Results from Different Adaptive Design Methods to Locate and Survey a Longitudinal Sample

Karen Grigorian, Jill Connelly, Tom Hoffer, and Michael Yang

April 11, 2017

Presentation Overview

Adaptive design methods used to prioritize a longitudinal sample to mitigate nonresponse bias

Reporting on methodological experiences from:

- High School and Beyond (HS&B) Study

  - Sophomore Cohort Fifth Follow-up 2014

  - Senior Cohort Fourth Follow-up 2015

Adoption of Adaptive Design in Survey Research

- Increasing use of Adaptive Design methods in survey research
  - Minimize nonresponse bias
  - Maximize representativeness of key domains
  - Maximize limited budgets
- An adaptive approach to survey data collection is one that uses information available prior to and during data collection to adjust the collection strategy for the remaining cases.
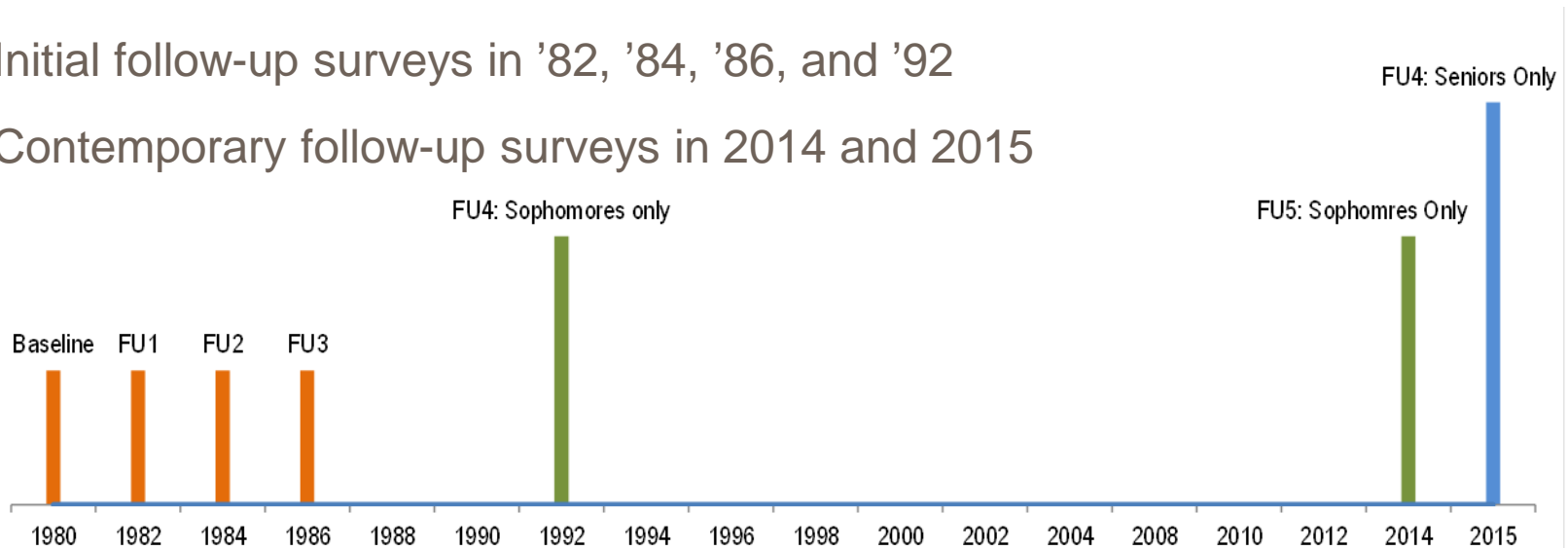
Traditional Adaptive Design Strategy

- Conduct Data Collection in phases

- Starting Phase: Treatment is consistent across all cases

- Response Assessment:
  - Statistical comparison of respondents to nonrespondents across key analysis domains
  - Assess bias and response propensity
  - Prioritize cases for differential treatment in the follow-up phase

- Follow-up Phase: Apply differential treatment(s) to mitigate bias to increase response from underrepresented cases

- Repeat the response assessment and follow-up phase as planned

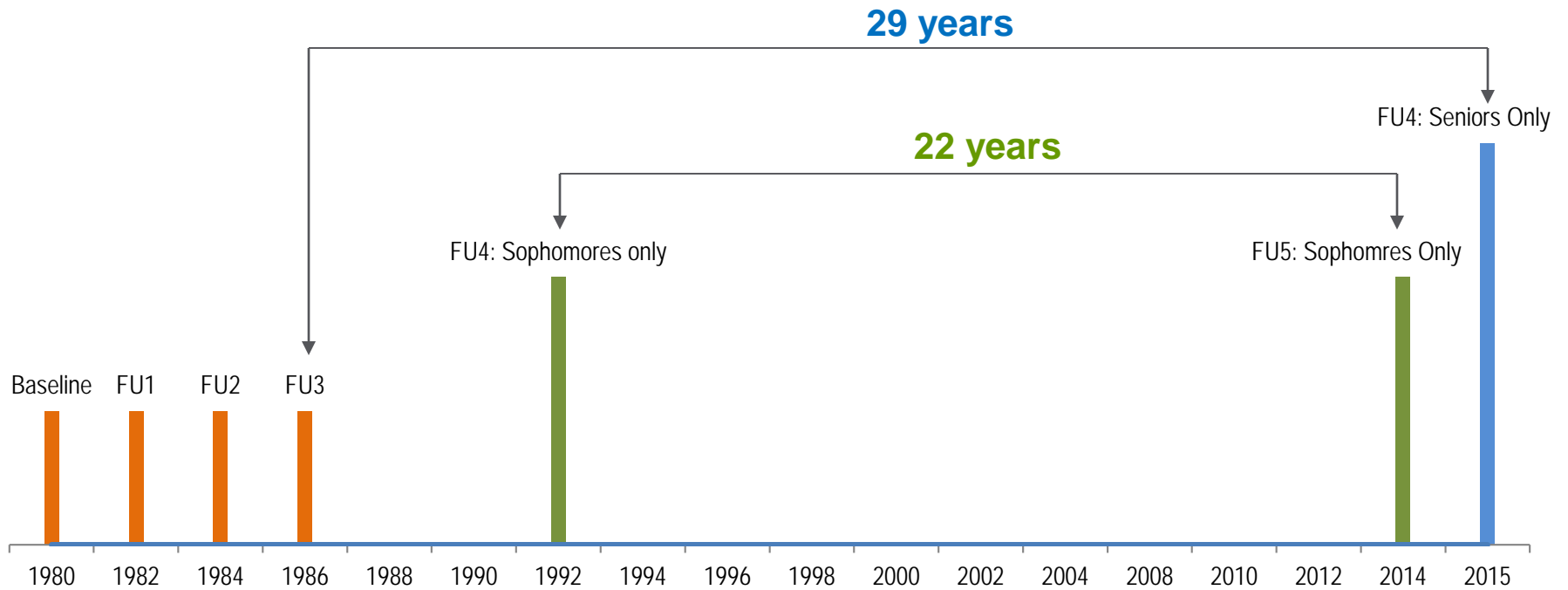# Adaptive Design Methods for Longitudinal Studies

- Longitudinal studies provide an opportunity to accelerate the adaptive design strategy

- Response data from past survey cycles can be used to predict future behavior

- Adaptive design strategies can be implemented *before* data collection begins – before there are respondents and nonrespondents for comparison

- To apply differential starting phase treatments, statistically assess the sample along the following dimensions

  a. Predicted locating difficulty

  b. Predicted cooperativeness

  c. Analytic domain size and/or base weight differences

Methods & Experiences from New HS&B Follow-Ups

- High School and Beyond Study started in 1980

- U.S. nationally-representative sample of 1,015 public and private high schools

- Students sampled from schools

    - Sophomores: ~16 years old

    - Seniors: ~18 years old

- Initial follow-up surveys in '82, '84, '86, and '92

- Contemporary follow-up surveys in 2014 and 2015

29 years

22 years

FU4: Seniors Only

FU4: Sophomores only

FU5: Sophomres Only

Baseline   FU1   FU2   FU3

1980   1982   1984   1986   1988   1990   1992   1994   1996   1998   2000   2002   2004   2008   2010   2012   2014   2015

2014 Data Collection Background

Sophomore Cohort: n = 14,825, N = 3,781,000

- 22 year hiatus between FU4 in 1992 and FU5 in 2014

- Available information on the sophomore cohort for 2014 follow-up:

  - Contacting information collected during FU3 in 1986

  - Sample member name, date of birth, and social security numbers
    - Sample member characteristics (e.g., race, gender, sex, education level from survey data, citizenship at birth, HS achievement)
    - 1986 address and phone number
    - Contacts listed at the end of the 1986 survey
    - Parent names and address
    - School information

Sophomore Cohort Data Collection Overview

Timeframe	Phase	Description
July- Sept. 2013	Prefield locating	• Locating of cases with incomplete or unreliable contact information
Sept. -Dec. 2013	Initial Protocol	• 40-minute telephone survey • Full sample • Refusal conversion effort
Jan. – June 2014	Nonresponse Follow-up	• 40-minute telephone survey and 10-minute critical-items-only survey • Multimode approach introduced (telephone, paper SAQ, and later web) • Effort focused on <u>half</u> of the sample due to cost constraints
Aug. – Oct. 2014	Late-stage	• Additional funding received • **Adaptive design approach introduced** • Multimode data collection continues • Full sample

NORC 75
at the UNIVERSITY of CHICAGO

Data Collection Dilemma

- Pursuing additional nonresponse follow-up would:
  - Increase the number of completed cases
  - Increase the overall survey response rate
  - Decrease the confidence interval widths for the total population estimates
- However, pursuing additional nonresponse follow-up <u>might</u>:
  - Lead to increased disparities in sample representation
  - Not reduce nonresponse bias
  - Not reduce confidence intervals for population subgroups
- Here's what we decided to do…

NORC 75
at the UNIVERSITY of CHICAGO

Adaptive Design Strategy Introduced

- Domains identified that would positively impact confidence interval widths and sample representation

    1. Non-Whites;

    2. Persons in the two lowest HS achievement quintiles;

    3. Persons not reporting a degree/certificate post-high school;

    4. Persons selected from schools in urban areas.

- Prioritize the sample based on membership in the domains:

    - Priority Group 1 (Highest Priority cases): Cases belonging to 3 or more domains

    - Priority Group 2: Cases belonging to 2 domains and Male

    - Priority Group 3: Cases belonging to 2 domains and Female

    - Priority Group 4 (Lowest Priority Cases): All other cases

Adaptive Design Treatment Results

Weighted Response Rate Improved in Late-Stage Phase: HS&B Sophomores

Characteristic	Category	Weighted Response Improvement
Race	Black	12.4
	Hispanic	12.9
	White	11.0
	Other	8.7
Sex	Male	12.5
	Female	10.1
High School Achievement Quintile	1	12.4
	2	11.8
	3	11.7
	4	10.1
	5	9.3
Post-secondary Education Attainment	Unknown/missing	14.5
	None	12.5
	Certificate	17.5
	Associates	11.1
	Bachelors	6.9
	Graduate/Professional	10.3
Urbanization	Rural	12.0
	Suburban	11.2
	Urban	10.6

NORC 75
at the UNIVERSITY of CHICAGO

Data Collection Background

Senior Cohort: n=11,995, N = 3,040,000

- 29 year hiatus between FU3 in 1986 and FU4 in 2015

- Available information on the senior cohort for 2015 follow-up

  - Survey records coming out of the 1984 follow-up
    - Sample member name, date of birth, and social security numbers
    - Sample member characteristics (race, gender, education level from survey data)
    - Contacts listed at the end of the 1986 survey
    - Parent names and address
    - School information

  - 1986 hardcopy records – face sheets from field work and future contacting page from the survey form for respondents

Adaptive Design Methods from the Start

Sample prioritized for data collection activities at 3 time points:

1. Starting Phase: Prioritized full sample for differential treatments in prefield locating and the initial data collection protocol.

2. Follow-up Phase: Reprioritize nonresponse cases before the follow-up data collection phase which included differential incentives.

3. Late-Stage Phase: Reprioritize remaining nonresponse cases before implementing the final gaining cooperation contacting strategy.

NORC 75
at the UNIVERSITY of CHICAGO

Starting Priority Score Calculation Method

- For the Starting Phase, the priority score was based on propensity to require locating

- To estimate propensity, restricted the Sophomore cohort to High School Graduates and estimated likelihood of being a locating problem with logistic regression model

- Apply the Sophomore logistic results to the Senior sample

- Aggregate into 3 different priority groups for differential locating treatments and the order in which cases were worked

Starting Priority Score Logistic Regression Results

Likelihood of Needing Locating for Subgroups

Variable	Category	% Likely to Require Locating	Direction
Gender	Males	12%	More
Race/Ethnicity	Hispanic	14%	More
	Blacks	23%	More
	Asian, Native American, Othe	42%	More
Education	Associates Degree	20%	*Less*
	BA/BS	33%	*Less*
	MS/PhD	23%	*Less*
Achievement Level	Lowest Scores (1st quntile)	106%	More
	2nd Lowest Scores	60%	More
	3rd Lowest Scores	38%	More
Parent's Native Born Status	Missing	35%	More
Census Division	South Atlantic (Region 5)	41%	More

NORC 75
at the UNIVERSITY of CHICAGO

Follow-up and Late-Stage Priority Score

- The priority score is based on a multivariate comparison of respondents to nonrespondents

- Specifically, the Mahalanobis distance of each nonrespondent from the weighted mean value of the respondents was calculated

- The Mahalanobis distance is based on data from the sampling frame (e.g., race, gender, region, etc.) and past survey data available for both respondents and nonrespondents

- Nonrespondents that are more distant from the respondents are assigned a higher priority score as they would contribute more to bias reduction if they become respondents

Senior Cohort Data Collection Overview

Timeframe	Phase	Description
Jan. – March 2015	Prefield locating	• Begin with adaptive design approach • Locating of cases with incomplete or unreliable contact information • Cases worked in priority order
April – June 2015	Starting Protocol	• 10-minute survey • Multimode approach (phone, web, & paper SAQ)
June – early Aug. 2015	Follow-up	• Sample evaluated, new priority assignments made • Multimode data collection continues • Cases worked in priority order • Differential incentives introduced
Aug. – Oct. 2015	Late-stage	• Sample evaluated, new priority assignments made • Multimode data collection continues • Cases worked in priority order • Differential incentives updated

NORC75
at the UNIVERSITY of CHICAGO

Preliminary Comparison Results

**Response Rate Differences Between Subgroups
for Key Characteristics: HS&B Seniors**

Subgroup Comparison	Sophomore Final Weighted RR	Senior Penultimate Weighted RR
Racial Ethnic Group: Hispanic to White	14.6%	7.7%
Racial Ethnic Group: Black to White	14.0%	12.9%
Sex: Male to Female	5.2%	6.6%
Educational Attainment: HS or less to Bachelors	17.6%	16.8%

NORC 75
at the UNIVERSITY *of* CHICAGO

Future Research

- Evaluate the final weighted Senior nonresponse results

- Create R-indicators for Sophomore FU5 and Senior FU4 sample data at key time points to evaluate the prioritization and adaptive treatments

- Evaluate ways to improve a Starting Priority Score for a panel sample incorporating propensity to respond and domain size in addition to the likely need for locating

NORC 75
at the UNIVERSITY of CHICAGO

References

Coffey, A., Reist, B., and White, M. (2013). Monitoring Methods for Adaptive Design in the National Survey of College Graduates. Presentation at the August 7, 2013 Joint Statistical Meeting.

Curtin, T. R., Ingels, S. J., Wu, S. and Heuer, R. (2002). User's Manual: NELS:88 Base-Year to Fourth Follow-up: Student Component Data File. NCES 2002323.

Gambino, J., Laflamme, F., and Wrighte, D. (2010) . Responsive Design at Statistics Canada: Development, Implementation and Early Results. This paper was prepared for the 21st International Workshop on Household Survey Nonresponse, Nürnberg, Germany, 30 August – 1 September, 2010.

Schouten, B., Calinescu, M., and Luiten, A. (2013) Optimizing Quality of Response through Adaptive Survey Designs. *Component of Statistics Canada Catalogue,* no. 12-001-X, Business Survey Methods Division. June 2013.