Responsive Design Using Mahalanobis Distancing: Application on Two National Center for Education Statistics Longitudinal Surveys

Ted Socha & Elise Christopher

National Center for Education Statistics

FedCASIC – March 20, 2013

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

B&B

ELS 2002 STUDENTS PARENTS SCHOOLS

Roadmap

1. Responsive design model choice
2. Application of responsive design models
3. Implications for monetary and non-monetary incentives
4. Substantive and paradata model variables
5. Preliminary results

ies NATIONAL CENTER FOR EDUCATION STATISTICS
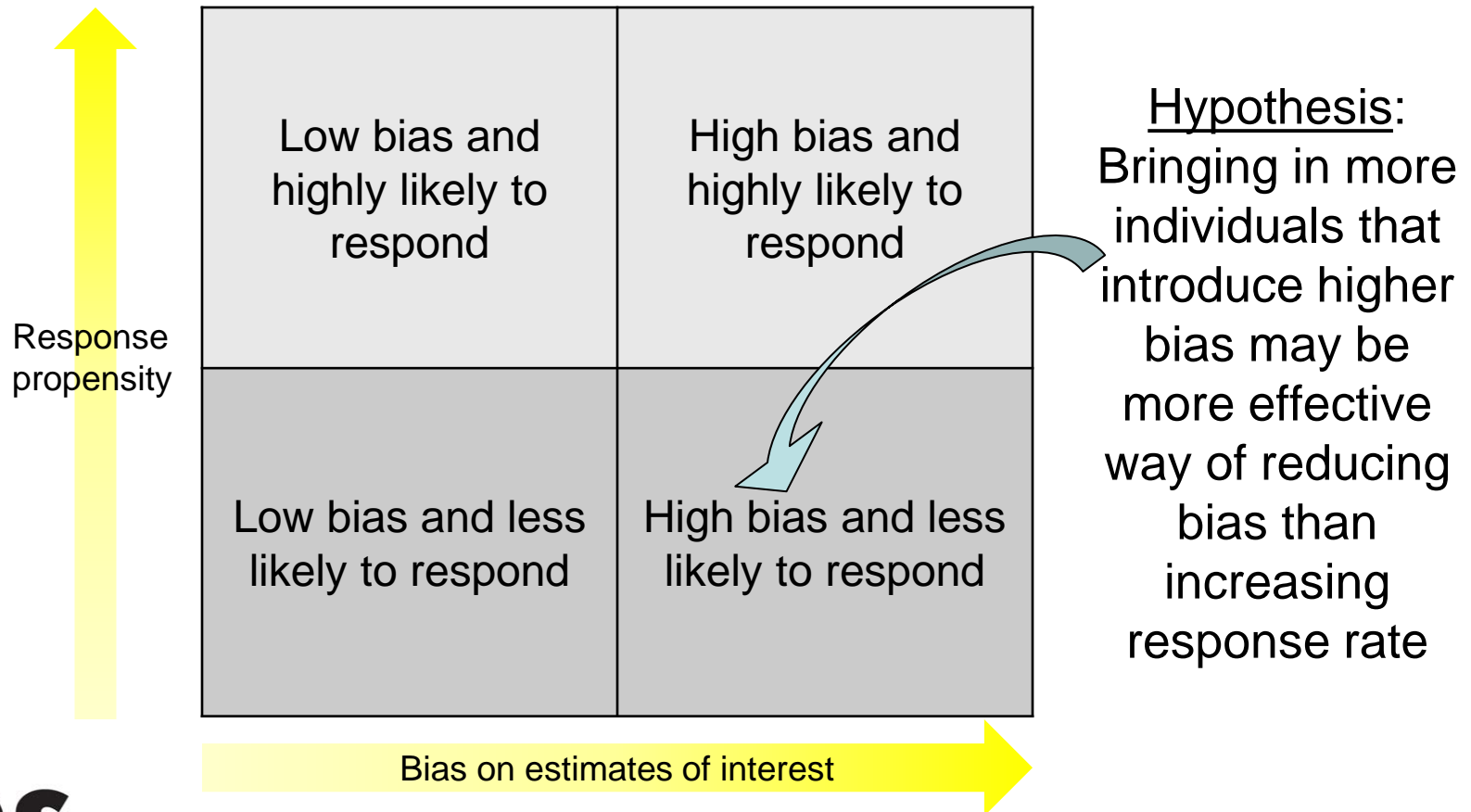
Institute of Education Sciences

Responsive Design Model Choice - Background

- The 2012 B&B is the second follow-up of graduates of the Class of '08
  - Base year: sample of NPSAS seniors in 2008
  - First follow-up: one year after graduation in 2009

- Relatively high prior response rate (~86%)

- Conducted "response propensity" experiment in 2011 field-test which showed that conversion of late phase nonrespondents didn't result in bias reduction

ies NATIONAL CENTER FOR
EDUCATION STATISTICS
Institute of Education Sciences

Responsive Design Model Choice - Background

- ELS:2002/12 is the third follow-up collection from 16,000+ high schoolers sophomores 2002, now aged ~26
  - Base year: Sophomores in 2002
  - First follow-up: Seniors in 2004 (freshened)
  - Second follow-up: 2 years post modal HS grad date in 2006
- Observable lower prior response rate for cases identified as ever dropping/stopping out of HS (83% vs. 89% overall)
- Typically this round is most difficult for secondary longitudinal studies
  - NELS:88/2000 round: ~78% (Dropouts: ~65%)

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

4

Problem: Nonresponse introduces bias but will targeting propensity to respond necessarily reduce *bias*?

Response propensity →

Low bias and highly likely to respond	High bias and highly likely to respond
Low bias and less likely to respond	High bias and less likely to respond

Bias on estimates of interest →

Hypothesis:
Bringing in more individuals that introduce higher bias may be more effective way of reducing bias than increasing response rate

ies NATIONAL CENTER FOR
EDUCATION STATISTICS
Institute of Education Sciences

Responsive Design Model Choice – Hypothesis

- Dissimilarity of each nonrespondent, relative to dissimilarity of mean respondent, can de determined
    - I.e., Distances between groups can be calculated

- Based on known variables for all sample members
    - E.g., frame variables, administrative record data collections, and prior response/imputations

- Hypothesis: Dissimilarity is related to an individual case's contribution to bias were the person not to respond

- Model choice: Variables chosen that are believed to be related to items of interest measured in upcoming collection

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

Responsive Design Model Choice - Simulations

- Through simulations using 2009 Baccalaureate and Beyond Longitudinal Study (B&B) data, we tested two techniques:

  - R-Indicator – This method calculates distances of *subgroups* of nonrespondents from the characteristics of the respondent group.

  - Mahalanobis Distancing (M) – This method calculates a separate distance for *each* nonrespondent from the characteristics of the mean respondent.

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

Responsive Design Model Choice - Simulations

The results of the simulations showed the following:

- R-Indicator: No significant reduction in nonresponse bias on 23 key variables
- Mahalanobis: Significant reduction nonresponse bias on five variables:

> -Cumulative total amount borrowed;
> -Cumulative amount owed;
> -Cumulative federal loan amount borrowed;
> -Parent's highest education; and
> -Graduate school enrollment.

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

Responsive Design Model Choice - Variables

- Both B&B and ELS models used:

  - <u>Substantive variables</u> – Variables we care about, can be demographical or sample member's status
    - Enrollment status
    - Age, race/ethnicity, gender of sample member

  - <u>Paradata variables</u> – Needed because Mahalanobis function incorporates ultimate response propensity measure
    - Response status from previous rounds
    - Number of contact attempts

- Cases selected for treatment received all following treatments until response occurs

NATIONAL CENTER FOR **EDUCATION STATISTICS**

Institute of Education Sciences

Targeting monetary and non-monetary incentives

B&B Treatment of High Distance Cases

1. First three months of data collection: no add'l intervention offered to either treatment or control groups

    » Primarily self-administered web instrument data collection period will cream-skim the low-cost-to-complete cases

2. Treatment 1 – End of Month 3 – Additional monetary incentive

3. Treatment 2 – End of Month 4 – Switch to prepaid incentive, FedEx

4. Treatment 3 – End of Month 5 – Early Abbreviated Interview

ies **NATIONAL CENTER** FOR **EDUCATION STATISTICS**

Institute of Education Sciences

Targeting monetary and non-monetary incentives

ELS: Treatment of High Distance Cases

1. Treatment 0 – Entire collection – Dropout cases receive additional monetary incentive

2. Treatment 1 – End of Month 1 – Additional monetary incentive, Intensive tracing

3. Treatment 2 – End of Month 2 – Cases selected for Field Collection

4. Treatment 3 – End of Month 4 – Switch to prepaid incentive, FedEx envelope

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

11

B&B Model variables

- Substantive
  - Race/ethnicity
  - Age
  - Parental education
  - Marital status
  - Dependency status
  - Immigrant generation
  - Disability status
  - Undergrad college type
  - Undergrad college region
  - Undergrad college selectivity
  - Undergrad major field of study

  - Expected family contribution
  - SAT/ACT scores
  - Earned income since graduation
  - Undergrad GPA
  - Employment status since graduating
  - Time to undergrad degree
  - Postgrad enrollment history

- Paradata
  - Base-year response status
  - First follow-up (FFU) response status
  - FFU call count
  - Second follow-up call count
  - Sample member "located" in second follow-up

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

ELS Model variables

- Substantive
  - Race/ethnicity
  - Age
  - Gender
  - High school control
  - High school urbanicity
  - SES in 2004
  - Parents' highest education
  - High school GPA
  - Educational attainment expectation in 12th grade

- Paradata
  - Response mode for panel maintenance (web or paper)
  - Ever responded to panel maintenance (this or earlier wave)
  - Previous wave response status
  - Number of calls in previous wave

- **Note:** More paradata built into model as data collection progressed (e.g., number of 2012 call attempts)

NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

13

Implications for monetary and non-monetary incentives

- **Treatment matters –** no point to a responsive design if no treatment

- Type of treatment should be carefully selected
  - Monetary incentives vs. non-monetary incentives
    - FedEx envelope successful with ELS population
    - Earlier intensive tracing successful in finding more cases
  - Some cheaper treatments may be more effective than expensive ones
  - Timing of treatment within data collection period

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

Thank you! Questions?

Ted Socha	Elise Christopher
B&B	ELS:2002
ted.socha@ed.gov	elise.christopher@ed.gov