

Adaptive Design in the National Immunization Survey-Teen Provider Record Check Phase

Xian Tao¹, Megha Ravanam¹, Benjamin Skalland¹, Kirk Wolter¹, David Yankey², Zhen Zhao²

¹NORC at the University of Chicago

²National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention

Abstract¹

Adaptive design principles are applied to the National Immunization Survey-Teen (NIS-Teen), sponsored by Centers for Disease Control and Prevention, which monitors vaccination coverage of U.S. adolescents age 13-17 years. Data collection is ongoing in two phases: (1) a random-digit-dial telephone survey to interview parents/guardians with age-eligible adolescents, followed by (2) a mail survey to vaccination providers, called the provider record check (PRC), to obtain vaccination histories for the adolescents. A logistic regression model relating the probability that an Immunization History Questionnaire (IHQ) is returned for a teen-provider pair to characteristics of the adolescent, mother, household, and providers was fit. R-indicators and partial R-indicators for the PRC phase of the 2015 NIS-Teen are presented to evaluate the representativeness of response in the PRC. The indicators are visualized using interactive graphics embodied in an R Shiny application to track the real time changes. Programmatic interventions to improve representativeness are discussed, which include strategies for prompting providers and special treatment of certain subgroups.

Keywords: Adaptive Design, National Immunization Survey-Teen, R-indicators, Partial R-indicators

1. Introduction

With the decline in survey response rates (Miller, 2017) and the lack of a strong connection between the response rate and nonresponse bias (Groves, 2006), researchers are increasingly focusing on measures of representative response as alternatives to response rates when assessing survey quality. Adaptive/responsive (A/R) design is a method of sample management in which survey managers analyze the state of response in real time and use the results to adapt, target, and tailor their data collection strategies, both in the balance of the current round of the survey and in future rounds.

A/R design is related to the concept of representative response, which assesses the degree to which the respondents to the survey resemble the whole selected sample; response is representative if (1) the response propensities (probabilities) are the same for all units in

¹ The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Immunization and Respiratory Diseases or NORC at the University of Chicago.

the population, or (2) the variation in the response propensities within each category of a categorical variable is small.

A/R design brings with it a paradigm shift towards somewhat less emphasis on strategies to increase the response rate and somewhat greater emphasis on strategies to increase the representativeness of response or to interview the “right” units.

There are several managerial interventions used in A/R design to increase the representativeness of response, including additional call-backs to units in specific subgroups; providing late-stage incentives; preparing refusal converters; switching the mode of data collection; and using alternative contact strategies. The response rate and its components can be calculated in real time and are easy to interpret, but they alone may not provide insight into where to focus additional data-collection efforts. New indicators may be needed to support decisions about targeting and tailoring call-back efforts. The new indicators should have the capacity to be computed in real time, should be easily interpretable, and should be able to lead to effective managerial interventions. R-indicators were first introduced by Schouten, Cobben, and Bethlehem (2009) to assess the similarity between the response and the sample of a survey in order to serve as tools for monitoring and comparing survey quality.

In this paper, A/R design is applied to the National Immunization Survey-Teen (NIS-Teen), sponsored by the Centers for Disease Control and Prevention, which monitors vaccination coverage of U.S. adolescents age 13-17 years. Data collection is ongoing in two phases: (1) a random-digit-dial telephone survey to interview parents/guardians with age-eligible adolescents, followed by (2) a mail survey to vaccination providers, called the provider record check (PRC), to obtain vaccination histories for the adolescents. For the household phase, due to a lack of good auxiliary variables to model response propensities, R-indicators have not yet been operationalized. This paper will focus the application of R-indicators to the PRC phase. The PRC phase includes a richer set of covariates for estimating response propensities. Covariates can include information collected during the household phase, known characteristics of the provider, and paradata collected to date. A logistic regression model relating the probability that an Immunization History Questionnaire (IHQ) is returned for a teen-provider pair to characteristics of the adolescent, mother, household, and providers was fit. R-indicators and partial R-indicators are presented for the PRC phase of the 2015 NIS-Teen to evaluate the representativeness of response in the PRC. The indicators are visualized using interactive graphics embodied in an R Shiny application to track real-time changes. Potential programmatic interventions to improve representativeness are also discussed, including strategies for prompting providers and special treatment of certain subgroups.

2. Methods

Let $i = 1, 2, 3, \dots, N$ to indicate the units in the population, and let s be a selected sample from the population. X is a vector of known covariates such as geographic, demographic, and other characteristics that explain the survey’s response mechanism. $i \in s$ is a unit in the selected sample, and let ρ_i be the response propensity for unit i . Let x_i be the unit’s value of X , and let W_i be the sampling weight for the unit. The formula for the overall R-indicator (Schouten et al. 2009) is

$$R = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i \in S} W_i \left(\rho(x_i) - \frac{1}{N}\sum_{j \in S} W_j \rho(x_j)\right)^2}.$$

R is a measure of the balance of the predicted response propensities over the sampled units. R takes values between 0 and 1. Larger values signify greater representativeness of response; smaller values signify greater departures from representativeness. As the data collection period progresses, one hopes to find rising values of the R-indicator. In comparing one round of a survey to previous rounds, one hopes to find comparable or rising values of the R-indicator.

The original formula for the partial R-indicator (Schouten and Shlomo, 2015) is

$$P_u(Z) = \sqrt{\sum_{k=1}^K \frac{\bar{N}_k}{N} (\bar{\rho}_k - \bar{\rho})^2}.$$

P_u is a measure of the balance of the predicted response propensities across categories $k = 1, 2, 3, \dots, K$ of a particular categorical variable, Z , e.g., race/ethnicity in $K=3$ categories. P_u takes values between 0 and 0.5. Note that, unlike the overall R-indicator, small values of P_u are good. To make the partial R-indicator more comparable with the overall R-indicator, we rescale it as $P_u^* = 1 - 2P_u$. In this way, the partial R-indicator, P_u^* , has the same scale as the R-indicator, which is between 0 and 1. The larger the values of P_u^* are, the more balanced the sample is. Small values of P_u^* suggests that managerial intervention is needed.

3. Results

3.1 2015 NIS-Teen

A logistic regression model was fit using the 2015 NIS-Teen teen-provider pairs that were mailed an IHQ, relating information collected during the household interview to the IHQ return status. Covariates used in the model are shown in Table 1 and include socio-demographic characteristics of the teen, the teen's mother, and the household, as well as the type of provider that was nominated, as classified based on the provider name. The model is to predict whether a teen-provider pair will result in a returned IHQ. There were a total of 39,024 teen-provider pairs mailed an IHQ in 2015, and a total of 36,520 teen-provider pairs returned an IHQ, for an IHQ return rate of 93.6%. The distribution of the predicted response propensities for these teen-provider pairs is shown in Figure 1. Most of the predicted response propensities cluster in the range of 92% to 98%. The distribution has a longer left tail, indicating that a small number of teen-provider pairs have a lower predicted response propensity.

The overall R-indicator based on the predicted response propensities is 94.3%, which is quite high, indicating that the responding sample is very representative. Table 2 shows the partial R-indicators for different categorical variables (Metropolitan Statistical Area (MSA) status, provider type, poverty status, race/ethnicity of the teen, past doctor visits, mother's age group, housing tenure, number of children in the household, mother's marital status, mother's education, mobility status, relationship of the respondent to the teen, age of the teen, sample type, 11-12 year old check-up status, and sex of the teen). Overall, the partial R-indicators are quite high. The partial R-indicators are lowest for the provider type and

race/ethnicity of the teen; that is, among the mailed IHQs, there is more variation in the predicted response propensities between categories of these variables than for any of the other variables examined. Tables 3.1 and 3.2 show the IHQ return rates for each category of these two variables. The IHQ return rate was lower for pharmacies, hospitals, and schools; it was also lower for non-Hispanic Black teens, non-Hispanic Asian teens, and Hispanic teens.

Table 1: Covariates in Logistic Regression Model

Covariate	Definition
Age category of teen	Age category of teen (13-14 years, 15-17 years)
Household report of check-up at 11-12 years of age	Household report of check-up at 11-12 years of age (yes, no, don't know/missing)
Household reported past doctor visits	Household reported past doctor visits (none, 1, 2-3, 4+, don't know/missing)
Housing tenure	Housing tenure (owner, renter/other/don't know/missing)
Marital status of the mother	Marital status of the mother (widowed/divorced/separated/deceased, never married, married)
Maternal age group	Maternal age group (<=34 years, 35-44 years, 45+ years)
Mobility status	Mobility status (moved, did not move from different state)
Mother's education	Mother's education (<12 years, 12 years, > 12 years non-college graduate, college graduate)
MSA status	MSA status (MSA-central city, other MSA, non-MSA)
Number of children under 18 living in the household	Number of children under 18 living in the household (1, 2-3, 4+)
Poverty status	Poverty status (above poverty >75K, above poverty <=75K, below poverty, unknown)
Provider type	11 Levels (Pharmacy, Hospital, School, Medical Center, Large Healthcare Org, Military/Other, Missing, Clinic, Private Practice, Health Firm, Public)
Race/ethnicity of teen	Race/ethnicity of teen (Hispanic, White alone non-Hispanic, Black alone non-Hispanic, American Indian alone non-Hispanic, Asian alone non-Hispanic, Other/multi-racial non-Hispanic)
Relationship of household respondent to teen	Relationship of household respondent to teen (mother, father, other)
Sample type	Sample type (landline, cell-phone)
Sex of teen	Sex of teen (male, female)

Figure1: Estimated Response Propensity for Teen-Provider Pairs:
NIS-Teen, 2015

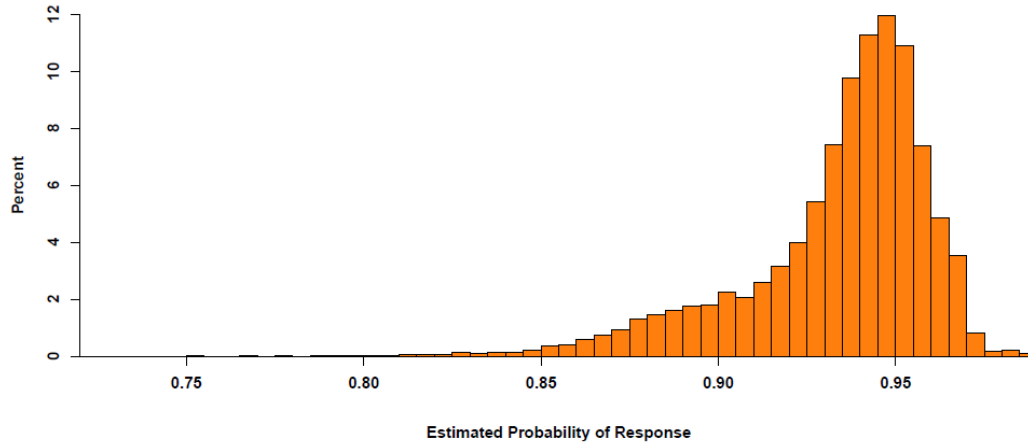


Table 2: Unconditional Partial R-Indicators, NIS-Teen, 2015

Definition	Partial R Indicator
Provider type	95.4%
Race/ethnicity of teen	98.1%
Poverty status	98.1%
MSA status	98.2%
Number of valid, unique providers identified by respondent	98.8%
Maternal age group	99.0%
Household reported past doctor visits	99.1%
Number of children under 18 living in the household	99.2%
Housing tenure	99.2%
Mother's education	99.3%
Marital status of the mother	99.5%
Mobility status	99.8%
Relationship of household respondent to teen	99.8%
Sex of teen	99.9%
Household report of check-up at 11-12 years of age	99.9%
Sample type	99.9%
Age category of teen	100.0%

Table 3.1: IHQ Return Rates by Provider Type: NIS-Teen, 2015

Provider Type	Number of Teen-Provider Pairs	IHQ Return Rate
Overall	39,024	93.6%
Pharmacy	575	89.9%
Hospital	2,320	90.7%
School	807	90.7%
Medical center	3,149	91.0%
Large healthcare organization	311	92.0%
Military/other	3,474	92.7%
Missing	2,402	93.3%
Clinic	4,477	94.1%
Private practice	12,767	94.5%
Health firm	8,521	94.6%
Public	221	96.8%

Table 3.2: IHQ Return Rates by Race/Ethnicity: NIS-Teen, 2015

Definition	Number of Teen-Provider Pairs	IHQ Return Rate
Overall	39,024	93.6%
Non-Hispanic Black alone	3,814	91.7%
Non-Hispanic Asian alone	1,377	92.8%
Hispanic	8,630	92.9%
Non-Hispanic American Indian alone	548	93.8%
Non-Hispanic other/multiple races	1,993	94.0%
Non-Hispanic White alone	22,662	94.2%

3.2 Real-Time Monitoring of Q3/2016 NIS-Teen

Since Q3/2016, the R-indicator and partial R-indicators for the NIS-Teen provider record check have been computed and tracked in real-time. As data collection progresses, more and more completed household interviews are achieved, and IHQs for the teen-provider pairs are mailed for the newly completed household interviews on a weekly basis. The basic idea is to fit a logistic regression model each week based on the teen-provider pairs with IHQs mailed as of that calendar date to produce a predicted response propensity for each teen-provider pair. Covariates collected during the household phase were used in the model, including socio-demographic characteristics of the teen, the teen's mother, and the household, as well as the type of provider that was nominated, as classified based on the provider name; these covariates are the same as those used in the 2015 model (Table 1). As of November 11, 2016, there were a total of 9,027 teen-provider pairs that had been mailed an IHQ and a total of 4,738 teen-provider pairs for which an IHQ had been returned, for an IHQ return rate of 52.5%.

Table 4 shows the cumulative number of mailed IHQs, the cumulative number of returned IHQs, the cumulative IHQ return rate, and the R-indicator by week. For example, by August 26, 2016, there were total of 1,545 teen-provider pairs with a mailed IHQ and 212

with an IHQ returned, for an IHQ return rate of 13.7%; the R-indicator as of that date is 83.6%.

An R Shiny Dashboard was created to visualize the real-time change in the R-indicator, partial R-indicators, and IHQ return rates. Figure 2.1 and Figure 2.2 in the Appendix are two examples from the R-Shiny Dashboard. The Dashboard can be used to chart trends in partial R-indicators and IHQ return rates for user-specified socio-demographic variables.

Table 4: IHQ Return and R-Indicator by Day, NIS-Teen, Q3/2016

Date	Cumulative Number of Returned IHQ	Cumulative Number of Mailed IHQ	Cumulative IHQ Return Rate	R Indicator
8/26/2016	212	1,545	13.7%	83.6%
9/2/2016	443	2,216	20.0%	84.1%
9/9/2016	715	2,851	25.1%	81.7%
9/16/2016	1,084	3,428	31.6%	80.5%
9/23/2016	1,386	4,192	33.1%	83.1%
9/30/2016	1,861	5,012	37.1%	80.0%
10/7/2016	2,328	5,587	41.7%	80.6%
10/14/2016	2,703	6,254	43.2%	82.3%
10/21/2016	3,210	7,026	45.7%	82.3%
10/28/2016	3,758	7,757	48.4%	80.9%
11/4/2016	4,316	8,496	50.8%	82.5%
11/11/2016	4,738	9,027	52.5%	82.4%

4. Discussion and Limitations

While R-indicators are now being produced and tracked for the PRC-phase, R-indicators for the household phase have not yet been operationalized because the household phase lacks good auxiliary variables for estimating the response propensities. For the PRC phase, potential managerial interventions could be taken for certain provider types and for teens with certain race/ethnicities. Examples of such changes may include inserting additional materials in the PRC mail packet to encourage participation and reiterate the importance of the study, offering to provide special handling practices for those providers (i.e., only mailing requests once a month or only calling on a specific day/time), providing monetary or non-monetary incentives to providers if protocol allows, and querying state Immunization Information Systems (IISs) in lieu of obtaining the data from providers. Interventions such as special handling practices and additional encouragement from CDC groups are actively being formulated and will be tested in future quarters.

References

- Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys, vol. 70, no. 5, special issue, 2006, pp. 646–675.
- Miller, P. (2017). Is there a future for surveys? *Public Opinion Quarterly*, vol. 81, special issue, 2017, pp. 205-212.

Schouten, B., Cobben, F., and Bethlehem, J. (2009). Indicators for the Representativeness of Survey Response. *Survey Methodology*, vol. 35, pp. 101–113.

Schouten, B., Sholomo, N. (2015). Selecting adaptive survey design strata with partial R-indicators. *International Statistical Review*, vol. 85, issue 1, 2017, pp. 143-163.

Appendix

Figure 2.1: Example of R-Shiny Dashboard for Visualization

IHQ Return Tab1 Tab2

IHQ Return Related Plots

Date range:
 to

Select Covariates to Show Partial R-Indicators in Figure1:

- Poverty status
- Race/ethnicity of teen
- Provider type
- MSA3 status
- Maternal age group
- Household reported past doctor visits
- Number of children under 18 living in the household
- Housing tenure
- Mother's education
- Marital status of the mother
- Mobility status
- Relationship of household respondent to teen
- Age category of teen
- Household report of check-up at 11-12 years of age
- Sex of teen
- Sample type

Select Covariate to Show IHQ Return Rates in Figure2:

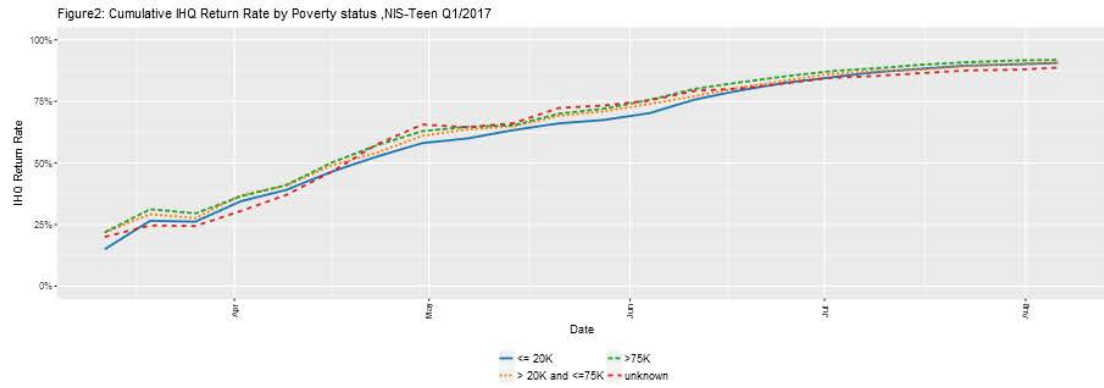
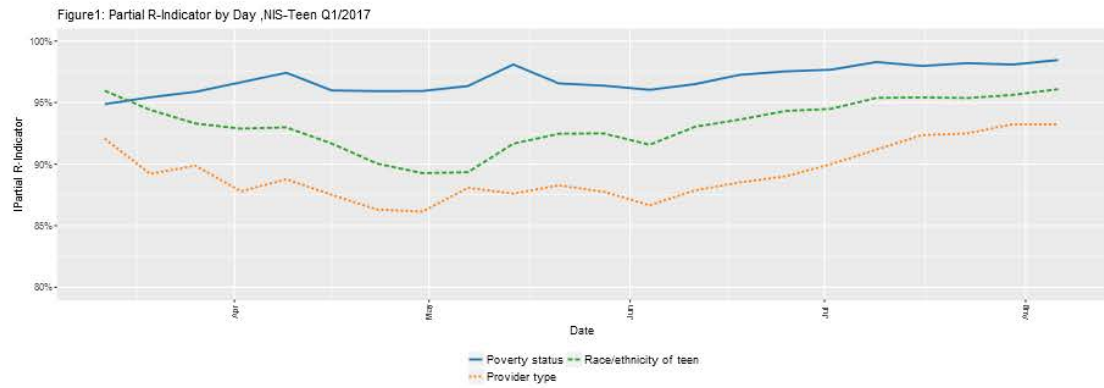


Figure 2.2: Example of R-Shiny Dashboard for Visualization

