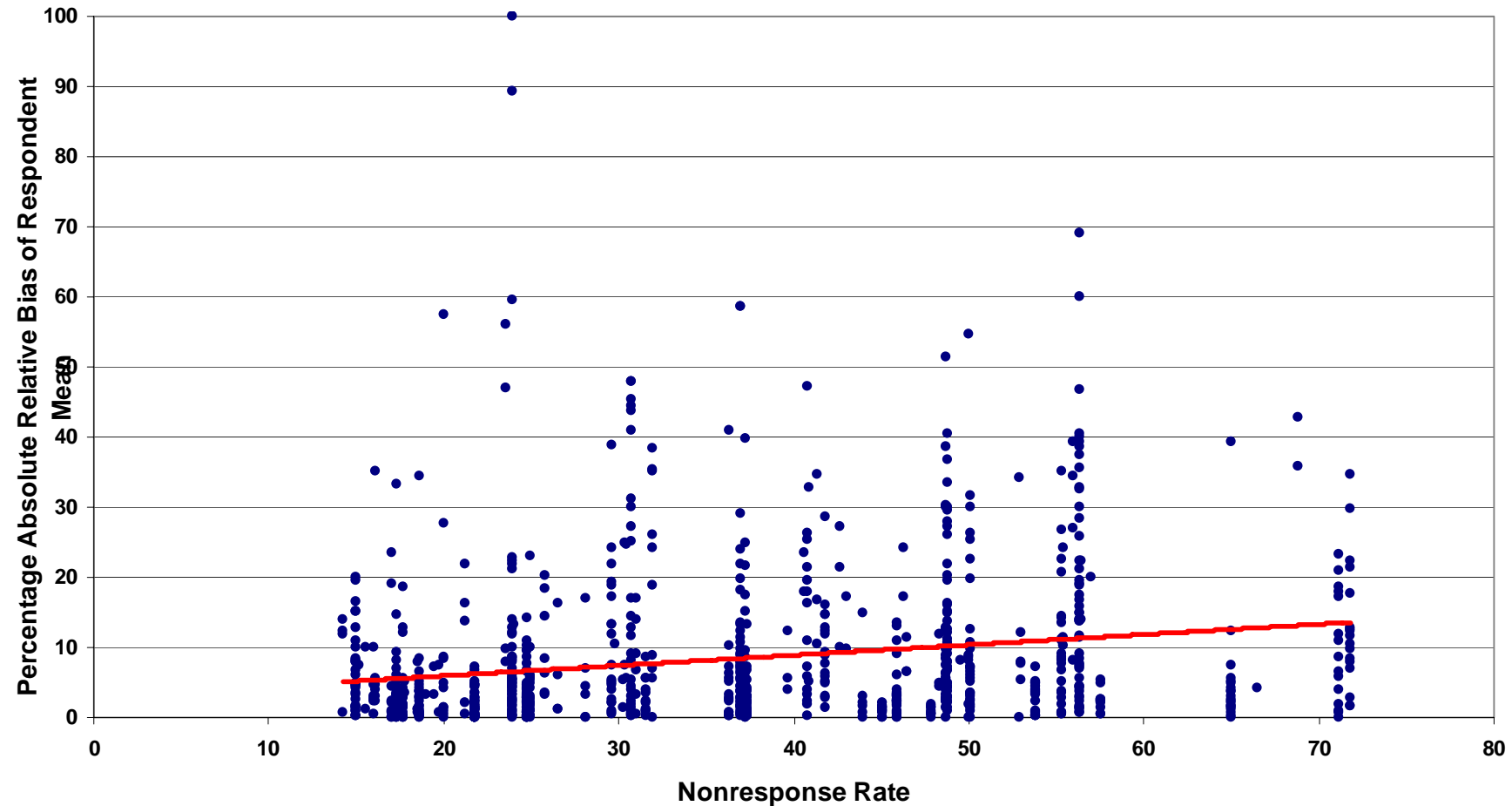
# 1. Nonresponse Bias Happens



## 2. Large Variation in Nonresponse Bias Across Estimates Within the Same Survey, or



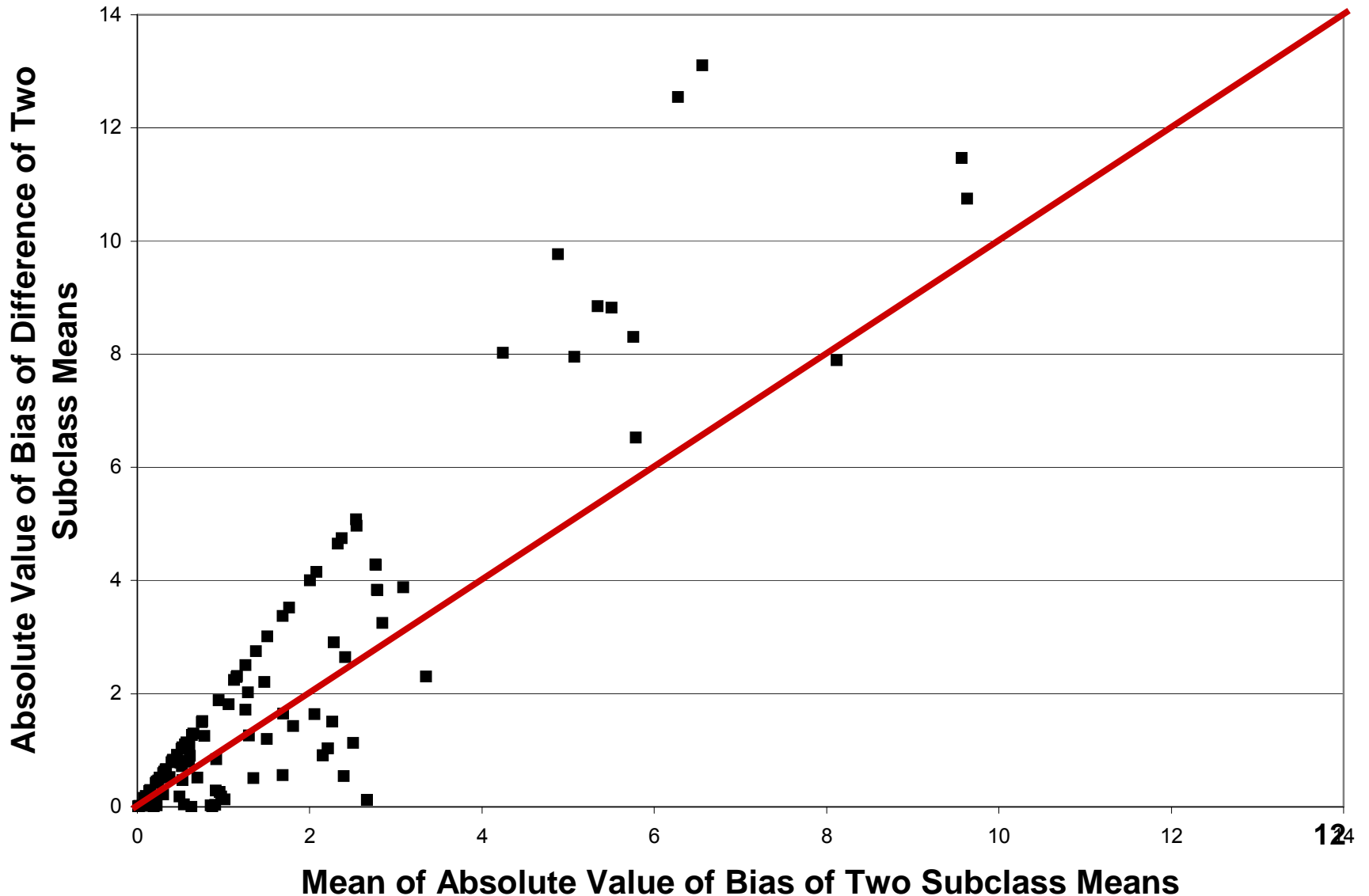
### 3. The Nonresponse Rate of a Survey is a Poor Predictor of the Bias of its Various Estimates (Naïve OLS, $R^2=.04$ )



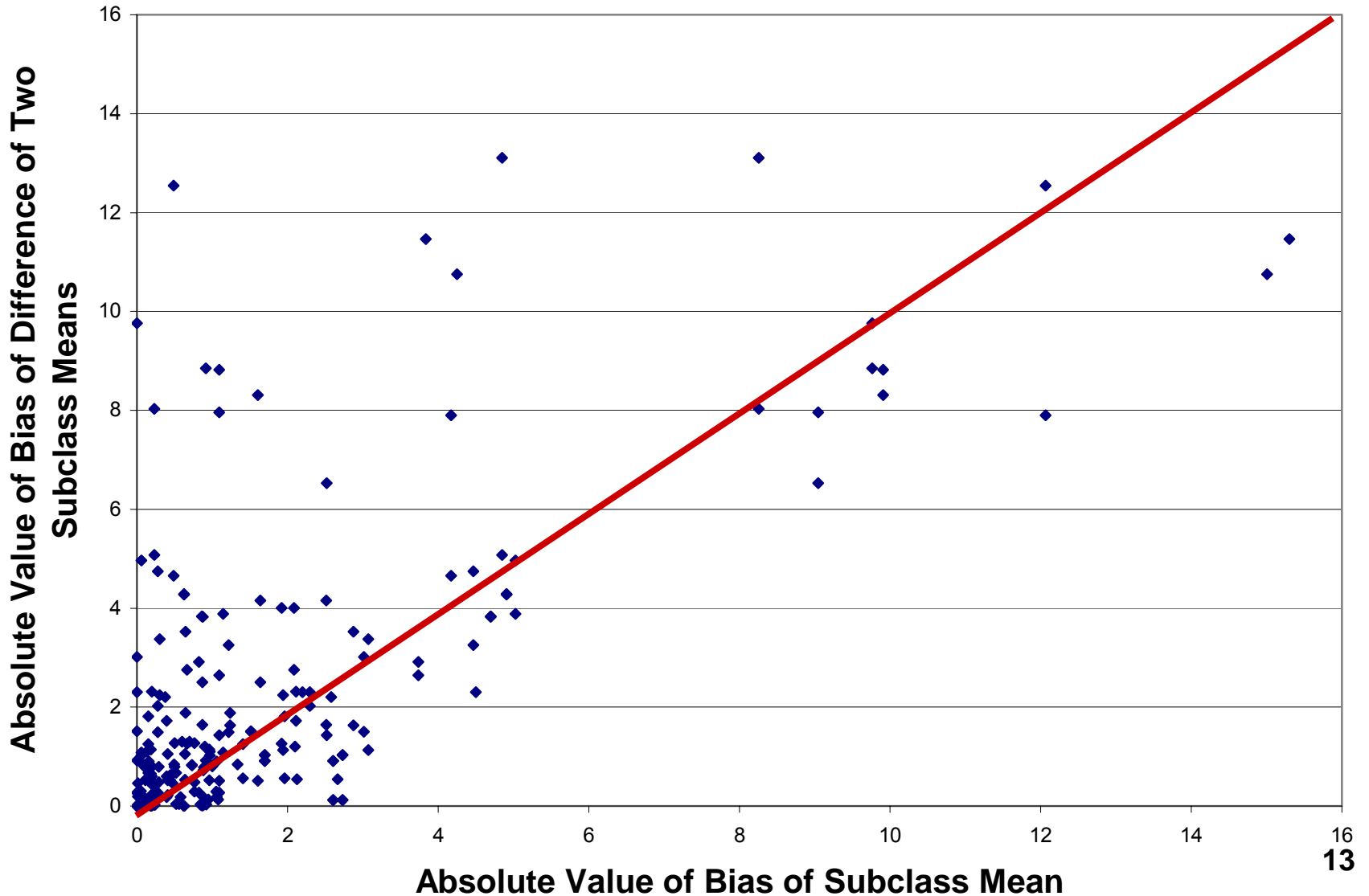
## **Do Differences of Subclass Means have Lower Nonresponse Bias?**

- **When estimating subclass differences, we hope that nonresponse biases of the two estimates cancel**
- **234 reported estimates of subclass means and their 117 differences**
- **Only 41 of the 117 differences have bias of the differences of subclass means lower than average bias of the two subclass means**

# Absolute Value of Bias of 117 Differences of Subclass Means by the Average Absolute Value of the 234 Subclass Means



# Absolute Value of Bias of Differences of Subclass Means by Absolute Value of the 234 Subclass Means



# Does the Bias Level of Demographic Variables Predict the Bias of Substantive Estimates?

- **Common to adjust for socio-demographic variables in hopes of eliminating nonresponse bias**
- **23 studies measure biases on both**
- **Bias levels of demographic variables only weakly related to bias levels of substantive variables ( $r = .28$ , “explained variance”=8%)**
- **Naïve selection bias models or propensity models probably have little effect on bias reduction**

# Nonresponse Error for Sample Mean

In simplest terms 
$$\bar{Y}_r = \bar{Y}_n + \left( \frac{m}{n} \right) [\bar{Y}_r - \bar{Y}_m]$$

OR

Respondent Mean = Full Sample Mean +  
(Nonresponse Rate)\*(Respondent Mean –  
Nonrespondent Mean)

OR

Survey Results = Desired Results + Error

OR

Nonresponse Error = f(Rate, Difference between  
Respondents and Nonrespondents)

# A Stochastic View of Response Propensities

$$\mathbf{Bias}(\bar{y}_r) = \frac{\sigma_{yp}}{\bar{p}} = \left( \frac{\rho_{yp}}{\bar{p}} \right) \sigma_y \sigma_p$$

where  $\sigma_{yp}$  = covariance between  $y$  and response propensity,  $p$

$\bar{p}$  = mean propensity over the sample

$\rho_{yp}$  = correlation between  $y$  and  $p$

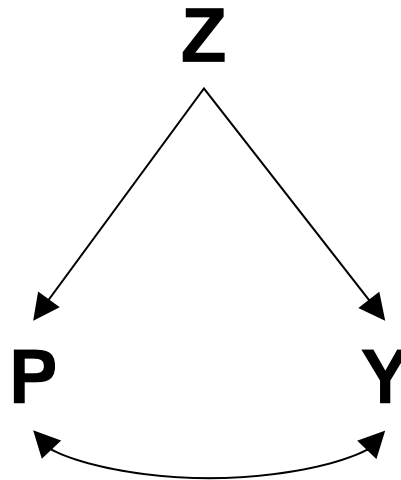
$\sigma_y$  = standard deviation of  $y$

$\sigma_p$  = standard deviation of  $p$

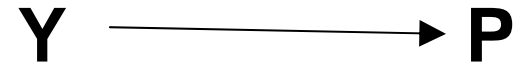
# Viewing Nonresponse Error from a Causal Modeling Perspective



Separate Cause Model



Common Cause Model



Survey Variable Cause Model

# So What to Do?

- **Reduce importance given to response rates**
- **Increase attention to identifying estimates most affected by mechanisms that produce nonresponse**
  
- **At Michigan we are mounting experiments attempting to systematically produce nonresponse bias**

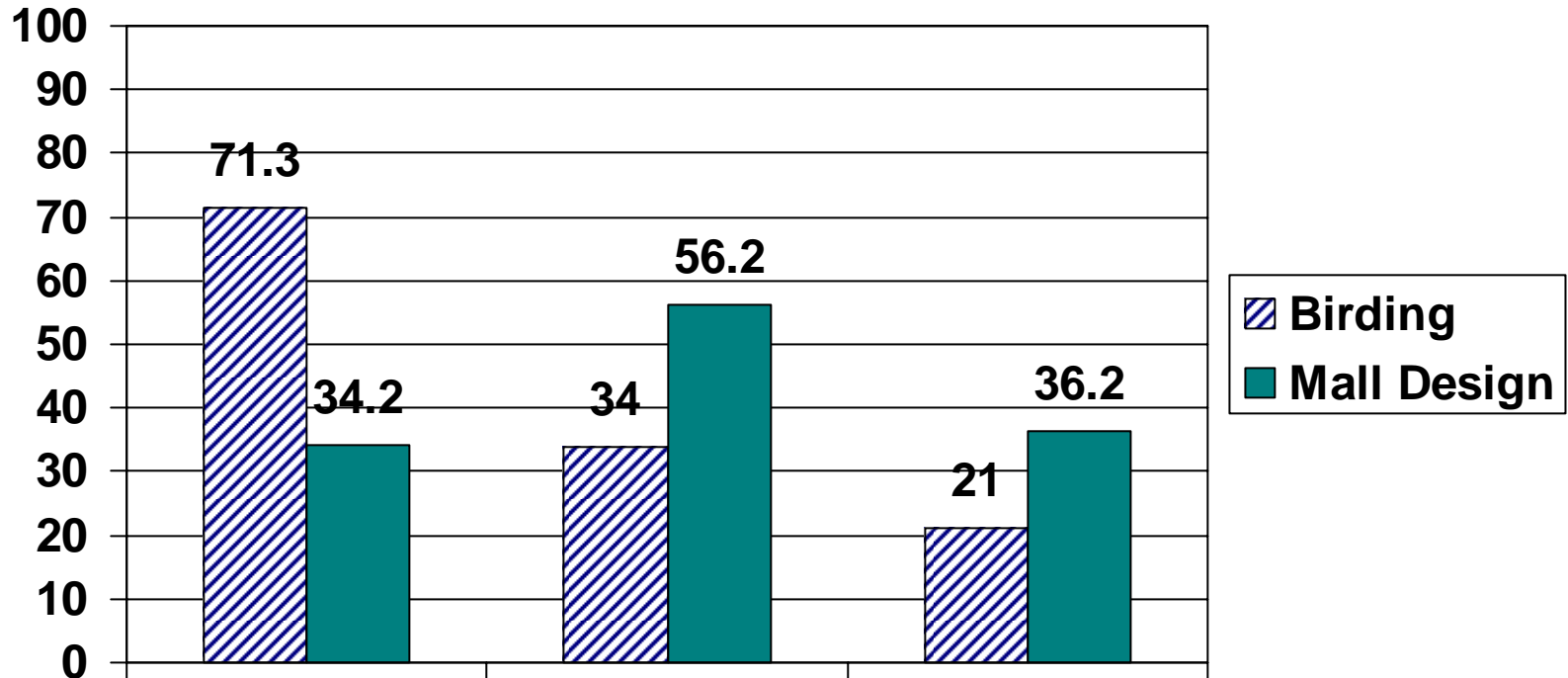
# Three Key Candidate Design Features to Produce Nonignorable Nonresponse

- **Topic interest**
- **Sponsorship**
- **Incentives**

# Experiments Producing Nonresponse Bias

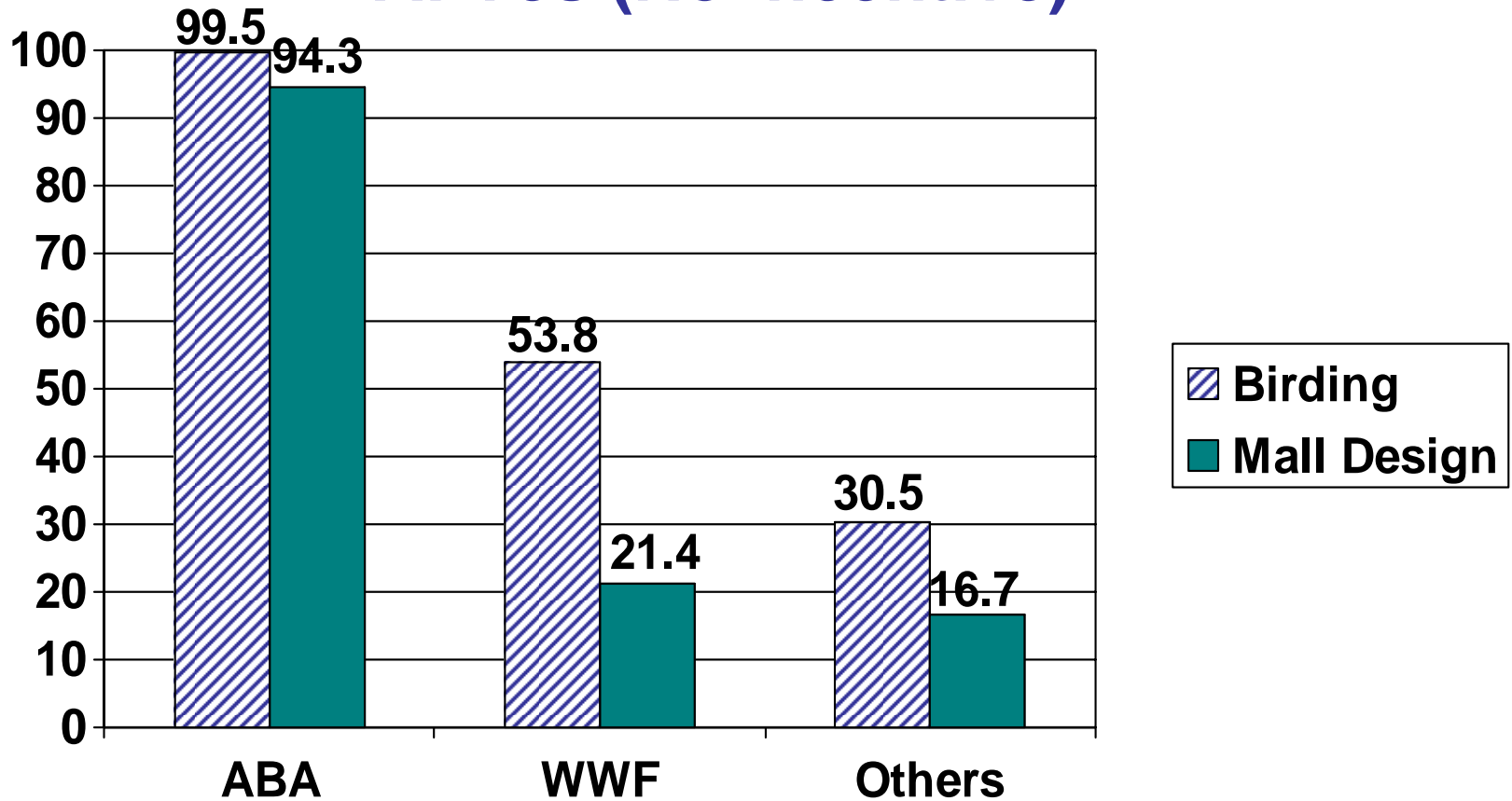
- **Survey topics**
  - **Birds, Bird-watching, and Birding**
  - **Design of Indoor Shopping Malls**
- **Target populations**
  - **members of American Birding Association (ABA)**
  - **donors to World Wildlife Fund (WWF)**
  - **other adults (Others)**

# Response Rate by Topic and Population (No Incentive)



**ABA**  $\chi^2 = 56.368, p < .000$  odds ratio = 5.20  
**WWF**  $\chi^2 = 19.539, p < .0001$  odds ratio = 0.41  
**Others**  $\chi^2 = 10.952, p < .001$  odds ratio = 0.47

# Q1: Did you participate in bird watching/ birding during the past 12 months? A: Yes (No Incentive)



$\chi^2 = 5.265, p < .022$   
odds ratio = 12.03

$\chi^2 = 20.103, p < .0001$   
odds ratio = 4.28

$\chi^2 = 3.023, p < .082$   
odds ratio = 2.19

# Five Summary Statements of What We Know

1. Large nonresponse biases can happen in surveys
2. **Most variation in nonresponse biases lie among estimates in the same survey**
3. The search for mechanisms that link nonresponse rates to error should be focused on the item-level
4. Differences of subclass means do not tend to have lower biases than the individual subclass means
5. Demographic variables are poor predictors of nonresponse bias in the meta-analysis
6. Topic interest, incentives, and sponsors appear to be influences affecting bias

# So, Where's the Research Field Going Now?

- **Reducing emphasis on response rate; increasing emphasis on nonresponse error**
  - OMB statistical standards
  - papers at International Workshop on Nonresponse
- **Building in nonresponse bias studies into surveys**
  - rich frames
  - supplementary records
  - seeded samples
  - interviewer observations of respondents and nonrespondents
  - nonresponse followup studies

# So, Where's the Research Field Going Now?

- **Rethinking right-side variables in propensity and selection models**
  - **Collect auxiliary data helpful to understanding linkage between response propensities and survey variables**
  - **Proxy indicators of the  $y$  variable**