Incorporating Level of Effort Paradata in Nonresponse Adjustments

Paul Biemer
RTI International
University of North Carolina – Chapel Hill

Acknowledgements

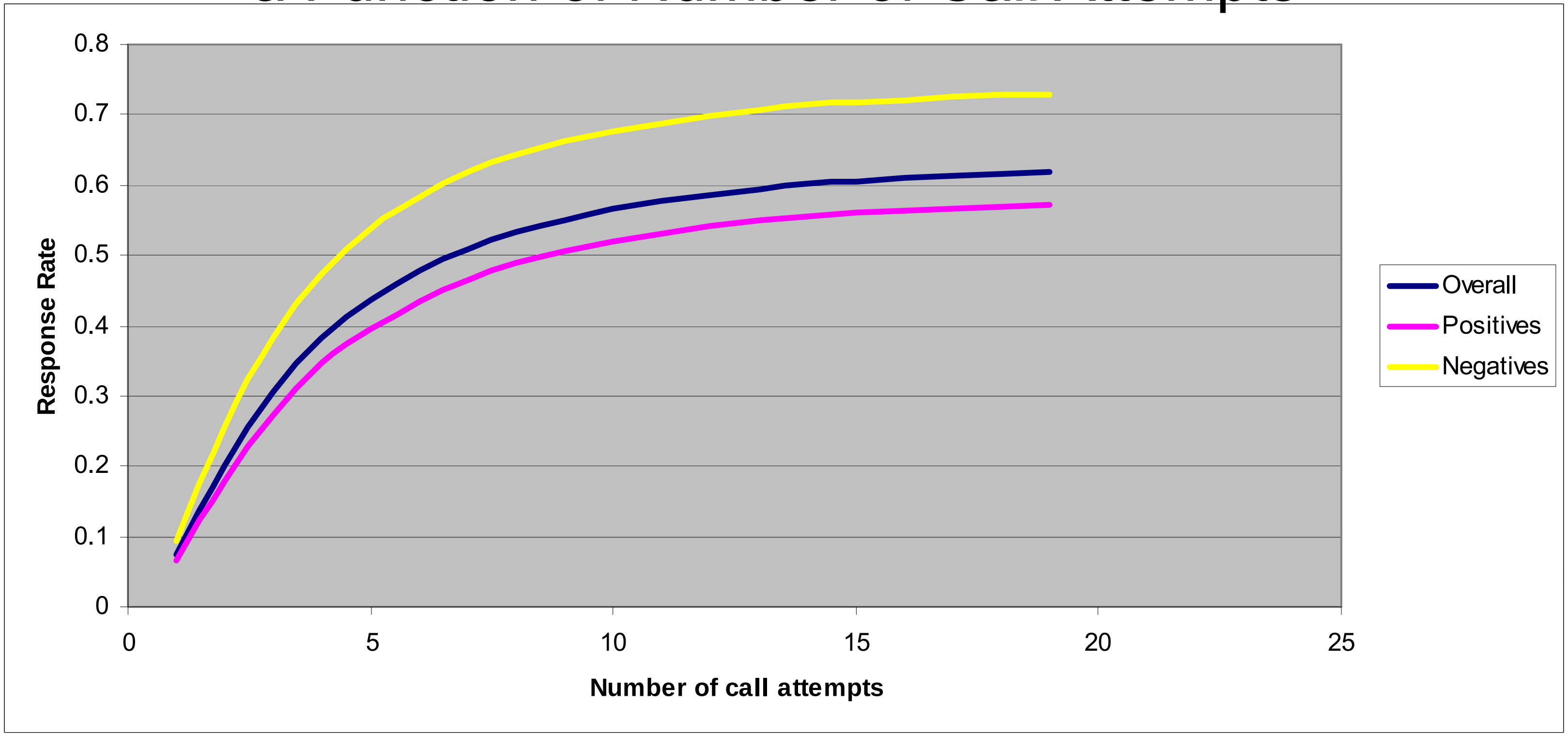
- Patrick Chen, RTI International
- Kevin Wang, RTI International
- SAMHSA

Outline of this talk

- Brief review of the literature
- Discussion of some issues in using LOE paradata
- Simple callback model for dichotomous variables
- Estimation via the EM algorithm
- Testing and adjusting for nonignorable nonresponse (NINR) bias
- Application to the National Survey on Drug Use and Health (NSDUH)
- Summary and Future Directions

What information does LOE paradata contain about response propensity?

Response Rate for Positives, Negatives and Overall as a Function of Number of Call Attempts



How can these data be used to inform the nonresponse (NR) bias mitigation process?

- Determine which variables are most subject to NR bias.
- Determine the severity of nonignorable nonresponse (NINR) bias in estimates adjusted for ignorable NR.
- Monitor NR bias during data collection for optimal reallocation of NR conversion resources.
- Adjust for NR bias in the absence of auxiliary data used as adjustment controls
- Adjust for NINR bias

What prior research has been conducted on the use of callback data for NR adjustment?

Early papers

- H.O. Hartley (1946); Politz-Simmons (1949); Simmons (1954) – Based upon retrospective reports of availability. Crude but sometimes effective.
- Remaining literature is divided between two approaches
  - Regression modeling
  - Probability modeling

Modeling Approaches

Regression Modeling

- (Alho, 1990; Anido-Valdez, 2000; Wood,White & Hotopf, 2006)
- Models response propensity at each attempt as a function of partially missing and fully observed predictors
- Inverse predicted propensities are used as weight adjustments
- Uses an modified conditional likelihood method of esitmation.
- Provides weights that adjust for NINR

Modeling Approaches (cont'd)

Probability Modeling

- (Drew-Fuller,1980; Potthoff, Manton, Woodbury, 1993; Biemer & Link, 2007)
- Models the probability of observation in each cell of the data summary table
- Simultaneously estimates the prevalence of the cross-classifications variables along with the NR parameters
- Partially observed variables completed via the EM algorithm
- Estimates NR propensity components that can be used to build weights that adjust for NINR

What results have been achieved?

- Little has been done to address real-world complexities.
  - Biemer/Link approach was a step in this direction
- Models specifying contact probabilities only have limited applicability for interview surveys.
- No rigorous evaluation of bias reduction capability
  - Validity of the approaches demonstrated on artificial populations or from model fit statistics.
- Studies have shown that NR bias reduction can be dramatic albeit at the cost of increasing variance.

Typical Interview Survey Data Summary Table for a Binary Response Variable, *y*

Call Attempt	Interviewed $y=1$	$y=2$	Ref/Oth NR $y=1$ or 2	NC (noncontact) $y=1$ or 2	Censored NC $y=1$ or 2
1	n_{111}	n_{211}	n_{+12}	n_{+13}	n_{+1c}
2	n_{121}	n_{221}	n_{+22}	n_{+23}	n_{+2c}
…	…	…	…		…
a	n_{1a1}	n_{2a1}	n_{+a2}	n_{+a3}	n_{+ac}
…	…	…	…		…
A (truncated)	n_{1A1}	n_{2A1}	n_{+A2}	n_{+A3}	n_{+Ac}

Censored case → (prematurely) finalized as NC cases

Call Attempt	Interviewed $y=1$	$y=2$	Ref/Oth NR $y=1$ or 2	NC (noncontact) $y=1$ or 2	Censored NC $y=1$ or 2
1	n_{111}	n_{211}	n_{+12}	n_{+13}	n_{+1c}
2	n_{121}	n_{221}	n_{+22}	n_{+23}	n_{+2c}
…	…	…	…		…
a	n_{1a1}	n_{2a1}	n_{+a2}	n_{+a3}	n_{+ac}
…	…	…	…		…
A (truncated)	n_{1A1}	n_{2A1}	n_{+A2}	n_{+A3}	n_{+Ac}

Truncated case → completed after the maximum number of attempts specified by the model; treated as NCs

Call Attempt	Interviewed $y=1$	Interviewed $y=2$	Ref/Oth NR $y=1$ or 2	NC (noncontact) $y=1$ or 2	Censored NC $y=1$ or 2
1	n_{111}	n_{211}	n_{+12}	n_{+13}	n_{+1c}
2	n_{121}	n_{221}	n_{+22}	n_{+23}	n_{+2c}
…	…	…	…		…
a	n_{1a1}	n_{2a1}	n_{+a2}	n_{+a3}	n_{+ac}
…	…	…	…		…
A (max att.)	n_{1A1}	n_{2A1}	n_{+A2}	n_{+A3}	n_{+Ac}

NC cases at attempt *a* are available at attempt *a*+1.

Call Attempt	Interviewed $y=1$	$y=2$	Ref/Oth NR $y=1$ or 2	NC (noncontact) $y=1$ or 2	Censored NC $y=1$ or 2
1	n_{111}	n_{211}	n_{+12}	n_{+13}	n_{+1c}
2	n_{121}	n_{221}	n_{+22}	n_{+23}	n_{+2c}
…	…	…	…		…
a	n_{1a1}	n_{2a1}	n_{+a2}	n_{+a3}	n_{+ac}
…	…	…	…		…
A (truncated)	n_{1A1}	n_{2A1}	n_{+A2}	n_{+A3}	n_{+Ac}

What are some issues that arise in callback modeling?

- What constitutes an attempt?
  - Definition varies by mode of data collection
  - E.g., is several calls within several hours 1 attempt or multiple attempts
  - Use the definition that is most predictive of the model parameters
- What constitutes a contact?
  - First contact with anyone in the HH?
  - First contact with sample person or guardian?
  - The contact that determines final disposition of the case (i.e., interview, refused, other)
  - Other?

- How should censoring and truncation be modeled?
  - If censoring mechanism is independent of *y*, it can be ignored.
  - However, standard errors of model parameters will be larger.
- How should weighting be handled in the modeling process?
  - Unweighted data – e.g., Pr(*i*th unit responds | *i*th unit is sampled)
  - Weight for probabilities of selection
  - Weight for selection probs and NR using ignorable NR models
- What callback model should be used?
  - Regression model does not adapt well to these complexities
  - Probability model adapts well but has other shortcomings

Essential Idea: Model missing data mechanism and …

Call Attempt	Interviewed $y=1$	Interviewed $y=2$	Ref/Oth NR $y=1$ or 2	Censored NC $y=1$ or 2
1	n_{111}	n_{211}	n_{+12}	n_{+1c}
2	n_{121}	n_{221}	n_{+22}	n_{+2c}
…	…	…	…	…
a	n_{1a1}	n_{2a1}	n_{+a2}	n_{+ac}
…	…	…	…	…
A (truncated)	n_{1A1}	n_{2A1}	n_{+A2}	n_{+Ac}

…complete the table.

	Interviewed		Refusals/Other NR		Censored	
Attempt	y=1	y=2	y=1	y=2	y=1	y=2
1	n_{111}	n_{211}	n_{112}	n_{212}	n_{113}	n_{213}
…	…	…	…	…	…	…
a	n_{1a1}	n_{2a1}	n_{1a2}	n_{2a2}	n_{1a3}	n_{2a3}
…	…	…	…	…	…	…
A	n_{1A1}	n_{2A1}	n_{1A2}	n_{2A2}	n_{1A3}	n_{2A3}
Totals	n_{1+1}	n_{2+1}	n_{1+2}	n_{2+2}	n_{1+3}	n_{2+3}

This will yield a model unbiased estimate of prevalence.

	Interviewed		Refusals/Other NR		Censored	
Attempt	y=1	y=2	y=1	y=2	y=1	y=2
1	n_{111}	n_{211}	n_{112}	n_{212}	n_{113}	n_{213}
…	…	…	…	…	…	…
a	n_{1a1}	n_{2a1}	n_{1a2}	n_{2a2}	n_{1a3}	n_{2a3}
…	…	…	…	…	…	…
A	n_{1A1}	n_{2A1}	n_{1A2}	n_{2A2}	n_{1A3}	n_{2A3}
Totals	n_{1+1}	n_{2+1}	n_{1+2}	n_{2+2}	n_{1+3}	n_{2+3}

Estimate prevalence: $$\hat{\pi} = \frac{n_{1+1} + n_{1+2} + n_{1+3}}{n}$$

Basic Call-back Model for Binary Response

Notation

Let

$a =$ call attempt;

$= 1,\ldots,A$

$b =$ call outcome;

$=1$ for interview,

$=2$ for refusal/other nr

$=3$ for noncontact

Basic Call-back Model for Binary Response

$$\pi = \text{ true prevalence}$$

$$\alpha_{y,a} = \Pr(\text{contact at attempt } a | \, y)$$

$$\beta_{y,a} = \Pr(b = 1) = \Pr(\text{interview} | \, y, \text{ contact at attempt } a)$$

$$1 - \beta_{y,a} = \Pr(b = 2) = \Pr(\text{refusal/oth NR} | \, y, \text{ contact at attempt } a)$$

$$\delta = \Pr(\text{censored at attempt } a)$$

Simple Call-back Model for Binary Response

Assume:

$$\alpha_{y,a} = \alpha_y \text{ for all } a$$

$$\beta_{y,a} = \beta_y \text{ for } a$$

For a binary response variable, *y*, this results in six parameters: $\pi,\ \alpha_1, \alpha_2,\ \beta_1,\ \beta_2,$ and δ

Cell Frequencies and Probabilities

	Interviewed		Refusals/ Other NR	Censored NCs
	y=1	*y*=2	*y* is unknown	*y* is unknown
Cell freq $a = 1, \ldots, A$	n_{1a1}	n_{2a1}	n_{+a2}	n_{+a3}
, Prob $a = 1, \ldots, A$	π_{1a1}	π_{2a1}	$\pi_{+a2} = \pi_{1a2} + \pi_{2a2}$	$\pi_{+a3} = \pi_{1a3} + \pi_{2a3}$

Probabilities for Basic Model

Interview: *y*=1 $$\pi_{1a1} = (1-\delta)^{a-1}[\pi(1-\alpha_1)^{a-1}\alpha_1\beta_1]$$

Interview: *y*=2 $$\pi_{2a1} = (1-\delta)^{a-1}[(1-\pi)(1-\alpha_2)^{a-1}\alpha_2\beta_2]$$

Refused $$\pi_{+a2} = (1-\delta)^{a-1}[\pi(1-\alpha_1)^{a-1}\alpha_1(1-\beta_1) + (1-\pi)(1-\alpha_2)^{a-1}\alpha_2(1-\beta_2)]$$

NC – Censored $$\pi_{+a3} = (1-\delta)^{a-1}\delta[\pi(1-\alpha_1)^{a} + (1-\pi)(1-\alpha_2)^{a}]$$

EM Algorithm to Estimate Parameters

E-step at *t*th iteration

Refusals $\hat{n}_{1a2}^{[t]} = n_{+a2} \frac{\hat{\pi}_{1a2}^{[t]}}{\hat{\pi}_{1a2}^{[t]} + \hat{\pi}_{2a2}^{[t]}}, \quad \hat{n}_{2a2}^{[t]} = n_{+a2} \frac{\hat{\pi}_{2a2}^{[t]}}{\hat{\pi}_{1a2}^{[t]} + \hat{\pi}_{2a2}^{[t]}}$

NC-censored $\hat{n}_{1a3}^{[t]} = n_{+a3} \frac{\hat{\pi}_{1a3}^{[t]}}{\hat{\pi}_{1a3}^{[t]} + \hat{\pi}_{2a3}^{[t]}}, \quad \hat{n}_{1a3}^{[t]} = n_{+a3} \frac{\hat{\pi}_{2a3}^{[t]}}{\hat{\pi}_{1a3}^{[t]} + \hat{\pi}_{2a3}^{[t]}}$

EM Algorithm to Estimate Parameters

M-step at (t+1)th iteration

$$\hat{\alpha}_1^{[t+1]} = \frac{\sum_a (n_{1a1} + \hat{n}_{1a2}^{[t]})}{\sum_a a(n_{1a1} + \hat{n}_{1a2}^{[t]} + \hat{n}_{1a3}^{[t]})}; \quad \hat{\alpha}_2^{[t+1]} = \frac{\sum_a (n_{2a1} + \hat{n}_{2a2}^{[t]})}{\sum_a a(n_{2a1} + \hat{n}_{2a2}^{[t]} + \hat{n}_{2a3}^{[t]})}$$

$$\hat{\beta}_1^{[t+1]} = \frac{\sum_a n_{1a1}}{\sum_a (n_{1a1} + \hat{n}_{1a2}^{[t]})}; \quad \hat{\beta}_2^{[t+1]} = \frac{\sum_a n_{2a1}}{\sum_a (n_{2a1} + \hat{n}_{2a2}^{[t]})}$$

$$\hat{\delta} = \frac{\sum_a n_{+a3}}{n}$$

Estimate of Prevalence

$$\pi^{[t+1]} = \frac{\sum_a (n_{1a1} + \hat{n}_{1a2}^{[t]} + \hat{n}_{1a3}^{[t]})}{\sum_y \sum_a (n_{ya1} + \hat{n}_{ya2}^{[t]} + \hat{n}_{ya3}^{[t]})}$$

Tests of Ignorability

$$H_0 : \alpha_1 = \alpha_2 \text{ and } \beta_1 = \beta_2$$

$$H_1 : \alpha_1 \neq \alpha_2 \text{ and } \beta_1 = \beta_2 \quad \text{or} \quad H_1' : \alpha_1 \neq \alpha_2 \quad \text{or} \quad \beta_1 \neq \beta_2$$

Test statistics

$$G^2(M_0) - G^2(M_1) \sim \chi_1^2 \qquad G^2(M_0) - G^2(M_1') \sim \chi_2^2$$

$$G^2(M) = 2\sum_c n_c \ln \frac{n_c}{\hat{n}_{c,M}}$$

Tests of Ignorability

$$H_0 : \alpha_1 = \alpha_2 \text{ and } \beta_1 = \beta_2$$

$$H_1 : \alpha_1 \neq \alpha_2 \text{ and } \beta_1 = \beta_2 \quad \text{or} \quad H_1' : \alpha_1 \neq \alpha_2 \quad \text{or} \quad \beta_1 \neq \beta_2$$

Test statistics

Preferred

$$G^2(M_0) - G^2(M_1) \sim \chi_1^2 \qquad G^2(M_0') - G^2(M_1) \sim \chi_2^2$$

$$G^2(M) = 2\sum_c n_c \ln \frac{n_c}{\hat{n}_{c,M}}$$

Indicators of Nonignorable Bias Based on M_1

Designed to measure the magnitude of the NR bias prior to, during and after traditional NR weighting

$$\Delta G^2 = G^2(M_0) - G^2(M_1)$$

$$R = \frac{G^2(M_0) - G^2(M_1)}{G^2(M_0)}$$

$$D = \frac{G^2(M_0)/df_0 - G^2(M_1)/df_1}{G^2(M_0)/df_0}$$

Indicators of Nonignorable Bias Based on M_1

Designed to measure the magnitude of the NR bias prior to, during and after traditional NR weighting

$$\Delta G^2 = G^2(M_0) - G^2(M_1)$$

Preferred

$$R = \frac{G^2(M_0) - G^2(M_1)}{G^2(M_0)}$$

$$D = \frac{G^2(M_0)/df_1 - G^2(M_1)/df_0}{G^2(M_0)/df_0}$$

Application – National Survey on Drug Use and Health

- NSDUH is quarterly survey to estimate the prevalence of drug, alcohol and tobacco use in U.S.
- 170,000 households are screened and 67,500 interviews are conducted per year
- Response rates are approx. 90% (screener) and 78% (interview)
- Only screener respondents are used in our analysis (i.e., adjustments pertain only to interview survey nonresponse)

Current NSDUH NR Adjustment

- Uses the GEM (generalized exponential model), a logistic regression response propensity adjustment
- Incorporates 13 grouping variables and their interactions including a number of state specific components
- This model will be referred to as the GEM model
- We also considered the GEM+ model obtained by simply adding the LOE variable to the GEM model.

Definitions

- Call attempt – attempt to contact recorded by I'er; similar to call slots (morning, afternoon, evening of same day)
- Contact attempt – first call attempt resulting in face to face contact with the sample member
- Contact outcomes – interview, refused, other NR and NC (censored)

Some Research Questions Addressed in this Research

- Will the test for ignorability be rejected for key estimates?
- Is the probability callback model a valid approach remove the NINR bias?
- Which model works best?
- How do the probability model results compare with simply adding the LOE variable (i.e., number of call attempts) to the GEM model?

Some Callback Models Considered

Model		
Mod0 –	$\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$	Homogeneous contact and interview probabilities (GEM)
Mod1 –	α_1, α_2 and $\beta_1 = \beta_2 = \beta$	Homogeneous contact probs Heterogeneous interview probs
Mod2 –	$\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$ and $\beta_1 = \beta_2 = \beta$	Same as Mod1 except contact probs change after attempt 1
Mod3 –	$\alpha_1 = \alpha_2 = \alpha$ and β_1, β_2	Homogeneous contact probs Heterogeneous interview probs
Mod4 –	α_1, α_2 and β_1, β_2	Heterogeneous contact and interview probs

Our Approach

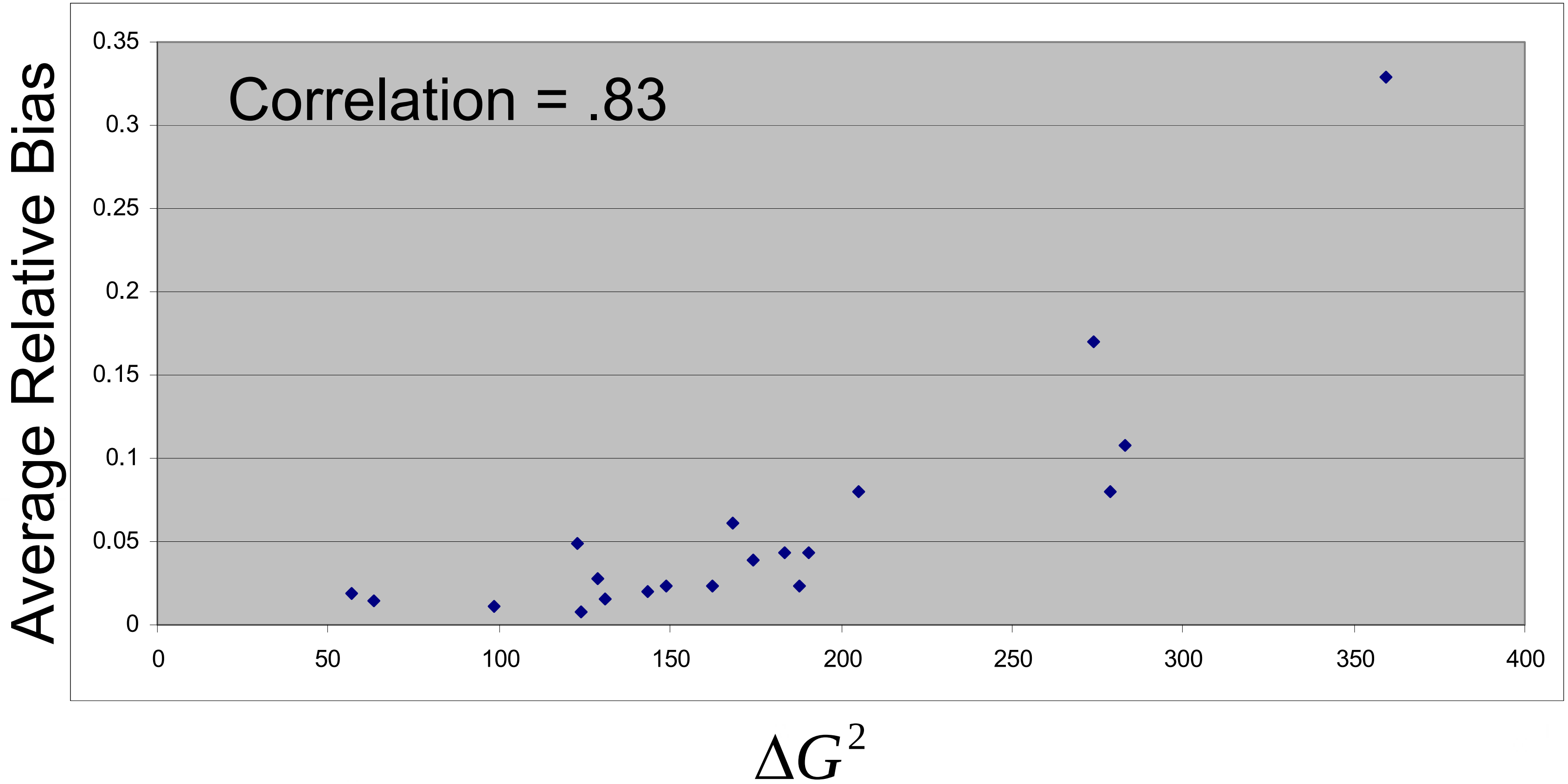
- Fit the standard GEM model to obtain estimated response propensities for each unit
  - To create NINR bias for some variable *x**, *x** was omitted from the GEM model.
  - e.g., $\text{logit}(p_i) = \mathbf{X}'_{-x^*}\boldsymbol{\beta}$ makes *x** an ignored variable in the estimation of response propensity, p_i
  - Choices for *x** included age, race and sex
- Divide the sample into 20 strata based upon the propensity, p_i
- Compare the estimates of the missing variable, x*, across the NINR adjustment models

Tests of Ignorability

Tests of H_0 vs. H_1

Ignored Variable, *x**	ΔG^2	**df**	χ^2_{df}	$\alpha_y = \alpha$
Age	174	4	9.5	Rejected
Race	172	4	9.5	Rejected
Sex	62	1	3.8	Rejected
Alcohol Use	60	1	3.8	Rejected
Marijuana Use	135	1	3.8	Rejected
Cocaine Use	165	1	3.8	Rejected

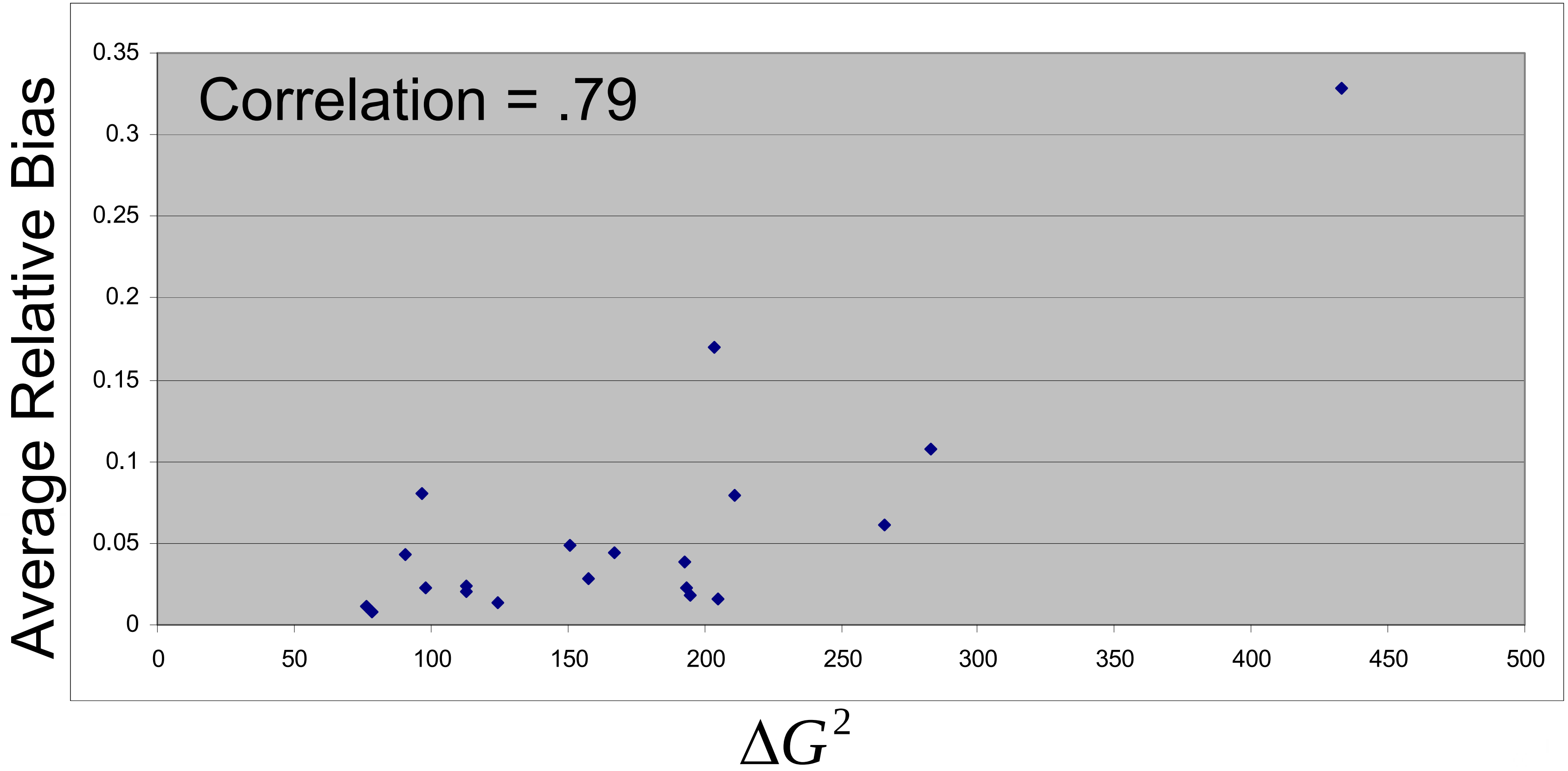
Average Relative Bias as a Function of ΔG^2 for 20 Propensity Strata: AGE



RTI INTERNATIONAL 50 YEARS

The ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Average Relative Bias as a Function of ΔG^2 for 20 Propensity Strata: RACE



RTI INTERNATIONAL 50 YEARS

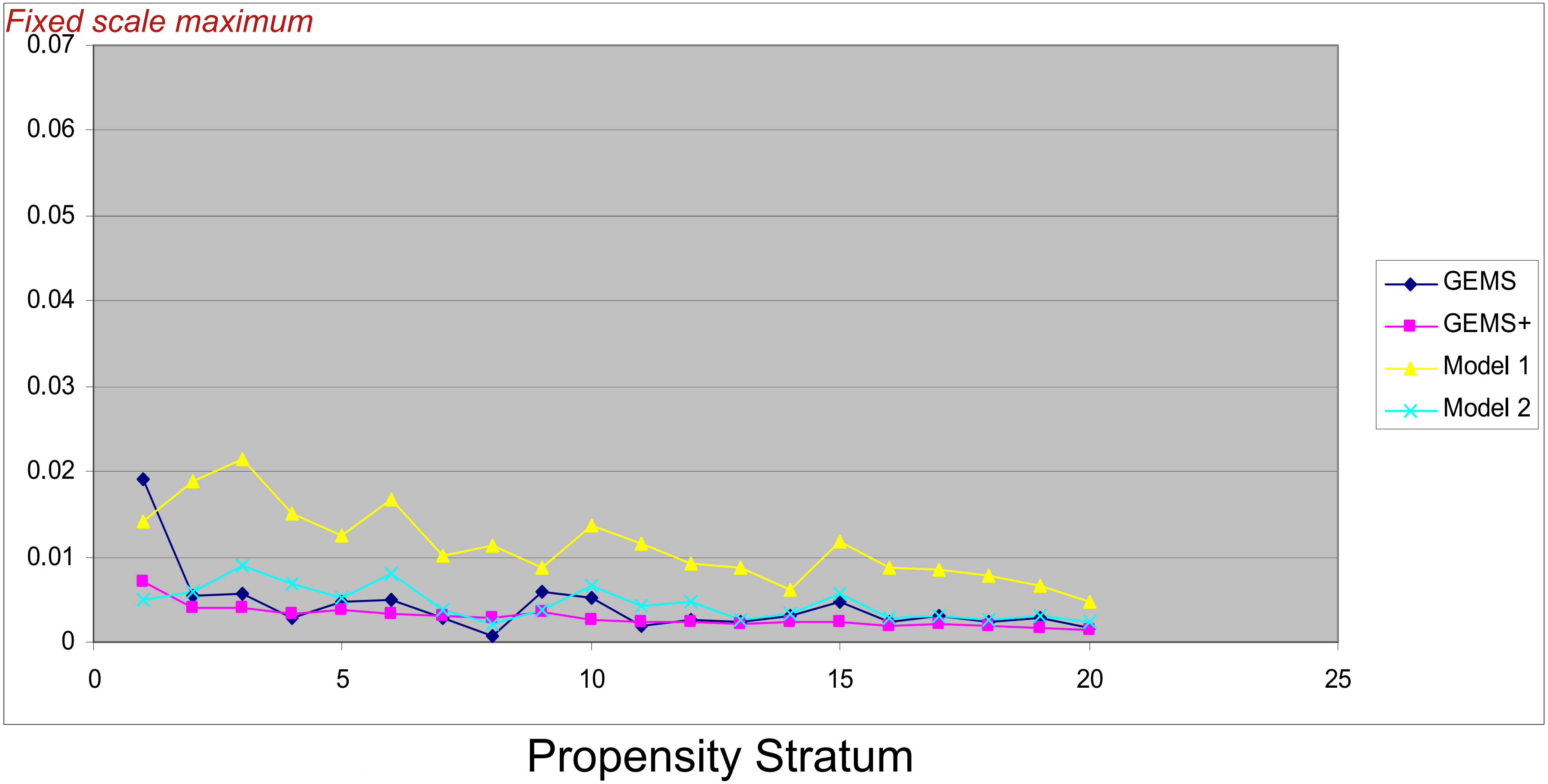
The ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

Comparison of Estimates

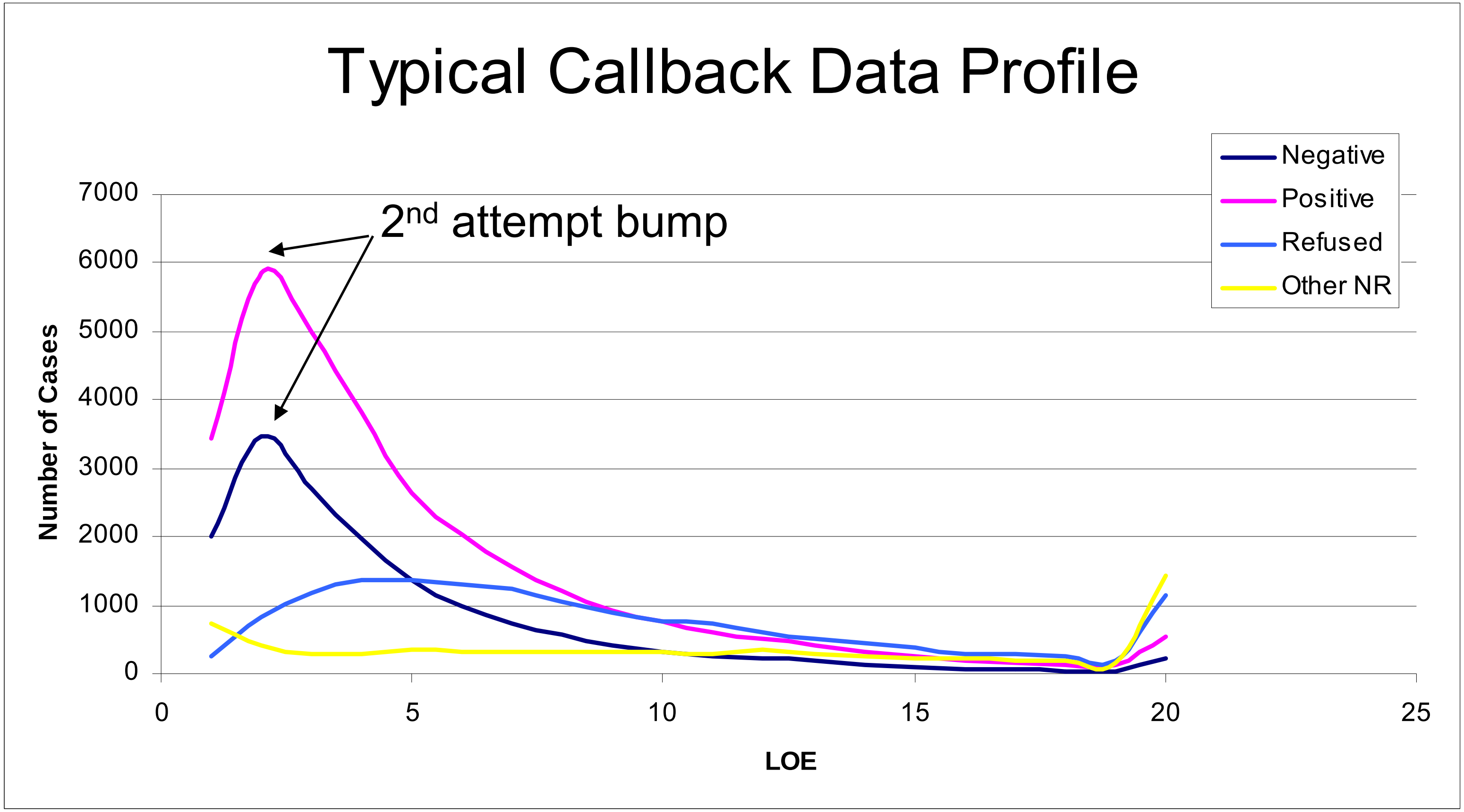
Problems with Heterogeneous β's Models

- Heterogeneous outcome probability models (e.g., Mod3 and Mod4) performed quite poorly
- Based upon a simulation study, behavior of these models appears consistent with callback data recording errors
  - Suppose some proportion of callbacks are not recorded
  - It can be shown that estimates of the callback model parameters are biased.
  - Biases were generally low for the homogeneous β's models
  - Biases were quite large for the heterogeneous β's models – particularly for small values of π
- These models will not be considered further in the results.

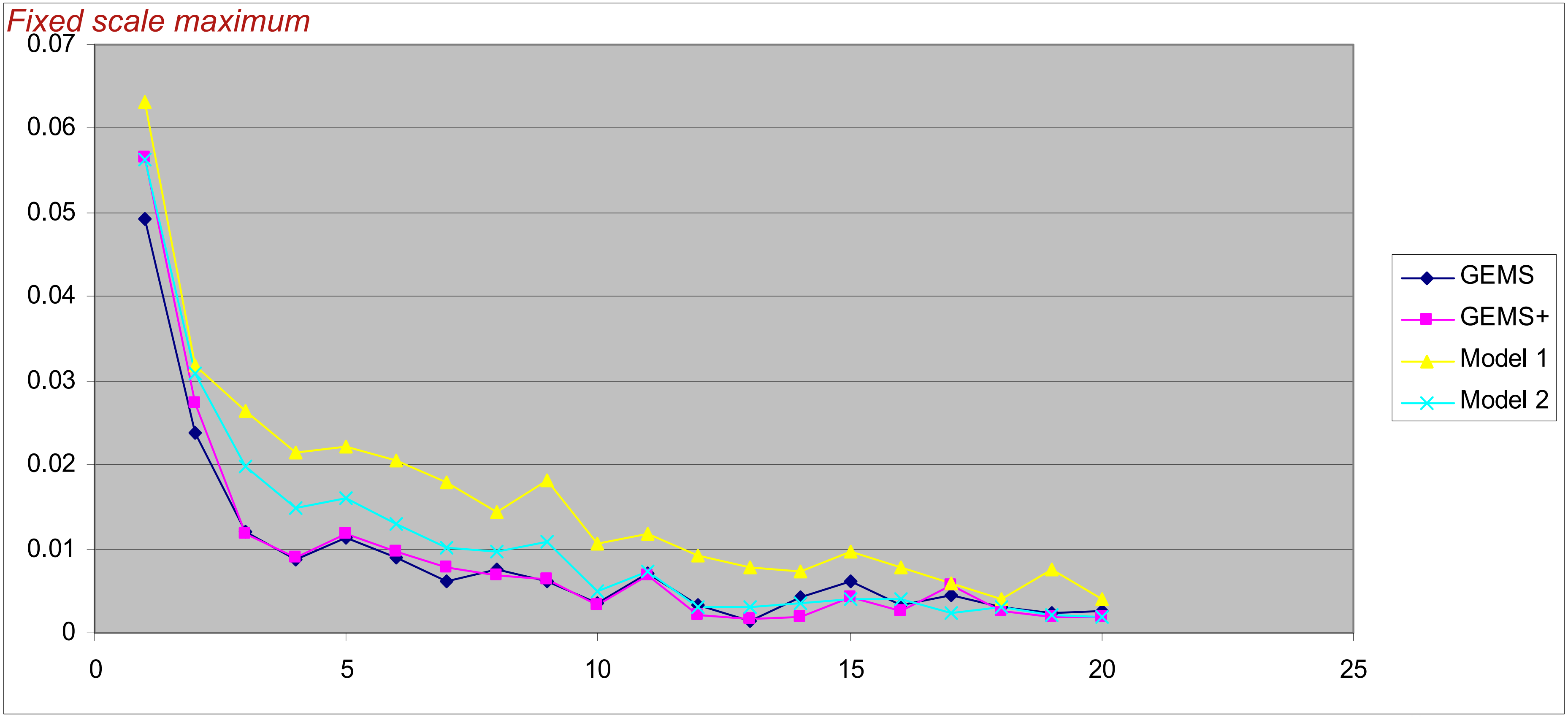
Average Bias for Four Models: RACE



Mod2 Provides for Change in Contact Probabilities after Initial Attempt

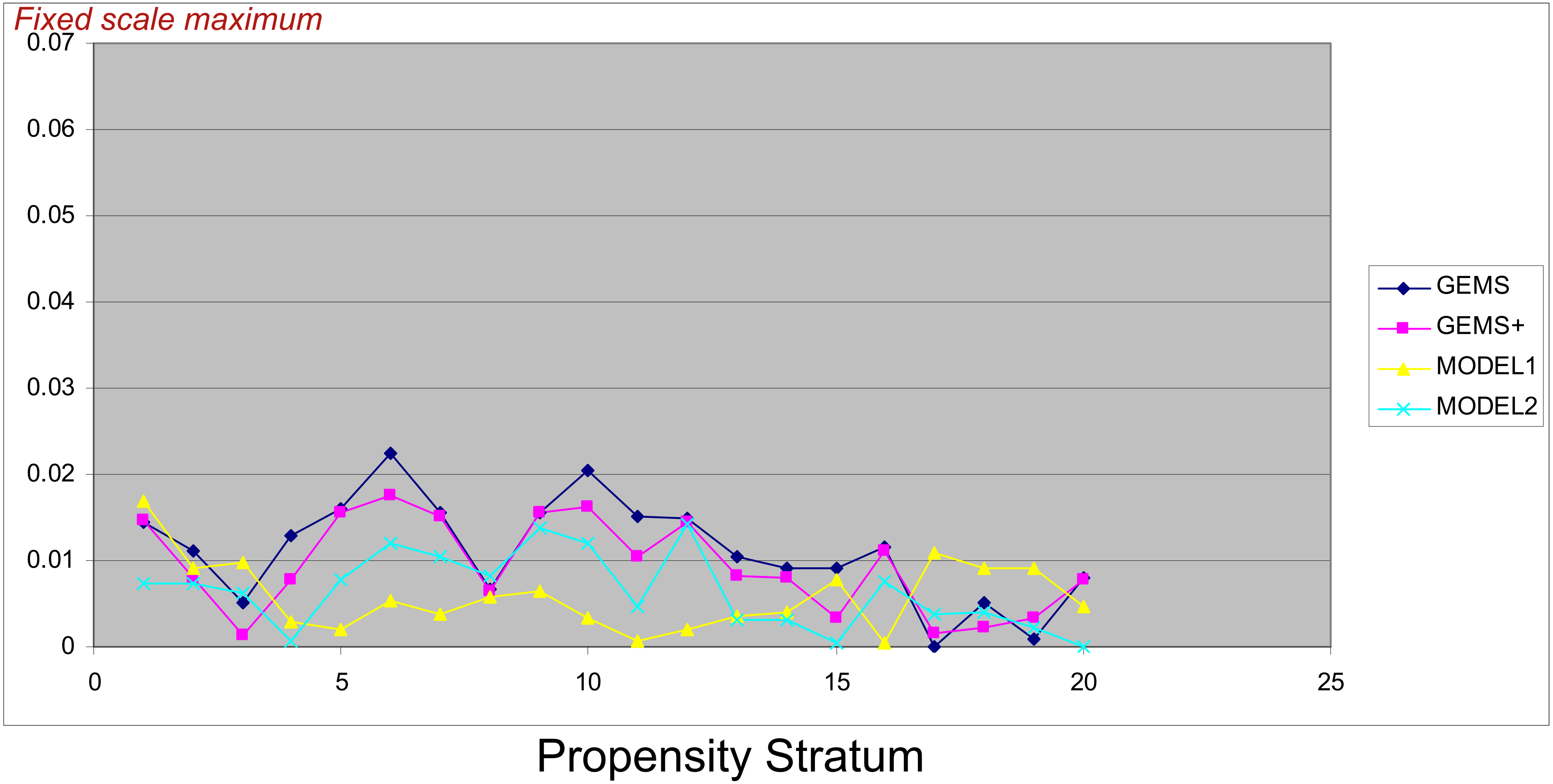


Average Bias for Four Models: AGE

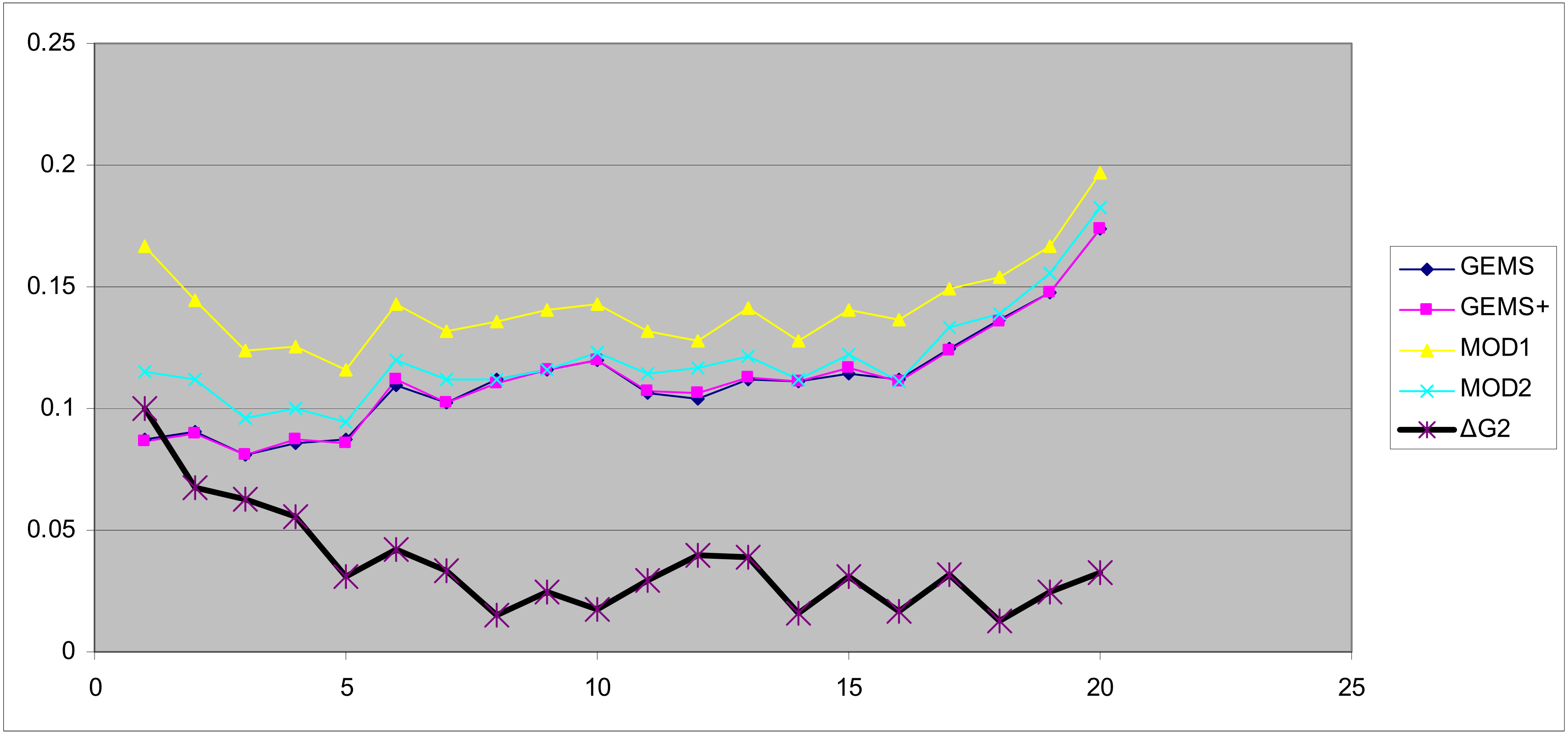


Propensity Stratum

Average Bias for Four Models: SEX

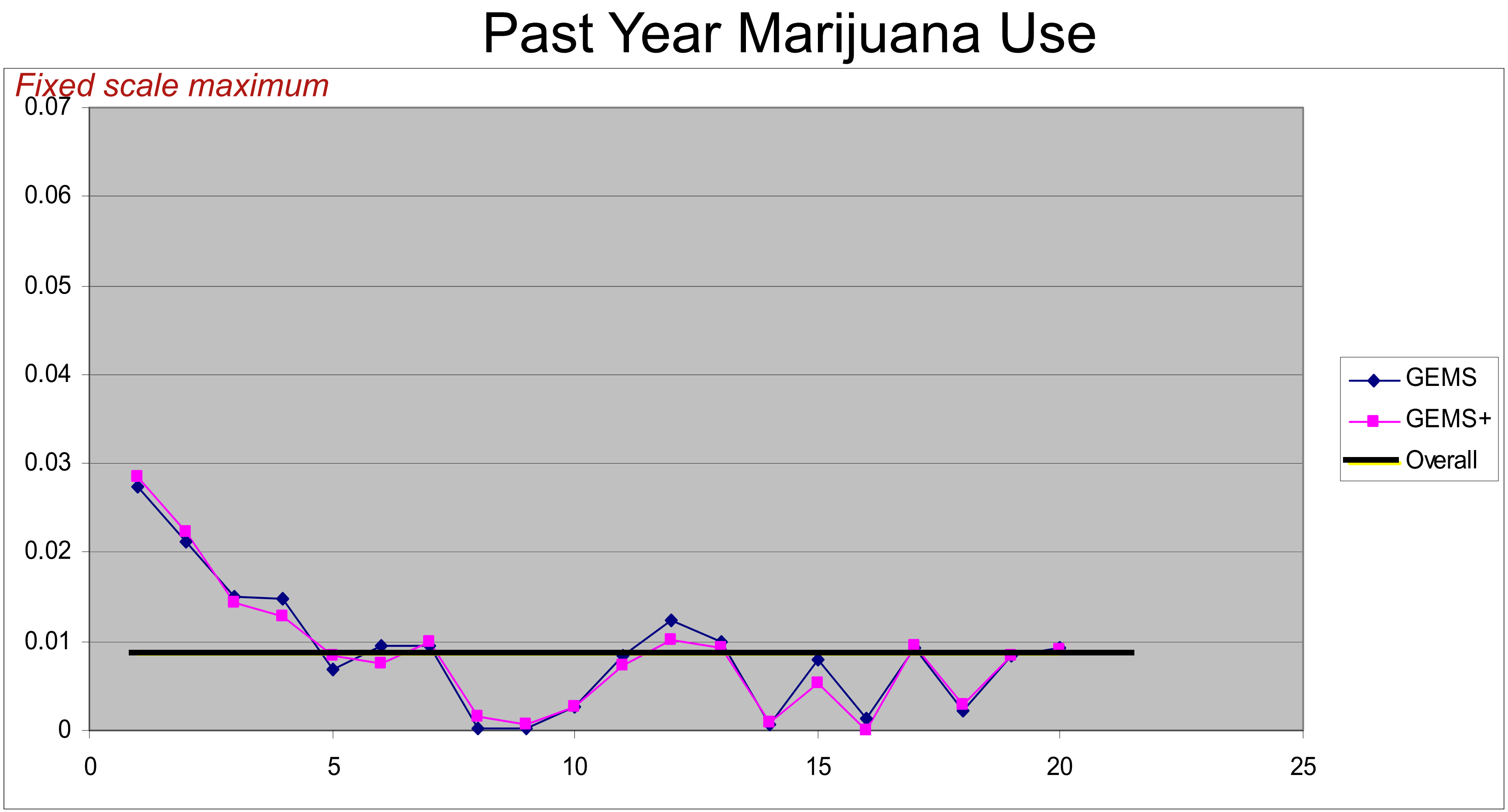


Estimates of Past Year Marijuana Use and ΔG^2 (Rescaled) by Propensity Group



Propensity Stratum

Bias in GEM and GEM+ Adjustments Using Mod2 as the Gold Standard



Summary and Future Directions

Summary

- χ^2 NINR indexes performed well in these tests
- Can be applied post-survey weighting as a check on residual NINR bias
- Models with heterogeneous contact probabilities performed well
- Models with heterogeneous interview probabilities did not perform well
- Suspect the problem is errors in the callback data
- Surprisingly, the GEM+ model was also effective at eliminating NINR bias.
  - It outperformed the probability callback models for some estimates

Future Directions

Model Improvements

- Alternative definitions for call attempt, contact attempt and contact outcome
- Expand application of the models to other variables
- Consider modifications to field procedures for recording of call attempts

Other Applications

- Optimization of fieldwork (e.g., responsive design)
- Representivity measures for international surveys