

Request for Approval of a Nonsubstantive Change:

**National Health Interview Survey (NHIS)
ADAPTIVE DESIGN EXPERIMENT**

OMB No. 0920-0214
(Expiration Date 12/31/2017)

Contact Information:

Marcie Cynamon
Director, Division of Health Interview Statistics
National Center for Health Statistics/CDC
3311 Toledo Road
Hyattsville, MD 20782
301.458.4174 (voice)
301.458.4035 (fax)
mlc6@cdc.gov

May 23, 2016

NATIONAL HEALTH INTERVIEW SURVEY (NHIS)
ADAPTIVE DESIGN EXPERIMENT

A1. Circumstances Making the Collection of Information Necessary

This request is for a nonsubstantive change to an approved data collection (OMB No. 0920-0214; exp. 12/31/2017), the National Health Interview Survey (NHIS). With this nonsubstantive change request, the Division of Health Interview Statistics seeks clearance to test the use of adaptive design in the NHIS. The aim is to investigate the impact of adaptive case prioritization on sample representativeness and nonresponse bias, while maintaining survey costs and minimizing any possible negative effect on the overall response rate.

The National Health Interview Survey (NHIS), like many other large national sample surveys, has experienced a steady decline in response rates over the past 15 to 20 years (see Figure 1). Low participation rates in surveys matter to the extent that they introduce the potential for nonresponse bias in survey estimates. Nonresponse bias assessments of 2013 and 2014 NHIS Early Release (ER) Program estimates have been performed, with the results suggestive that bias may be present in some key health estimates. In this section, we more fully describe the methods behind and the results from these assessments.

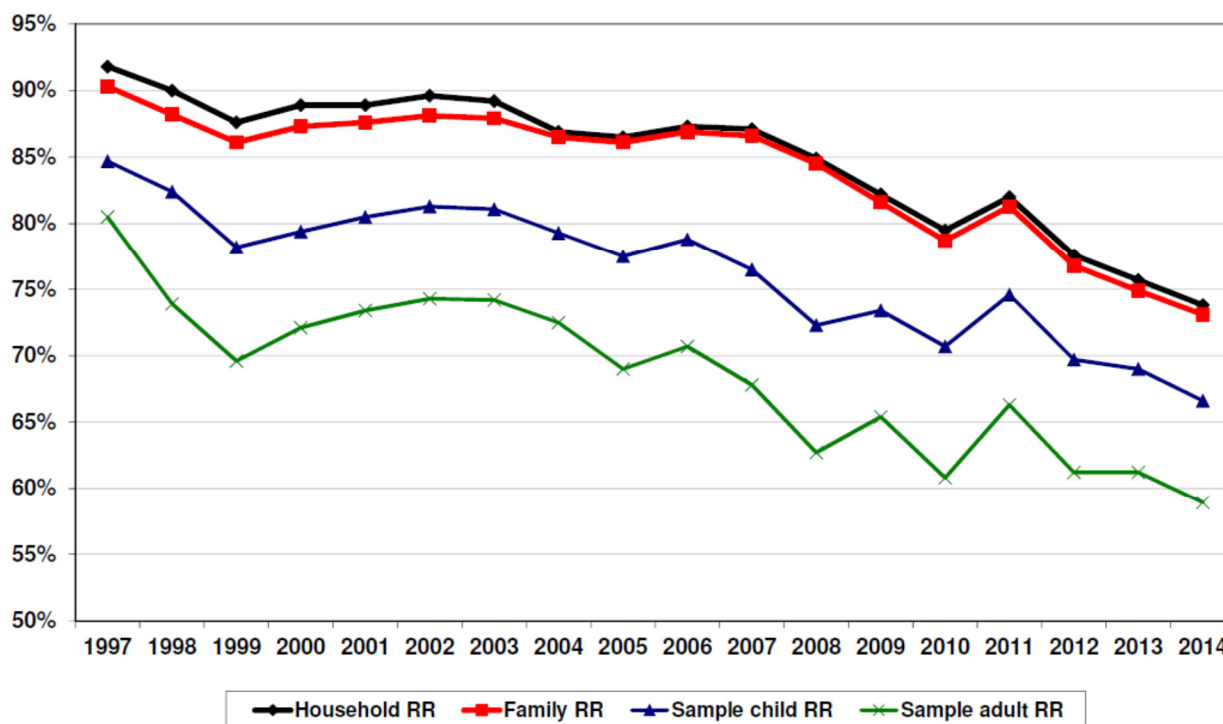


Figure 1. Final Household, Family, Sample Child, and Sample Adult Response Rates: NHIS, 1997-2014

To assess if nonresponse bias may be present in ER measures, a logistic regression of family response that included several covariates culled from the NOI and CPD was first estimated. The covariates in the model were those found to be associated with several health outcomes on the NHIS. Examples include the interviewer assessment of whether or not one or more residents may be disabled, handicapped, or have a chronic health condition; interviewer assessment of whether or not the sample unit may include smokers; average aggregated household income for the block group or tract in which the household is

located (ACS 2009-13); and the proportion of persons aged 25 and over with a college degree in the block group or tract in which the household is located (ACS 2009-13). The two-level logistic regression model included random interviewer effects and was estimated twice, once on 2013 data and once on 2014 data. The predicted values of response propensities from each run were used to group responding families into response propensity quintiles. We then examined estimates for 20 ER measures in two ways. First, we observed the health estimates for each of the response propensity quintiles. And second, we observed the estimates cumulatively moving from the high response propensity quintile to the low response propensity quintile. Note that the quintile-specific and cumulative estimates were weighted using base weights (inverse of the probability of selection). Comparisons of estimates by quintile, as well as systematic changes in the cumulative estimate as families with lower response propensities are added to the sample, provide clues as to possible nonresponse bias with these key health measures. The approach is akin to level-of-effort (LOE) analyses [1], where change in a statistic over increased levels of effort, or in this case over response propensity quintiles, is indicative of the risk of nonresponse bias. Conversely, little to no change in the statistic is suggestive of the absence of nonresponse bias [2].

Table 1 presents person-level ER variables broken out by year. Starting with the percentage of persons of all ages who had excellent or very good health, we can consider the persons in quintile 1 (high response propensities) to be from families that were the easiest to recruit. In 2013, the estimate for this group was 57.5%. This is roughly 16 percentage points lower than the estimate for quintile 5 (low response propensities), or persons from families that were the least likely to participate. Table 1 also presents the cumulative estimate moving from the high response propensity quintile to the low response propensity quintile (left to right). As we recruit families with lower and lower response propensities, the estimate of persons who had excellent or very good health increases from 57.5% to 65.9%. Furthermore, we can treat the persons from the low response propensity families as proxies for nonresponders, and then compare the estimate for this group (73.2%) to the estimate for the remainder of the sample (64.4%). The difference between the two estimates is statistically significant (two-tailed t-test conducted at the .05 level). Together, the quintile-specific and cumulative estimates suggest that persons from nonresponding families may have a higher rate of reporting excellent or very good health. Therefore, we may be underestimating the percentage of persons who have excellent or very good health. The pattern identified for 2013 appears to hold for 2014.

For the three remaining person-level estimates, evidence of possible nonresponse bias is observed for both years and the patterns are consistent. Generally speaking, we may be overestimating the percentage of persons who failed to obtain need medical care due to cost in the past 12 months and the percentage of persons who need help with personal care needs, while underestimating the percentage of persons with health insurance coverage. The final column in Table 1 presents the final weighted estimate for each measure. With the exception of health insurance coverage, applying the final person weights appears to move the estimates in the anticipated direction, although the movements are small and may be insufficient. For health insurance coverage, the final weighted estimate moved slightly in the opposite direction from what the bias analysis suggests.

Table 1. Nonresponse Bias Analysis of Person-Level Early Release Estimates: NHIS, 2013-2014^a

	Quintile 1: high response propensities	Quintile 2	Quintile 3	Quintile 4	Quintile 5: low response propensities	Quintile 5 vs. Top 4 Quintiles (t-test)	Final Weighted Estimate
2013							
Percentage of persons of all ages who had excellent or very good health							
Quintile estimate (s.e.)	57.5 (0.63)	64.1 (0.63)	67.9 (0.59)	69.3 (0.63)	73.2 (0.62)	*	66.3 (0.29)
Cumulative estimate (s.e.)	57.5 (0.63)	60.7 (0.44)	63.0 (0.37)	64.4 (0.32)	65.9 (0.30)		
Percentage of persons of all ages who failed to obtain needed medical care due to cost in the past 12 months							
Quintile estimate (s.e.)	7.6 (0.23)	6.5 (0.24)	5.1 (0.20)	5.1 (0.23)	4.9 (0.29)	*	5.9 (0.11)
Cumulative estimate (s.e.)	7.6 (0.23)	7.0 (0.17)	6.4 (0.14)	6.1 (0.12)	5.9 (0.11)		
Percentage of persons of all ages who need help with personal care needs							
Quintile estimate (s.e.)	3.2 (0.17)	1.9 (0.12)	1.8 (0.12)	1.5 (0.11)	1.2 (0.11)	*	1.9 (0.06)
Cumulative estimate (s.e.)	3.2 (0.17)	2.6 (0.10)	2.3 (0.08)	2.2 (0.07)	2.0 (0.06)		
Percentage of persons of all ages with health insurance coverage							
Quintile estimate (s.e.)	81.5 (0.48)	84.4 (0.46)	86.9 (0.41)	88.0 (0.42)	89.1 (0.47)	*	85.5 (0.22)
Cumulative estimate (s.e.)	81.5 (0.48)	82.9 (0.36)	84.2 (0.28)	85.1 (0.24)	85.8 (0.22)		
2014							
Percentage of persons of all ages who had excellent or very good health							
Quintile estimate (s.e.)	59.5 (0.62)	66.1 (0.62)	67.5 (0.58)	68.9 (0.61)	71.0 (0.65)	*	66.5 (0.28)
Cumulative estimate (s.e.)	59.5 (0.62)	62.7 (0.46)	64.1 (0.37)	65.1 (0.32)	66.1 (0.28)		
Percentage of persons of all ages who failed to obtain needed medical care due to cost in the past 12 months							
Quintile estimate (s.e.)	5.9 (0.24)	5.2 (0.24)	5.3 (0.22)	5.0 (0.23)	4.7 (0.25)	*	5.3 (0.12)
Cumulative estimate (s.e.)	5.9 (0.24)	5.6 (0.19)	5.5 (0.15)	5.4 (0.13)	5.3 (0.12)		
Percentage of persons of all ages who need help with personal care needs							
Quintile estimate (s.e.)	3.0 (0.17)	1.9 (0.12)	1.5 (0.11)	1.5 (0.11)	1.5 (0.11)	*	1.9 (0.06)
Cumulative estimate (s.e.)	3.0 (0.17)	2.5 (0.11)	2.2 (0.08)	2.1 (0.07)	2.0 (0.06)		
Percentage of persons of all ages with health insurance coverage							
Quintile estimate (s.e.)	86.0 (0.43)	87.6 (0.47)	89.4 (0.35)	90.1 (0.36)	91.4 (0.39)	*	88.5 (0.19)
Cumulative estimate (s.e.)	86.0 (0.43)	86.8 (0.35)	87.6 (0.27)	88.1 (0.23)	88.6 (0.20)		

^a The quintile and cumulative estimates are base weighted.

* Indicates that a two-tailed t-test was significant at the .05 level.

Table 2. Nonresponse Bias Analysis of Sample Child Early Release Estimates: NHIS, 2013-2014^a

	Quintile 1: high response propensities	Quintile 2	Quintile 3	Quintile 4	Quintile 5: low response propensities	Quintile 5 vs. Top 4 Quintiles (t-test)	Final Weighted Estimate
2013							
Percentage of children aged 17 and under with a usual place to go for medical care							
Quintile estimate (s.e.)	94.0 (0.65)	94.9 (0.55)	96.3 (0.41)	96.6 (0.48)	97.5 (0.39)		95.9 (0.20)
Cumulative estimate (s.e.)	94.0 (0.65)	94.5 (0.42)	95.1 (0.31)	95.4 (0.27)	95.8 (0.22)		
Percentage of children aged 17 and under who had received an influenza vaccination in the past 12 months							
Quintile estimate (s.e.)	43.5 (1.26)	45.6 (1.25)	47.0 (1.32)	47.0 (1.36)	48.2 (1.42)		46.1 (0.62)
Cumulative estimate (s.e.)	43.5 (1.26)	44.6 (0.88)	45.4 (0.77)	45.7 (0.68)	46.2 (0.61)		
Percentage of children aged 17 and under who had an asthma attack in the past 12 months							
Quintile estimate (s.e.)	5.6 (0.60)	5.3 (0.55)	4.3 (0.46)	5.5 (0.63)	4.3 (0.47)	*	4.9 (0.25)
Cumulative estimate (s.e.)	5.6 (0.60)	5.4 (0.41)	5.1 (0.33)	5.2 (0.28)	5.0 (0.25)		
Percentage of current asthma among children aged 17 and under							
Quintile estimate (s.e.)	9.6 (0.69)	8.6 (0.69)	7.5 (0.61)	9.9 (0.79)	7.3 (0.61)	*	8.5 (0.31)
Cumulative estimate (s.e.)	9.6 (0.69)	9.1 (0.48)	8.6 (0.40)	8.9 (0.35)	8.6 (0.31)		
2014							
Percentage of children aged 17 and under with a usual place to go for medical care							
Quintile estimate (s.e.)	95.4 (0.45)	96.1 (0.51)	96.5 (0.51)	97.0 (0.45)	97.0 (0.44)		96.4 (0.22)
Cumulative estimate (s.e.)	95.4 (0.45)	95.7 (0.33)	96.0 (0.29)	96.2 (0.25)	96.3 (0.22)		
Percentage of children aged 17 and under who had received an influenza vaccination in the past 12 months							
Quintile estimate (s.e.)	46.4 (1.35)	48.6 (1.24)	48.7 (1.34)	48.9 (1.39)	50.0 (1.55)		48.8 (0.64)
Cumulative estimate (s.e.)	46.4 (1.35)	47.5 (0.93)	47.9 (0.78)	48.1 (0.69)	48.4 (0.63)		
Percentage of children aged 17 and under who had an asthma attack in the past 12 months							
Quintile estimate (s.e.)	3.7 (0.47)	4.3 (0.65)	4.8 (0.59)	4.8 (0.54)	3.9 (0.49)		4.3 (0.27)
Cumulative estimate (s.e.)	3.7 (0.47)	4.0 (0.40)	4.3 (0.35)	4.4 (0.30)	4.3 (0.26)		
Percentage of current asthma among children aged 17 and under							
Quintile estimate (s.e.)	8.1 (0.68)	9.0 (0.79)	8.9 (0.74)	9.2 (0.74)	8.9 (0.72)		8.8 (0.35)
Cumulative estimate (s.e.)	8.1 (0.68)	8.5 (0.52)	8.7 (0.44)	8.8 (0.38)	8.8 (0.34)		

^a The quintile and cumulative estimates are base weighted.

* Indicates that a two-tailed t-test was significant at the .05 level.

Table 3. Nonresponse Bias Analysis of Sample Adult Early Release Estimates: NHIS, 2013-2014^a

	Quintile 1: high response propensities	Quintile 2	Quintile 3	Quintile 4	Quintile 5: low response propensities	Quintile 5 vs. Top 4 Quintiles (t-test)	Final Weighted Estimate
2013							
Percentage of adults aged 18 and over who had an asthma attack in the past 12 months							
Quintile estimate (s.e.)	3.6 (0.27)	3.3 (0.26)	3.7 (0.29)	3.3 (0.27)	3.7 (0.34)		3.4 (0.13)
Cumulative estimate (s.e.)	3.6 (0.27)	3.5 (0.19)	3.6 (0.16)	3.5 (0.14)	3.5 (0.13)		
Percentage of current asthma among adults aged 18 and over							
Quintile estimate (s.e.)	7.4 (0.38)	6.8 (0.44)	7.7 (0.38)	7.2 (0.39)	7.3 (0.47)		7.1 (0.18)
Cumulative estimate (s.e.)	7.4 (0.38)	7.1 (0.29)	7.3 (0.23)	7.3 (0.20)	7.3 (0.19)		
Percentage of adults 18 and over with diagnosed diabetes							
Quintile estimate (s.e.)	12.4 (0.46)	9.9 (0.46)	10.0 (0.43)	8.4 (0.43)	8.3 (0.45)	*	9.3 (0.20)
Cumulative estimate (s.e.)	12.4 (0.46)	11.2 (0.32)	10.8 (0.25)	10.3 (0.22)	10.0 (0.20)		
Percentage of adults 18 and over who had 5 or more drinks in 1 day at least once in the past year							
Quintile estimate (s.e.)	18.1 (0.57)	21.4 (0.66)	22.7 (0.70)	22.7 (0.74)	23.9 (0.84)	*	22.6 (0.35)
Cumulative estimate (s.e.)	18.1 (0.57)	19.6 (0.46)	20.6 (0.39)	21.1 (0.35)	21.5 (0.33)		
2014							
Percentage of adults aged 18 and over who had an asthma attack in the past 12 months							
Quintile estimate (s.e.)	3.6 (0.25)	3.1 (0.24)	3.9 (0.34)	3.3 (0.27)	3.3 (0.33)		3.4 (0.14)
Cumulative estimate (s.e.)	3.6 (0.25)	3.4 (0.17)	3.5 (0.17)	3.5 (0.15)	3.4 (0.13)		
Percentage of current asthma among adults aged 18 and over							
Quintile estimate (s.e.)	7.7 (0.36)	7.1 (0.36)	8.7 (0.50)	7.5 (0.40)	7.5 (0.51)		7.6 (0.20)
Cumulative estimate (s.e.)	7.7 (0.36)	7.4 (0.26)	7.8 (0.25)	7.7 (0.22)	7.7 (0.19)		
Percentage of adults 18 and over with diagnosed diabetes							
Quintile estimate (s.e.)	12.7 (0.47)	10.1 (0.45)	9.5 (0.47)	7.7 (0.42)	8.6 (0.50)	*	9.1 (0.20)
Cumulative estimate (s.e.)	12.7 (0.47)	11.5 (0.33)	10.9 (0.27)	10.2 (0.24)	10.0 (0.21)		
Percentage of adults 18 and over who had (male=5/female=4) or more drinks in 1 day at least once in the past year							
Quintile estimate (s.e.)	19.2 (0.67)	24.4 (0.73)	25.0 (0.82)	26.2 (0.84)	26.9 (0.82)	*	25.3 (0.40)
Cumulative estimate (s.e.)	19.2 (0.67)	21.7 (0.50)	22.6 (0.46)	23.4 (0.42)	23.9 (0.37)		

Table 3. (continued)

	Quintile 1: high response propensities	Quintile 2	Quintile 3	Quintile 4	Quintile 5: low response propensities	Quintile 5 vs. Top 4 Quintiles (t-test)	Final Weighted Estimate
2013							
Prevalence of current cigarette smoking among adults aged 18 and over							
Quintile estimate (s.e.)	23.7 (0.68)	19.2 (0.63)	15.4 (0.57)	15.1 (0.60)	13.5 (0.63)	*	17.8 (0.30)
Cumulative estimate (s.e.)	23.7 (0.68)	21.6 (0.47)	19.7 (0.38)	18.7 (0.33)	17.8 (0.29)		
Percentage of adults aged 18 and over who had received an influenza vaccination in the past 12 months							
Quintile estimate (s.e.)	41.0 (0.79)	40.9 (0.79)	42.7 (0.76)	43.2 (0.86)	45.2 (0.93)	*	41.0 (0.39)
Cumulative estimate (s.e.)	41.0 (0.79)	40.9 (0.59)	41.5 (0.47)	41.9 (0.43)	42.4 (0.39)		
Percentage of adults aged 18 and over who met the federal physical activity guidelines							
Quintile estimate (s.e.)	38.8 (1.10)	46.9 (0.87)	51.1 (0.87)	51.6 (0.83)	54.8 (0.98)	*	49.4 (0.45)
Cumulative estimate (s.e.)	38.8 (1.10)	42.6 (0.73)	45.2 (0.60)	46.6 (0.50)	48.0 (0.45)		
Prevalence of obesity among adults aged 18 and over							
Quintile estimate (s.e.)	30.9 (0.81)	29.1 (0.71)	29.6 (0.70)	28.5 (0.75)	26.5 (0.78)	*	28.6 (0.36)
Cumulative estimate (s.e.)	30.9 (0.81)	30.1 (0.58)	29.9 (0.45)	29.6 (0.39)	29.1 (0.35)		
2014							
Prevalence of current cigarette smoking among adults aged 18 and over							
Quintile estimate (s.e.)	20.3 (0.68)	16.7 (0.59)	16.7 (0.75)	14.1 (0.60)	13.9 (0.73)	*	16.8 (0.33)
Cumulative estimate (s.e.)	20.3 (0.68)	18.6 (0.46)	18.0 (0.39)	17.2 (0.34)	16.7 (0.31)		
Percentage of adults aged 18 and over who had received an influenza vaccination in the past 12 months							
Quintile estimate (s.e.)	45.1 (0.79)	42.4 (0.77)	45.2 (0.85)	42.6 (0.84)	43.2 (0.97)		42.2 (0.42)
Cumulative estimate (s.e.)	45.1 (0.79)	43.8 (0.58)	44.2 (0.49)	43.9 (0.43)	43.8 (0.40)		
Percentage of adults aged 18 and over who met the federal physical activity guidelines							
Quintile estimate (s.e.)	41.2 (0.90)	47.9 (0.93)	50.5 (0.92)	51.8 (0.94)	53.0 (1.04)	*	49.3 (0.45)
Cumulative estimate (s.e.)	41.2 (0.90)	44.3 (0.68)	46.2 (0.56)	47.3 (0.49)	48.2 (0.45)		
Prevalence of obesity among adults aged 18 and over							
Quintile estimate (s.e.)	30.9 (0.65)	30.5 (0.78)	29.8 (0.70)	29.0 (0.75)	27.6 (0.88)	*	29.3 (0.37)
Cumulative estimate (s.e.)	30.9 (0.65)	30.7 (0.53)	30.4 (0.43)	30.1 (0.38)	29.7 (0.35)		

Table 3. (continued)

	Quintile 1: high response propensities	Quintile 2	Quintile 3	Quintile 4	Quintile 5: low response propensities	Quintile 5 vs. Top 4 Quintiles (t-test)	Final Weighted Estimate
2013							
Percentage of adults aged 18 and over who had ever been tested for HIV							
Quintile estimate (s.e.)	36.0 (0.85)	36.4 (0.79)	35.2 (0.79)	37.0 (0.84)	38.6 (0.93)	*	37.3 (0.41)
Cumulative estimate (s.e.)	36.0 (0.85)	36.1 (0.59)	35.8 (0.48)	36.1 (0.43)	36.5 (0.40)		
Percentage of adults aged 18 and older who experienced serious psychological distress during the past 30 days							
Quintile estimate (s.e.)	5.5 (0.33)	4.4 (0.33)	3.4 (0.27)	3.2 (0.28)	2.4 (0.29)	*	3.8 (0.15)
Cumulative estimate (s.e.)	5.5 (0.33)	5.0 (0.24)	4.5 (0.19)	4.2 (0.16)	3.9 (0.15)		
Percentage of adults aged 18 and over with a usual place to go for medical care							
Quintile estimate (s.e.)	81.4 (0.65)	83.2 (0.57)	84.7 (0.57)	86.3 (0.56)	87.7 (0.62)	*	83.7 (0.30)
Cumulative estimate (s.e.)	81.4 (0.65)	82.3 (0.44)	83.0 (0.36)	83.7 (0.31)	84.4 (0.28)		
Percentage of adults 18 and over who received the pneumococcal vaccination in the past 12 months							
Quintile estimate (s.e.)	27.2 (0.71)	23.1 (0.70)	22.0 (0.68)	21.3 (0.68)	18.7 (0.66)	*	21.0 (0.30)
Cumulative estimate (s.e.)	27.2 (0.71)	25.3 (0.53)	24.3 (0.43)	23.6 (0.37)	22.8 (0.33)		
2014							
Percentage of adults aged 18 and over who had ever been tested for HIV							
Quintile estimate (s.e.)	35.5 (0.84)	36.7 (0.77)	36.7 (0.87)	36.6 (0.89)	37.9 (0.96)		37.5 (0.41)
Cumulative estimate (s.e.)	35.5 (0.84)	36.1 (0.58)	36.3 (0.49)	36.3 (0.44)	36.6 (0.39)		
Percentage of adults aged 18 and older who experienced serious psychological distress during the past 30 days							
Quintile estimate (s.e.)	4.1 (0.27)	3.1 (0.25)	3.0 (0.25)	3.1 (0.29)	2.4 (0.26)	*	3.1 (0.12)
Cumulative estimate (s.e.)	4.1 (0.27)	3.6 (0.18)	3.5 (0.15)	3.4 (0.13)	3.2 (0.12)		
Percentage of adults aged 18 and over with a usual place to go for medical care							
Quintile estimate (s.e.)	85.3 (0.58)	85.3 (0.56)	86.7 (0.52)	86.3 (0.58)	87.7 (0.55)	*	85.3 (0.29)
Cumulative estimate (s.e.)	85.3 (0.58)	85.3 (0.43)	85.7 (0.33)	85.8 (0.30)	86.1 (0.27)		
Percentage of adults 18 and over who received the pneumococcal vaccination in the past 12 months							
Quintile estimate (s.e.)	29.8 (0.75)	23.8 (0.68)	22.6 (0.69)	20.2 (0.72)	20.0 (0.76)	*	21.8 (0.34)
Cumulative estimate (s.e.)	29.8 (0.75)	26.9 (0.53)	25.7 (0.44)	24.5 (0.38)	23.8 (0.35)		

^a The quintile and cumulative estimates are base weighted.

* Indicates that a two-tailed t-test was significant at the .05 level.

Table 2 presents the same bias analysis for four sample child ER estimates. Overall, we tend to see less evidence for potential nonresponse bias in these measures, especially for 2014. In 2014, there were no significant differences across these four measures when comparing the estimate for sample children in the low response propensity quintile to the estimate for the remainder of the sample. In 2013, we observe possible nonresponse bias in current asthma and having an asthma attack in the past 12 months. The data suggest we may be underestimating the prevalence of both of these measures. Looking at the final weighted estimate (person weights that include nonresponse adjustments), it appears that we may be mitigating the potential bias, but only slightly.

Finally, Table 3 presents the nonresponse bias analysis results for 12 sample adult ER estimates, again broken out by year. Across the two years, evidence of potential bias is observed for all but two of the 12 measures: current asthma and having an asthma attack in the past 12 months. The quintile-specific and cumulative estimates suggest that relatively large amounts of nonresponse bias could be present in some health measures. Current cigarette smoking and meeting the federal physical activity guidelines provide good examples. In 2013, the current smoking estimate for adults in quintile 1 (high response propensities) was 23.7%. The estimate for adults in quintile 5 (low response propensities) was 13.5%, roughly 15 percentage points lower. The corresponding quintile 1 and quintile 5 estimates for 2014 were 20.3% and 13.9% respectively. The cumulative estimate of current cigarette smoking in 2013, moving from the high response propensity quintile to the low response propensity quintile, drops from 23.7% to 17.8% (20.3% to 16.7% in 2014). Finally, the estimate for adults in quintile 5 (proxies for nonresponders) is significantly lower than the estimate for the other four quintiles combined (2013: 13.5% vs. 18.7%; 2014: 13.9% vs. 17.2%). Together, this information suggests that the estimate of current cigarette smoking could be an overestimate. For both years, applying final sample adult weights (that include nonresponse adjustments) has little impact on the estimates.

With regard to meeting federal physical activity guidelines, the analysis suggests a possible underestimate. The quintile 1 (high response propensities) estimate for 2013 was 38.8%, while the quintile 5 estimate was 54.8%. The corresponding figures for 2014 were 41.2% and 53.0%. Hence, interviewers encountered more difficulties in securing the participation of physically active sample adults. Not surprisingly, a steady increase in the cumulative estimates of exercise is observed for both years when moving from the high to low propensity quintile (2013: 38.8% to 48.0%; 2014: 41.2% to 48.2%). The estimate for adults in quintile 5 (again, proxies for nonresponders) is significantly higher than the estimate for the other four quintiles combined (2013: 54.8% vs. 46.6%; 2014: 53.0% vs. 47.3%). Unlike current cigarette smoking, however, applying the final sample adult weights has the effect of increasing the estimate for both years, as suggested by the bias analysis.

In sum, we found evidence for possible nonresponse bias in 16 of 20 key health outcomes in 2013 and 12 of 20 outcomes in 2014. The pattern of possible bias appears to be consistent across the two years of data, and may be substantial for estimates of needing help with personal care needs, diagnosed diabetes, serious psychological distress, binge drinking, current smoking, and exercise. Final weighted estimates suggest that the current nonresponse adjustment procedures used with the NHIS moves most estimates in directions suggested by the bias analysis. However, for some estimates where bias may be more substantial, the movement in estimates was small and may be insufficient. Hence, there appears to be room for improvements in minimizing nonresponse bias in NHIS health estimates. Additionally, by addressing nonresponse bias during data collection, via adaptive design, we may minimize the size of needed post-survey nonresponse weighting adjustments, producing reductions in variance and therefore gains in precision.

Declining response rates and increasing data collection costs—like those faced by the NHIS—have forced survey organizations to consider new data collection strategies. Adaptive design has emerged as a tool for tailoring contact to cases, based on data monitoring both between and during data collection periods. There is a range of data collection features that can be tailored, including mode of data collection, incentives, number of contacts, and case prioritization. The feature used for tailoring should be chosen to achieve specific survey goals. For example, one may want to increase response rate or sample representativeness for a given cost, or maintain the same response rate, while reducing cost. The planned case prioritization experiment is a test of the former; prioritization entails altering the amount of effort an interviewer expends on a case based on data monitoring metrics.

A2. Purpose and Use of the Information Collection

The Division of Health Interview Statistics, in collaboration with the Center for Adaptive Design (CAD) at the U.S. Census Bureau U.S. Census Bureau, is planning to test the use of adaptive design to increase sample representativeness and, therefore, reduce nonresponse bias in the National Health Interview Survey (NHIS). The planned experiment would be carried out beginning in July 2016, to encompass the third quarter of data collection. Data monitoring metrics and quality indicators will be used to assign a priority level to cases in each interviewer’s workload, with the aim of increasing sample representativeness and reducing nonresponse bias within a fixed cost.

This experiment requires no additional data collection from respondents, or changes to data collection procedures experienced by survey participants. No personally-identifiable information is used in building the response propensity models. Instead, response propensity and balancing propensity models will be constructed using Contact History Instrument (CHI) variables, Neighborhood Observation Instrument (NOI) variables, and Census 2010 or ACS 2009-13 variables (block group level) available on the 2015 Census Planning Database (CPD). Mid-month, each incomplete case will be assigned a response propensity score and a balancing propensity score based on these models and, using these scores, assigned a priority (low, medium, high) for contact by field representatives. Thus, this prioritization involves altering the amount of effort an interviewer expends on a case based on data monitoring metrics.

Experimental Design

Figure 2 provides a high-level visual overview of the proposed experiment. The experiment will be randomized at the interviewer level, meaning an entire interviewer’s workload will either be a control workload or an experimental workload. Each arm of the experiment will include 50% of interviewers, so the survey will be split evenly between experiment and control groups. For information on the minimum detectable differences that we expect to find, given conservative assumptions, see Appendix A.

While cases will be assigned to either the experimental or control groups before data collection starts, all cases will be treated the same for the first 15 days of data collection in each month. During this time, interviewers (referred to as “FR” in the figure, short for “field representative”) must ensure that at least one personal visit attempt is made for each case, and that contact history data and interviewer observations are documented for each case in the Contact History Instrument (CHI) and Neighborhood

Observation Instrument (NOI). The NOI data¹ will be combined with covariate data that is available before the start of data collection (CPD variables) and contact history instrument (CHI) data² that is collected during every contact attempt for each case to estimate two models: a response propensity (likelihood) model and a balancing propensity (or bias likelihood) model.

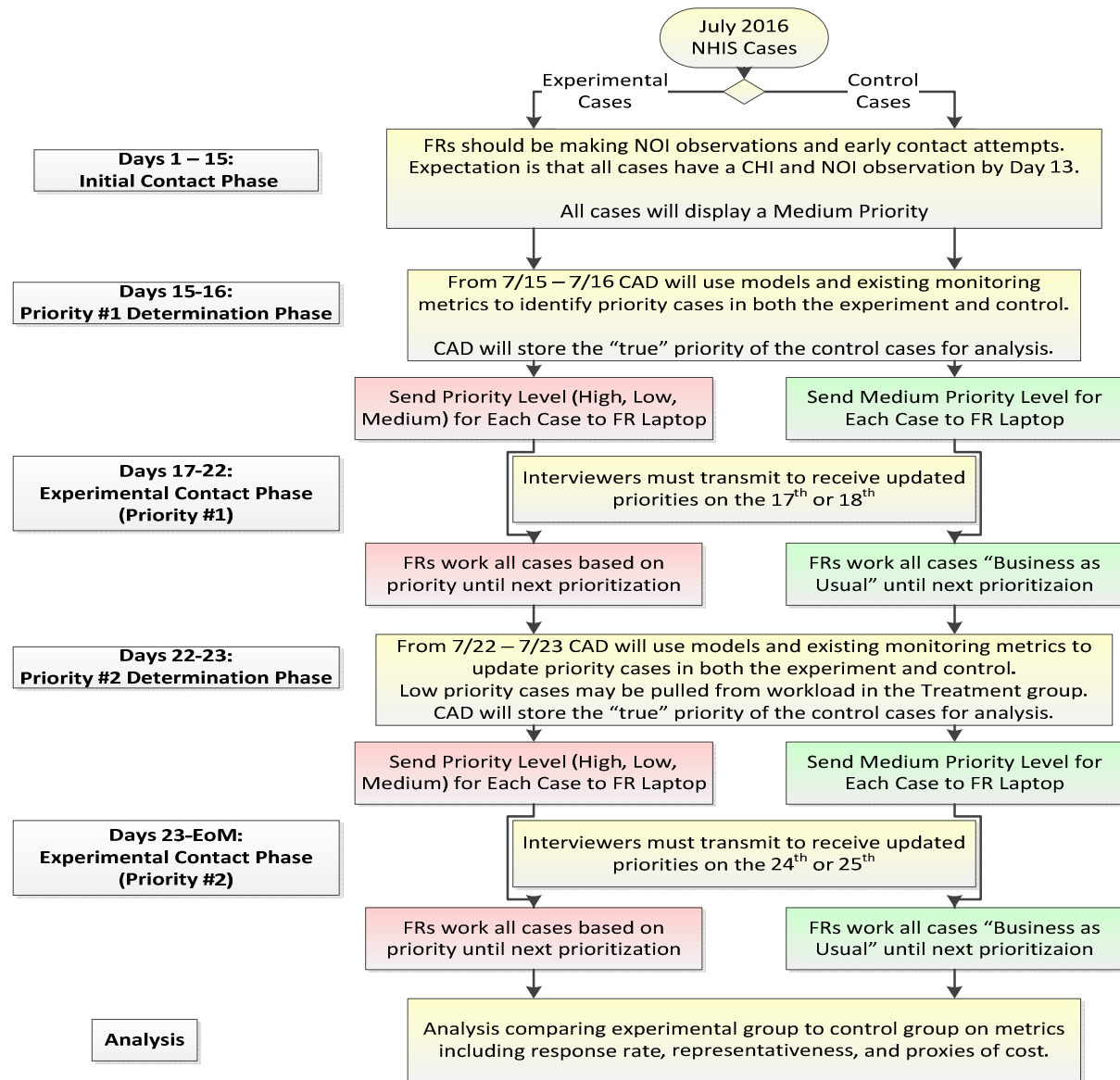


Figure 2. Conceptual Overview of NHIS Adaptive Design Experiment

Response Propensity (Likelihood) Model. A response propensity or likelihood model (logistic regression) will be run on the 15th to 16th of each month in quarter three. Response status will be the dependent variable, and variables from CHI, NOI, and the CPD will serve as predictors. Variable selection relied

¹ These data include information the field representative observes about the housing unit that may be predictive of key survey outcomes (as well as response).

² These data include information about the date, time, and outcomes of contact attempts, contact strategies used, and respondent reluctance.

heavily on survey theories of response and past assessments of CHI and NOI variables related to NHIS response [10, 11]. Unlike propensity models used in weighting adjustments [12], where we are concerned about finding variables that predict both response and the outcomes of interest, here we are concerned solely with predicting likelihood of response. After the model is run, each case will be assigned a response propensity or likelihood score. Table 4 presents the set of variables to be used in the response propensity model.

Table 4. Variables Selected for Inclusion in the Response Propensity Model

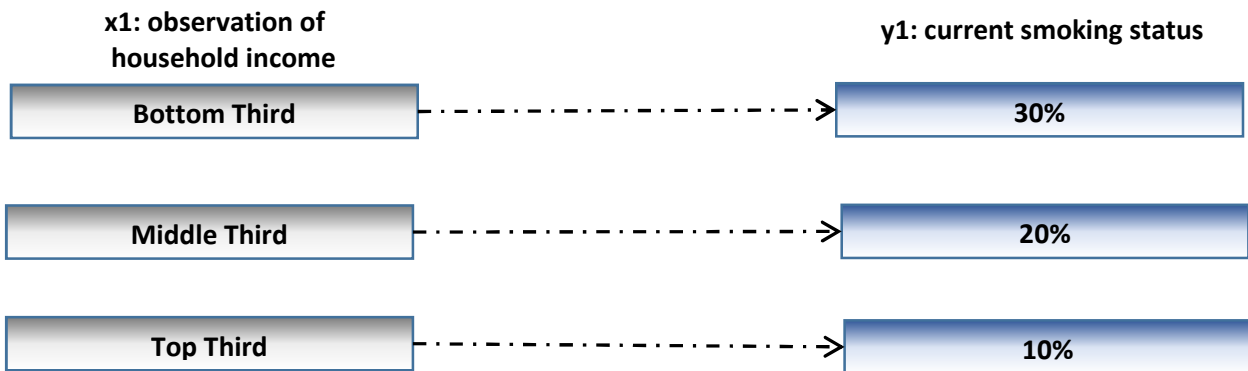
CHI Variables
Access barriers
Prior peak contact
Prior hard refusals
Prior privacy concerns
Total prior personal visits
Total prior telephone attempts
Total prior incoming contacts
Day of month
NOI Variables
Interviewer assessment of the condition of the sample unit
Interviewer assessment of whether or not the sample unit has 3+ door locks
Interviewer assessment of whether or not all household occupants are over the age of 65
Interviewer assessment of whether or not a language other than English is spoken by residents
Interviewer assessment of whether or not one more residents is disabled, handicapped, or has a chronic health condition
Interviewer assessment of whether or not residents may be smokers
Interviewer assessment of whether or not one or more adults of the household are employed
CPD Variables
% of 2010 Census housing units with no registered occupants on Census day (vacant)

Balancing Propensity (Bias Likelihood) Model. A balancing propensity model (logistic regression) will also be estimated on the 15th and 16th of the month. Like the response propensity model, response status will be the dependent variable and NOI and CPD variables will be used as predictors. (Table 5 presents the set of variables to be used in the balancing propensity model.) What separates the balancing propensity model from a standard response propensity model is the inclusion of predictors or covariates related to key survey outcomes, in this case health variables. (The selection of variables for inclusion in the balancing propensity model is discussed in Appendix B.) Based on model output, each case will be assigned a balancing propensity score that will be used to identify nonrespondents most likely to reduce bias in key survey variables if converted to respondents. Essentially, the balancing propensity model will be used to identify cases that are most unlike the set of sample members that have already responded. In essence, we are attempting to balance response across groups defined by variables, available for both responding and nonresponding households, related to health outcomes on the NHIS.

Table 5. Variables Selected for Inclusion in the Balancing Propensity Model

NOI Variables
Interviewer assessment of the household’s income relative to the general population
Interviewer assessment of whether or not all household occupants are over the age of 65
Interviewer assessment of whether or not a language other than English is spoken by residents
Interviewer assessment of whether or not one or more adults of the household are employed
Interviewer assessment of whether or not one more residents is disabled, handicapped, or has a chronic health condition
Interviewer assessment of whether or not residents may be smokers
Interviewer assessment of whether or not the sample unit has a well-tended yard or garden
Interviewer assessment of the condition of the sample unit
Interviewer assessment of whether or not the walls of the sample unit are damaged
CPD Variables
% of ACS pop. 25+ that have a college degree or higher
% of ACS pop. 25+ that are not high school grads
% of 2010 Census occ. housing units with female householder(s) and no husband
% of 2010 Census occ. housing units where householder and spouse in same household
Average aggregated household income of ACS occupied housing units
Average aggregated value for ACS occupied housing units
% of ACS population that is uninsured
% of ACS civilians 16+ that are unemployed
% of ACS population classified as below the poverty line
% of 2010 Census housing units with no registered occupants on Census day (vacant)

As to how sample balancing helps to improve representativeness, take the following example. We know from past analyses of NHIS data that interviewer observations of the household’s income relative to the general population (bottom third, middle third, top third) is related to current smoking status among adults (polychoric correlation from 2013: 0.2745). Assume that the **true** relationship between interviewer assessment of household income and current smoking status is as follows:



Finally, assume that the survey achieves a 75% response rate. To get the most accurate estimate of current smoking status, how would we want the response rate to be distributed across the household income groups? In Figure 4, three response rate options are presented. In option A, all of the nonresponse is coming from the top third income group. Based on our knowledge of the true

relationship between income and smoking status (above), this would likely lead to an overestimate of current smoking among adults. In option B, while nonresponse is more equitably distributed, both the middle and top third income groups are under-represented. As a result, we are likely still overestimating current smoking status. Ultimately, option C is our goal. In this option, all three income groups have a 75% response rate. By ensuring that households from these groups respond at the same rate, our estimate of current smoking status should be a very close reflection of the true current smoking rate. This is what is meant by sample balancing.

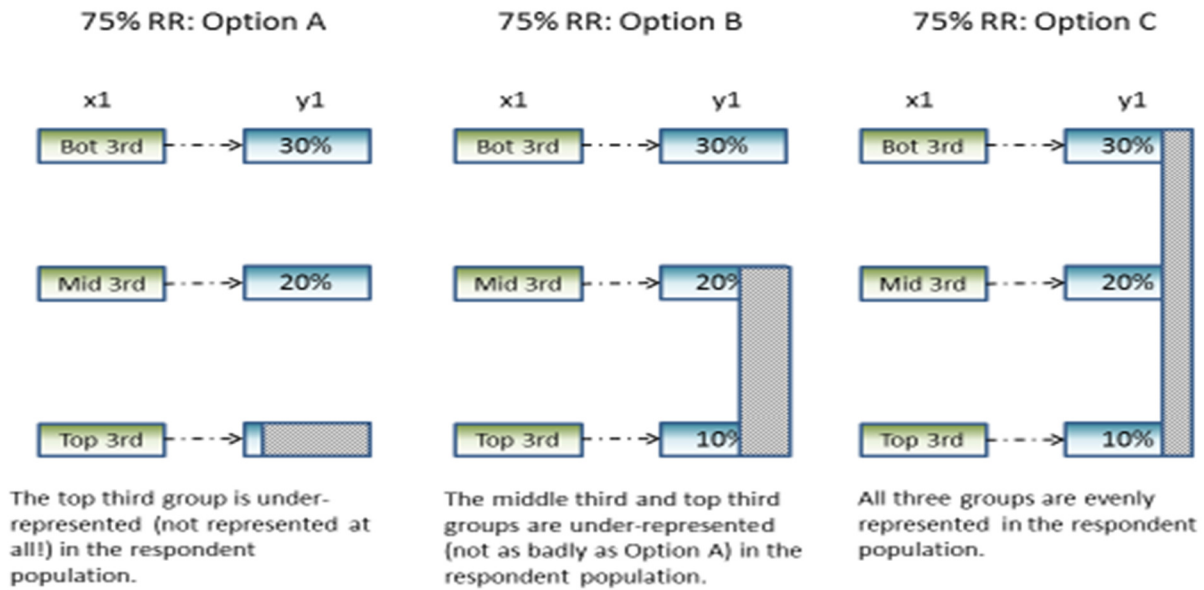


Figure 3. Balancing Sample to Ensure Representativeness

For a more thorough discussion of sample balancing and how it may be optimal for some estimators and can protect against nonresponse bias, see: Royall and Herson 1973[13]; Royall and Herson 1973a[14]; Royall 1992[15]; Sarndal and Lundquist 2015[16]; Shouten et al. 2015[17]; and Valliant et al. 2000[18]. Also, for a discussion of how the balancing propensity model covariates were selected for this experiment and response rates for groups defined by these covariates, see Appendix B.

Using the combination of response propensity and balancing propensity scores, the priority (H: high, M: medium, L: low) for each incomplete case will be determined. Together, the respective scores allow us to calculate how strategic a case is to achieving the experimental goal of improving sample representativeness and managing costs without negatively affecting overall response. For example, we would want to assign high priority to a case that, if completed, would improve representativeness. However, this decision would have to be balanced by how likely the case is to be completed. This is conceptually visualized in Figure 4.

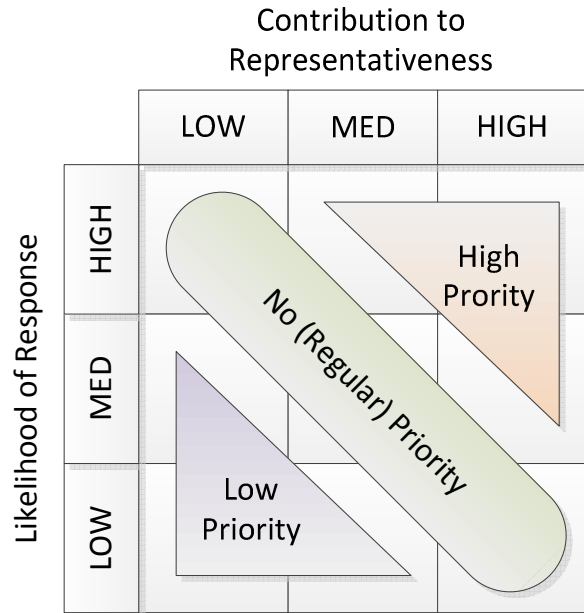


Figure 4. Conceptual Visualization of Prioritization

If the case would contribute moderately to representativeness but has a very low likelihood of responding based on the response propensity model, then assigning the case to receive greater priority may simply lead to increased costs with no gains in sample representativeness. Similarly, if a case had a very high likelihood of response, but would not improve representativeness because we already had an overabundance of respondents with the same characteristics, we would not want to assign a high priority to the case. All cases (both experimental and control) will be assigned a priority status, which may be useful later for propensity matching or direct comparisons among subgroups in the experimental vs. control groups. However, only experimental cases will have the priority statuses displayed in field management reports and on interviewer laptops. Interviewers will then be asked to work their cases according to the assigned priorities.

It is important to note that the response and balancing propensity models, as shown in Figure 2, will be re-estimated on the 22nd to 23rd of each month. Updated priorities on remaining incomplete cases will be pushed out to FR’s laptops at that time. The experimental cases with low priority status may be pulled from the interviewer’s workload in the final week of data collection.

At each prioritization point in data collection (the 15th-16th and 22nd-23rd), the execution of business rules to determine each case’s priority classification will occur in two steps. First, we will execute the initial model-based prioritization, and then we will evaluate those initial priorities against four specific exceptions to determine the final prioritization. On the 15th-16th, we will assign 20% of open cases a high priority, and 20% of cases a low priority. At the 22nd-23rd prioritization, we will evaluate remaining cases against the thresholds set on the 15th-16th to update priorities.

As previously described, each case will be assigned a response propensity score and a balancing propensity score. The response propensity score estimates the likelihood that a case will be completed (high versus low likelihood) and the balancing propensity score estimates how unlikely a case is to be in the respondent population (high versus low value). Because these scores are bounded by zero and one

and are the result of logistic regressions, the distributions of estimated propensities are often skewed. As a result, we transformed the propensity estimates to obtain distributions closer to standard normal distributions, and then we standardized them so the 50th percentiles of each propensity lie at zero.

Creating a scatterplot of the standardized response and balancing propensities *for cases still in the field* results in a plot with four quadrants, as follows: TR (top right: higher value, higher likelihood); TL (top left: higher value, lower likelihood); BR (bottom right: lower value, higher likelihood); and BL (bottom left: lower value, lower likelihood). Again, the axes are the “average” value and likelihood. Figure 5 shows these scatterplots for February and March of 2016. Approximately, though not exactly, 25% of the sample resides in each quadrant.

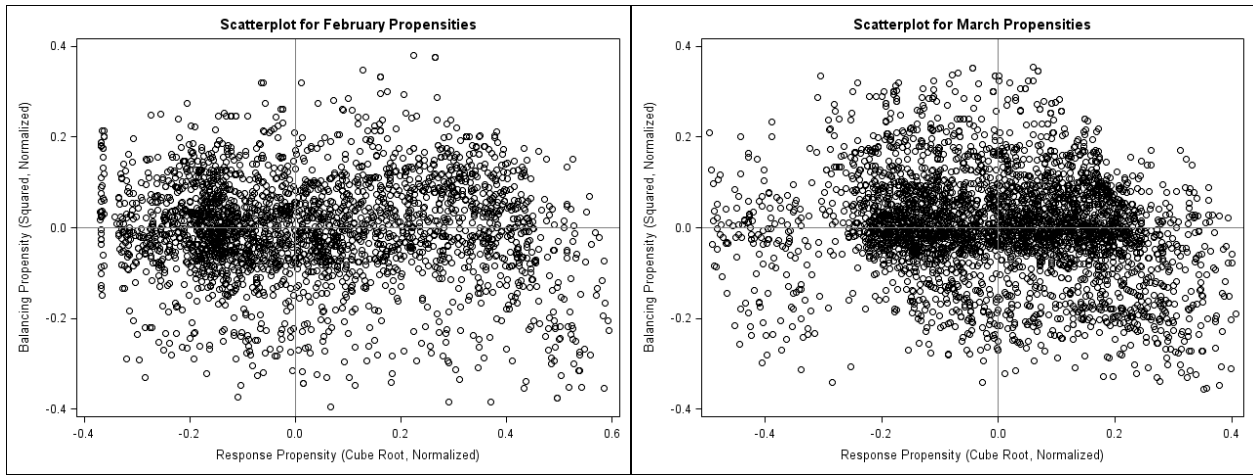


Figure 5. Scatterplots for February and March Standardized Response and Balancing Propensities

Given the fact that we want to assign high priority to cases that have higher than average value and higher than average likelihood of response, we want to focus our efforts on the TR (Top Right) Quadrant. Within that quadrant, cases farthest away from the origin (farthest top and farthest right) are going to be the cases with the highest value and the highest likelihoods, so we want to select cases farthest away from the origin first for prioritization. Similarly, we want to assign low priority to cases that have lower than average value and lower than average likelihood of response, so we want to focus on the BL (Bottom Left) Quadrant, first selecting cases farthest away from the origin in the negative direction.

To operationalize the selection, we can use a simple Cartesian distance metric to estimate the distance of each case from the origin and rank cases within their quadrant:

$$\hat{d}_i = \sqrt{(\hat{\rho}_i^R)^2 + (\hat{\rho}_i^B)^2} ,$$

where \hat{d}_i is the distance we can rank on (where larger distances are ranked higher), $\hat{\rho}_i^R$ is the standardized estimate of the response propensity and $\hat{\rho}_i^B$ is the standardized estimate of the balancing propensity. We can then rank cases based on their distances in the TR and BL quadrants and prioritize so that 20% of cases are assigned a high priority and 20% of cases are assigned a low priority. Figure 6 shows the same scatterplots above with the high and low priority cases identified.

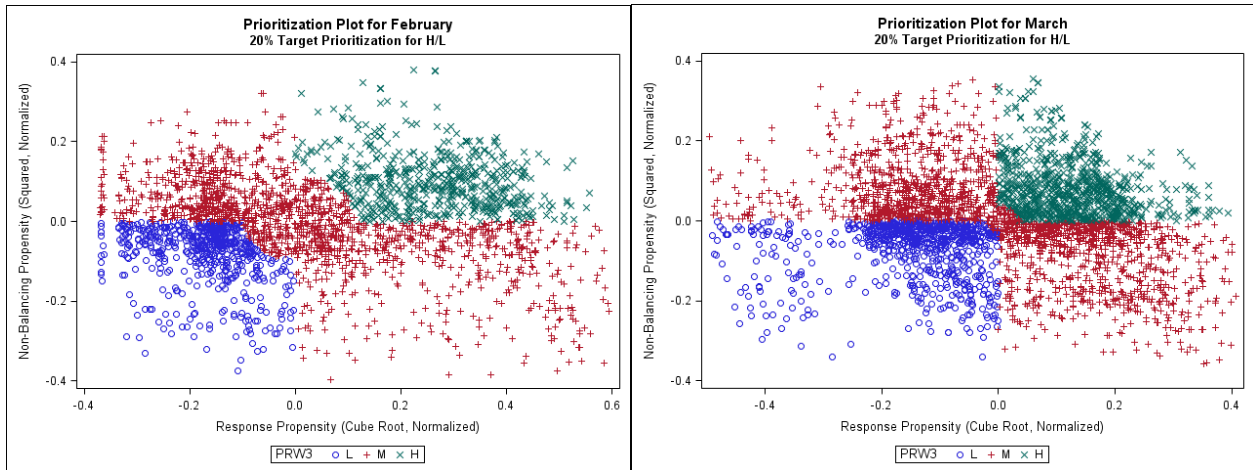


Figure 6. Figure 4 Scatterplots with Assigned Prioritization

Based on the Day 16 prioritization, the minimum distance for which a case was assigned a high or low priority will be retained for use in the Day 23 reprioritization that will be described later.

After prioritization has been assigned with respect to the models, as shown above, five business rules will be evaluated which could override this priority assignment for individual cases. Those rules are:

- (1) If a CHI and or a NOI entry has not been completed at the time of prioritization, these cases will receive an automatic high priority. CHI and NOI data are crucial to our ability to prioritize, since they comprise the variables in our propensity models. Further, NOI and CHI data provide insight into how a case is progressing in the field. Lacking these data means no progress has been made, and the case should be prioritized to rectify that situation. FRs will be evaluated partially based on their NOI and CHI completion rates by the 15th of the month, reducing the number of cases that should be affected by this rule.
- (2) If there are multiple sample units in a single Group Quarter (GQ) or Multi-Unit Structure (MU), and an access barrier is encountered, all cases in that GQ or MU will receive the highest priority of all cases in the GQ or MU. In other words, if there are 3 cases within a GQ, and two are medium priority and one is high, we will assign all three cases a high priority. Due to the additional effort commonly required by FRs to gain entry to GQs and MUs, we want to be as efficient about this additional effort as possible. We expect this rule to be executed rarely given the low level of clustering in the NHIS. In addition, not all sample months have GQ sample cases, further reducing the frequency of this rule being executed.
- (3) Extra units, additional units, and spawns will all inherit the priority of their parent case until we have a scheduled reprioritization. These cases are usually geographically co-located with the parent unit, and similarly to the units in GQs and MUs, we want the FRs to be efficient about their effort. Therefore, we will assign the same priorities so the cases are worked with the same level of effort. Their priorities may diverge when we reprioritize cases at a later date, given information in the NOI or CHI.
- (4) Partial interviews at the time of prioritization will not be assigned a low priority (they might be medium or high). These cases have already been worked, and the interviewer has collected partial data from the household. In order to remain efficient and obtain completed interviews

where possible, these cases will not be assigned a low priority, and if we pull cases in Week 4 from interviewer laptops, these cases will not be pulled.

- (5) If an appointment has been scheduled with a household, we will not allow the case to be a low priority in Week 4. We do not want to pull a case from the interviewer's laptop if an appointment has been scheduled.

Once these rules have been evaluated for all cases, priorities will be pushed down to interviewer laptops. Interviewers in the experiment will receive the "true" priorities, while interviewers in the control will receive only "medium" priorities.

Reprioritization will happen on the 22nd-23rd. Rather than prioritize from scratch by selecting a new minimum distance for prioritization as described earlier, we will use the thresholds from the prioritization that occurred on the 15th-16th. This ties closely with the idea of balance. If interviewers work cases according to priorities, spending more time in Week 3 on high priority cases and no to very little effort on low priority cases, more high priority cases will be completed. This will make the low priority cases which were over-represented less over-represented, and the under-represented cases less under-represented. By improving sample balance, more cases may be considered medium priority during Week 4, and those cases can be worked normally for the remainder of data collection. This will reduce the number of cases eligible to be pulled from interviewer laptops, while also reducing the number of cases that require additional effort late in data collection. From the perspective of contact burden, having a flexible prioritization that allows cases to move from high priority to medium priority will help avoid cases being overworked. Similarly, allowing cases to move from low priority to medium priority, or medium priority to high priority ensures cases that will help meet dynamic data collection goals are targeted for additional effort.

To understand how cases might be reassigned to different priorities during Week 4, and the effect this would have on resolution rates during Week 3, we simulated how cases could move across different priority classifications as interviewers complete more high priority cases successfully. To do this, we made some simple but reasonable assumptions:

- We assumed the response rate of the low priority cases would not increase during Week 3 (as the cases are not being worked);
- We assumed the response rate of the medium priority cases would increase by 15% during Week 3; and
- We assumed the response rate of the high priority cases could increase by 20%, 25%, or 30%, depending on how successful interviewers' efforts were at converting high priority cases.

Table 6 below includes cases that were open in Week 4, so cases that would not have been completed during the prioritization of Week 3. It is clear that there is some movement between the different strata. In February, across the two prioritization times (Week 3 and Week 4), both the high and low priority strata had net losses of unresolved cases, while the medium priority experienced a net gain. Conversely, in March, while cases still shifted from the low priority into medium, some cases in the medium strata actually moved to the high strata. These differences may be due to varying compositions of nonresponders at weeks 3 and 4 for the two months.

Table 6. Week 4 Reprioritization - Net Case Movement by Strata for Open Cases in Week 4

		Case Totals by Priority Category and Completion Pattern			
		February, 2016		March, 2016	
Completion Pattern	Priority Level	Week 3 Priority	Week 4 Priority	Week 3 Priority	Week 4 Priority
High Increases 20%	High	462	314	662	756
	Medium	1483	1819	2150	2118
	Low	588	400	848	786
High Increases 25%	High	429	321	621	739
	Medium	1483	1789	2150	2105
	Low	588	390	848	775
High Increases 30%	High	396	298	572	714
	Medium	1483	1776	2150	2097
	Low	588	393	848	759

Of particular interest is the fact that the number of cases in the low priority strata in Week 4 exceed, if only slightly, the number of cases in the high priority strata in Week 4. This is important for cost control, as we need to reduce contact attempts on some cases in order to reallocate resources to other cases. Cases that were assigned a high priority status in Week 3 and were considered completed for the sake of our simulations are not included in these tables, as we are showing the movement of open cases across strata.

Finally, we can estimate the effect of each of these completion patterns. Table 7 shows that the biggest increase in response rate that we achieve is when we prioritize cases and the high priority cases are completed at a rate of 20%. After that, the response rate does not appreciably increase, likely because the number of high cases remaining is diminishing. This also supports our theory that the low priority cases are less likely to respond – assuming they will not respond does not diminish the response rate significantly.

Table 7. Week 3 Effect on Resolution Rates by Completion Pattern

Response Pattern	Resolution Rate (end of Week 3)	
	February, 2016	March, 2016
Actual Response Pattern	53.87%	51.99%
High Increases 20%	60.33%	58.66%
High Increases 25%	60.75%	59.20%
High Increases 30%	61.28%	59.74%

Evaluation of the Experiment

Both process and outcome measures will be examined as part of the evaluation of the experiment.

Distribution of Interviewer Effort

The success of the experiment is predicated on interviewers adjusting their effort according to the prioritizations assigned to their cases. Once the case prioritizations are assigned and pushed out to FR laptops on the 16th of each month, the distribution of contact attempts by case prioritization will be monitored. Focusing on the experimental group, for example, are interviewers shifting their contact attempts away from low priority cases toward high priority cases? When interviewers make their contact attempts will also be monitored. Are interviewers in the experimental group moving attempts on high priority cases to more lucrative time slots (e.g., evening hours and weekends)?

Sample Representativeness

As noted previously, the primary objective of this experiment is to improve sample balance or representativeness (minimize nonresponse bias) within current cost constraints while maintaining current response rates.

During each month of the quarter, the R-indicator (a measure of variation in response propensities that ranges between 0 and 1) will be monitored for both the control and experimental groups using the aforementioned balancing model [19, 20]. The R-indicator provides an assessment of sample representativeness conditional on the covariates included in the balancing model. (Again, the balancing model will include covariates that are a mixture of interviewer observational measures along with decennial Census and ACS measures captured at the Census block group or tract level. The covariates selected for the model are those that are highly correlated with key health outcomes.) A decreasing R-indicator represents an increase in the variation in response propensities, suggesting less sample balance or representativeness. An increasing R-indicator indicates less variation in response propensities and better sample balance. Comparing R-indicators for the experimental and control groups during and at the end of data collection will provide an assessment of whether the experimental case prioritizations led to improved representativeness (indicated by a higher value on the R-indicator for the experimental compared to the control group).

Post-data collection, demographic characteristics of the two groups, including age, sex, race/ethnicity, and education level, will be compared. Next, point and variance estimates of 16 select health variables included in NHIS's Early Release (ER) indicator reports (e.g., health insurance coverage, failure to obtain needed medical care, cigarette smoking and alcohol consumption, and general health status) will be compared between the two groups. If increased representativeness of the sample has been achieved, differences in the estimates across the two groups should be observed. Using past analyses of nonresponse bias (see Section 2), statements can be made as to whether the experimental group estimate is moving in a direction that minimizes nonresponse bias.

Reductions in variance would also suggest increased representativeness. A more representative sample should result in less extreme nonresponse and poststratification adjustments, reducing the variability of sample weights. Design effects and standard errors for all estimates examined will be compared between the control and experimental groups.

All comparisons in this section would be performed overall and by age, sex, and race/ethnicity subgroups.

Response Rates and Survey Costs

Since the goal is to improve representativeness within current cost constraints while maintaining current response rates, overall and module-specific response rates, completed interview rates, sufficient partial interview rates, and refusal rates will be compared by control and experimental groups. In addition, the total number of contact attempts (in-person visits and phone calls) for the two groups will serve as a proxy for survey costs.

Other Data Quality Indicators

Additional measures of data quality will also be examined by adaptive design group, including, but not limited to, item nonresponse rates (“don’t know” and “refused” responses), item response times, and survey breakoffs.

A12. Estimates of Annualized Burden Hours and Cost

Requiring no additional data collection nor changes to data collection procedures experienced by NHIS respondents, the planned adaptive design experiment would not alter the previously-approved estimates of annualized burden hours and survey administration costs. Costs associated with implementing the experiment are limited to interviewer training expenses, and are covered by funds designated to Methodological Projects listed in Line 5 of the previously-approved burden table.

References

- [1] Curtin, R.; Presser, S.; Singer, E. (2000). "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly*, 64: 413-428.
- [2] Olson, K. (2006). "Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias." *Public Opinion Quarterly*, 70: 737-758.
- [3] Groves, R.; Heeringa, S. "Responsive design for household surveys: tools for actively controlling survey errors and costs." *Journal of the Royal Statistical Society, Series A*. 169 (2006): pp 439-457.
- [4] Axinn, W.; Link, C.; Groves, R. "Responsive Design Demographic Data Collection and Models of Demographic Behavior." *Demography*. 48 (2011): pp 1127 – 1149.
- [5] Calinescu, M.; Bhulai, S.; Schouten, B. "Optimal Resource Allocation in Survey Designs." *European Journal of Operations Research*. 226:1 (April, 2013): pp 115-121.
- [6] Mercer A.; Caporaso, A.; Cantor, D.; Townsend, R. "How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys." *Public Opinion Quarterly*. 79: 1 (2015): pp 105-129.
- [7] Coffey, S.; Reist, B.; Zotti, A. "Static Adaptive Design in the NSCG: Results of a Targeted Incentive Timing Study." Presentation at the Joint Statistical Meetings, 2015. Seattle, WA.
- [8] Konicki, S. "Adaptive Design Research for the 2020 Census." Presentation at the Joint Statistical Meetings, 2015. Seattle, WA.
- [9] Wagner, J.; West, B.; Kirgis, N.; Lepkowski, J.; Axinn, W.; Ndiaye, S. "Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection." *Journal of Official Statistics*. 28:4 (2012): pp. 477-499.
- [10] Erdman, C. and Dahlhamer, J. (2013) "Evaluating Interviewer Observations in the National Health Interview Survey: Associations with response propensity." Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods. Montreal, CN.
- [11] Walsh, R.; Dahlhamer, J.; Bates, N. (2014) "Assessing Interviewer Observations in the NHIS." Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods. Boston, MA.
- [12] Little, R. J.; Vartivarian, S. L. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology*. 31 (2005): pp. 161-168.
- [13] Royall, R. M.; Herson, J. "Robust estimation in finite populations I." *Journal of the American Statistical Association* 68. (1973): pp. 880-889.
- [14] Royall, R. M.; Herson, J. "Robust estimation in finite populations II: Stratification on a size variable." *Journal of the American Statistical Association* 68. (1973): pp. 890-893.
- [15] Royall, R. M. "Robustness and optimal design under prediction models for finite populations" *Survey methodology* 18 (1992): pp. 179-185.
- [16] Sarndal, C-E; Lundquist, P "Accuracy in estimation with nonresponse: A function of degree of imbalance and degree of explanation." *Journal of Survey Statistics and Methodology*. 2 (2014): 361-387.
- [17] Schouten, B.; Cobben, F.; Lundquist, P.; Wagner, J "Does more balanced survey response imply less non-response bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* (2015) doi: 10.1111/rssa.12152.
- [18] Valliant, R.; Dorfman, A. H.; Royall, R.M. *Finite Population Sampling and Inference: A Prediction Approach* (2000) John Wiley & Sons, New York
- [19] Schouten, B. Cobben, F. Bethlehem, J. "Indicators for representativeness of survey response." *Survey Methodology*. 35.1 (June 2009): pp 101 – 113.
- [20] Schouten, B. Shlomo, N. Skinner, C. "Indicators for monitoring and improving representativeness of response." *Journal of Official Statistics*. 27.2 (2011): pp 231 – 253.
- [21] Snedecor, G. W.; Cochran, W.G. (1989). *Statistical Methods*. 8th ed. Ames, IA: Iowa State University Press; 8ed.

[22] Dahlhamer, J; Gindi, R.; Bates, N.; Walsh, R. (2014) "Judgments under Uncertainty: Assessing the Quality of Interviewer-Generated Paradata." Paper prepared for the 26th International Workshop on Household Survey Nonresponse, Leuven, Belgium.

Appendix A Minimum Detectable Differences

The power of this experiment to make generalizations about case prioritization rests on having sample sizes that result in reasonable standard errors for our analytic outcomes of interest. We assume that there are approximately 6,400 cases in the NHIS in a given month, and so approximately 3,200 would be randomized (through interviewer selection) into each of two groups, experiment and control.

For the purposes of these calculations, we make a simple assumption that by the 15th of the month, which is 50% of the way through data collection, interviewers will have resolved 50% of their cases. This is a conservative assumption, resulting in a lower number of unresolved cases than we would expect to see halfway through data collection.

This leaves us with 1,600 cases in each of the treatments. If we assign 20% of cases a high priority and 20% of cases a low in each of the treatments, we have the following case breakdown:

- H: 320 cases
- M: 960 cases
- L: 320 cases

Minimum Detectable Difference Formula [21]:

$$\delta = \left[(Z_\alpha + Z_\beta)^2 \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) D \right]^{\frac{1}{2}}$$

We assume a design effect of 1.4 based on the 2015 NHIS. Given that the new design is less clustered than the old design, the design effect we are using is conservative. In addition, we will use an alpha-crit of 0.10 (Census Requirements), and a beta-crit of either 0.10 or 0.20. Finally, we will assume that there is no difference between the two groups, and that the proportions we are comparing are $p_1=p_2 = 0.50$. This will maximize the variance and make a conservative estimate of the required true difference required to see a change.

Minimum Detectable Differences for Strata Response Rates

Here, we are looking only at cases that are unresolved in the second half of data collection. In addition, note that these percent differences will hold for all items we may compare at the strata level including attempts per case, variability of contact times, and others.

Low or High Priority Strata (n=320 cases):

$$A1: \beta = 0.10: \delta = \left[(1.96^2 + (-1.28)^2) \left(\frac{0.5^2}{320} + \frac{0.5^2}{320} \right) 1.4 \right]^{\frac{1}{2}} = 10.95\%$$

$$A2: \beta = 0.20: \delta = \left[(1.96^2 + (-0.84)^2) \left(\frac{0.5^2}{320} + \frac{0.5^2}{320} \right) 1.4 \right]^{\frac{1}{2}} = 9.97\%$$

Medium Priority Strata (n=960 cases):

$$B1: \beta = 0.10: \delta = \left[(1.96^2 + (-1.28)^2) \left(\frac{0.5^2}{960} + \frac{0.5^2}{960} \right) 1.4 \right]^{\frac{1}{2}} = 6.32\%$$

$$B2: \beta = 0.20: \delta = \left[(1.96^2 + (-0.84)^2) \left(\frac{0.5^2}{960} + \frac{0.5^2}{960} \right) 1.4 \right]^{\frac{1}{2}} = 5.76\%$$

We would need to see differences of around 8%-10% in the High and Low identified groups to be able to make statistically significant comparisons between the treatment and control groups. This, however, assumes a 50% response rate across all groups, which maximizes the variance in the formula and also means that a larger difference is required to detect a statistical difference. If, say the response rates in the high group were closer to 70%, the requirement for statistical significance in A1 above would drop to: 10.05% and 8.48% depending on whether the design effect is incorporated.

Minimum Detectable Differences for R-Indicators (Full Sample or Variable Level):

Here, we are looking at all cases in each treatment. Note: These percent differences will hold for all items we may compare at the full sample level.

$$A1: \beta = 0.10: \delta = \left[(1.96^2 + (-1.28)^2) \left(\frac{0.5^2}{3200} + \frac{0.5^2}{3200} \right) 1.4 \right]^{\frac{1}{2}} = 3.46\%$$

$$A2: \beta = 0.20: \delta = \left[(1.96^2 + (-0.84)^2) \left(\frac{0.5^2}{3200} + \frac{0.5^2}{3200} \right) 1.4 \right]^{\frac{1}{2}} = 3.15\%$$

We are including all cases from the treatments in these comparisons, so we will be able to detect much smaller differences at a statistically significant level.

Appendix B: Identifying Variables for Use in a Balancing Propensity Model

Introduction

To identify variables for inclusion in the balancing propensity model, we performed a correlation analysis involving all 15 of the available NOI variables and 35 Census 2010 or ACS 2009-13 variables available on the 2015 CPD. We then correlated each of the 50 NOI and CPD variables with 84 health variables from the 2013 NHIS: 19 person-level variables from the person file, 22 sample child measures, and 43 sample adult measures. Included in the 84 health variables were all measures included in the NHIS Early Release (ER) Program.

Methods

Selection of Health-Related Variables

Health outcomes from the person, sample child, and sample adult files with universes of “all persons,” “all sample children,” or “all sample adults,” were included in the analysis, with some exceptions. Given their importance, ER variables were included by default. As a starting point, this produced a total of 19 variables from the person file, 49 from the sample child file, and 86 from the sample adult file.

All 19 person file variables were retained for the analysis. To narrow the list of sample child and sample adult variables, SAS PROC VARCLUS was used. (Note that the ER measures were not included in the VARCLUS procedures as all were retained in the subsequent analysis.) Clusters of similar health measures were identified using the VARCLUS procedure and the variable with the smallest $1 - R^2$ ratio was selected from each cluster. This reduced the number of sample child measures for analysis to 22 (including ER measures), and the number of sample adult variables to 43 (including ER measures). In total, 84 health outcomes were included in subsequent analyses exploring correlations with NOI and CPD variables. Table A1 lists the 84 health variables.

Selection of CPD Variables

The 2015 CPD was first merged with the 2013 NHIS at the block group level, and then the tract if block group was missing on the NHIS. Thirty-five (35) of the Census 2010 or ACS 2009-2013 measures available on the CPD (or recoded from existing CPD variables) were selected for analysis. Within each of nine Census Divisions, the 35 measures were recoded into deciles, quintiles, quartiles, or tertiles depending on the observed distribution of the original measure. The list and description of the 35 measures can be found in Table A2.

Selection of NOI Variables

All 15 of the available NOI variables were included in the analysis (see Table A3).

Statistical Procedures

Each selected CPD and NOI variable was correlated with the full set of 84 health measures. Given that the variables in this analysis were nominal (most health outcomes were dichotomous) or ordinal,

Table B1. Health Variables Used in the Correlation Analysis

Person/Family-Level Variables
Person limited in the kind or amount of work he/she can do
Person receives special educational or early intervention services
Person needs help with routine needs
Person has difficulty walking without special equipment
Person has difficulty remembering because he/she experiences periods of confusion
Person needs help with personal care needs (ER measure)
Persons reported health status is excellent or very good (ER measure)
Person had a medically-attended injury or poisoning in the past 3 months
Person did not receive care due to cost in past 12 months (ER measure)
Person was hospitalized overnight in the past 12 months
Person received care at home in the last 2 weeks
Person received medical advice or test results care over the phone in the last 2 weeks
Person received care at a doctor's office, clinic, ER, or other place in last 2 weeks
Person received care 10 or more times in past 12 months
Person has health insurance coverage (ER measure)
Family is deemed to be food insecure
Anyone in family has a flexible spending account
Family has problems paying medical bills
Family out-of-pocket expenditures for medical care
Sample Child Variables
Sample child took medications for difficulties with emotions, concentration, or behaviors in past 6 months
Seen or talked to a specialist about sample child's health in past 12 months
Delayed getting care for sample child in past 12 months
Did not get care for sample child due to cost in past 12 months
Time since last saw or talked to a doctor
Sample child has many worries, or often seems worried
Sample child had hay fever in past 12 months
Told by doctor that sample child has any other developmental delay
Sample child has difficulties with emotions, concentration, behavior, etc.
Sample child had stomach or intestinal illness that started in past 2 weeks
Seen or talked to a nurse practitioner, physician assistant, or midwife about sample child's health in past 12 months
Sample child's hearing without a hearing aid
Does sample child have any trouble seeing
Sample child's health compared to 12 months ago
Sample child ever has chicken pox
Told that doctor's office/clinic would not accept sample child's health insurance coverage in past 12 months
Sample child has a regular source of care (ER measure)
Sample child received a flu shot in the past 12 months (ER measure)
Sample child had an asthma attack in the past 12 months (ER measure)
Sample child still has asthma (ER measure)

Time since last saw or talked to a dentist
Sample child had stuttering or stammering during the past 12 months
Sample Adult Variables
Sample adult ever told by doctor that he/she had coronary heart disease
Sample adult had an asthma attack in the past 12 months (ER measure)
Sample adult still has asthma (ER measure)
Sample adult ever been told by doctor he/she had an ulcer
Sample adult ever told by doctor that he/she had cancer
Sample adult ever told by doctor that he/she had cancer (ER measure)
Sample adult told by doctor he/she had hay fever in past 12 months
Sample adult told by doctor he/she had bronchitis in past 12 months
Sample adult had symptoms of joint pain in past 30 days
Sample adult had neck pain in past 3 months
Sample adult had head/chest cold in past 2 weeks
Sample adult had high cholesterol in the past 12 months
Sample adult's hearing without a hearing aid
Does sample adult have any trouble seeing
Sample adult lost all upper and lower natural teeth
Number of days illness/injury kept sample adult in bed in past 12 months
Sample adult's health compared to 1 year ago
Sample adult has movement difficulties
Sample adults has social limitations
Sample adult had 5 or more drinks in 1 day in the past 12 months (ER measure)
Sample adult is a current smoker (ER measure)
Sample adult met federal physical activity guidelines (ER measure)
Sample adult is obese (ER measure)
Sample adult worries over being able to pay medical bills
Sample adult ever been tested for HIV (ER measure)
Sample adult had serious psychological distress in past 30 days (ER measure)
Average number of hours of sleep sample adult gets
Sample adult has a regular source of care (ER measure)
Sample adult saw an eye doctor in the past 12 months
Sample adult saw a specialist in the past 12 months
Sample adult received home care in the past 12 months
Number of doctor visits in the past 12 months
Sample adult delayed care in the past 12 months
Sample adult did not receive care due to cost in the past 12 months
Sample adult was denied care in the past 12 months
Sample adult received a flu shot in the past 12 months (ER measure)
Sample child received a pneumococcal vaccination in the past 12 months (ER measure)
Sample adult ever had hepatitis
Sample adult ever traveled outside of the United States to countries other than Europe, Japan, Australia, New Zealand or Canada, since 1995

Sample adult volunteers or works in a health-care facility
Sample adult used health information technology in the past 12 months
Sample adult tried to purchase health insurance directly in past 3 years
Sample adult's health insurance coverage compared to 1 year ago

Table B2. List of Recoded 2015 Census Planning Database Measures Used in This Analysis

Prediction of low Census mail return rate
% of 2010 Census total population < than 5 years old
% of 2010 Census total population identifying as NH white
% of 2010 Census total population identifying as NH black or Hispanic
% of 2010 Census total population < than 18 years old
% of ACS pop. 5+ that speaks language other than English at home
% of ACS pop. 25+ that are not high school grads
% of ACS pop. 25+ that have a college degree or higher
% of 2010 Census occ. housing units where householder and spouse in same household
% of 2010 Census occ. housing units where householder lives alone
% of 2010 Census occ. housing units that were owner-occupied
% of ACS housing units in a multi-unit structure
Average aggregated household income of ACS occ. housing units
Average aggregated value for ACS occ. housing units
% of 2010 Census total population 65+
% of 2010 Census total population identifying as Hispanic
% of 2010 Census total population identifying as NH black
% of 2010 Census total population identifying as NH AIAN
% of 2010 Census total population identifying as NH Asian
% of ACS population classified as below the poverty line
% of ACS civilians 16+ that are unemployed
% of ACS population that is uninsured
% of 2010 Census occ. housing units where householder lives alone or with a non-relative
Average # of persons per 2010 Census occ. housing unit
% of ACS pop. 5+ that speaks Spanish/Spanish Creole at home
% of ACS population that was not born a U.S. citizen
% of ACS pop. that are not citizens of the U.S.
% of 2010 Census occ. housing units w/ female householder and no husband
% of 2010 Census family-occ. housing units with a related child under 6
% of ACS occ. housing units where the householder moved in 2010 or later
% of ACS occ. housing units that received public assistance
% of 2010 Census housing units with no reg. occupants on Census day
% of ACS occ. housing units with > 1.01 persons per room
% of ACS occ. housing units with no working telephone
% of 2010 Census occ. housing units with a child

Table B3. Neighborhood Observation Instrument (NOI) Variables

Variable	Question	Response Options
GRAFFITI	Did you observe graffiti or painted-over graffiti on buildings, sidewalks, walls, or signs in the block face of the sample unit or building within which the sample unit resides?	Yes No
ADDR_COND	How would you describe the condition of the sample unit or the building within which the sample unit resides?	Very poor Poor Fair Good Excellent
ACCESS	Based on your observation, does the sample unit or the building within which the sample unit resides have: ...a security buzzer, key code, doorman, or any other barrier that may prevent access (for example dogs, locked gate, etc.)?	Yes No
YARDS	Based on your observation, does the sample unit or the building within which the sample unit resides have: ...a well-tended yard or garden?	Yes No Unable to observe
WALLS	Based on your observation, does the sample unit or the building within which the sample unit resides have: ...peeling paint or damaged exterior walls?	Yes No
BARS	Based on your observation, does the sample unit or the building within which the sample unit resides have: ...window bars or grating on the doors or windows?	Yes No
LOCKS	If this is a multiunit structure, answer based on the sample unit, not the building within the sample unit resides. Based on your observation, does the <u>SAMPLE UNIT</u> have: ...3 or more door locks?	Yes No Unable to observe
CHILDREN	Based on your observation, does the <u>SAMPLE UNIT</u> have...indication that children under 6 (including babies) may live at the unit (visible toys, car seat, strollers, outdoor swing/play set for example)?	Yes No Unable to observe
WHEELCHAIR	If this is a multiunit structure, answer based on the sample unit, not the building within the sample unit resides. Based on your observation, does the <u>SAMPLE UNIT</u> have: ...a wheel chair ramp or other indicators that the residents of the sample unit are handicapped, disabled, or may have a chronic health condition (deaf, blind, use oxygen, etc.)?	Yes No Unable to observe
BICYCLE	If this is a multiunit structure, answer based on the sample unit, not the building within the sample unit resides. Based on your observation, does the <u>SAMPLE UNIT</u> have: ...an adult-sized bicycle?	Yes No Unable to observe
SMOKER	If this is a multiunit structure, answer based on the sample unit, not the building within the sample unit resides. Based on your observation, does the <u>SAMPLE UNIT</u> have: ...any indication that the residents of the sample unit are smokers (cigarette/cigar butts, ashtrays, smell smoke, etc.)?	Yes No Unable to observe

HHINC	Relative to the general population and based on your observations, would you judge this sample unit to have a household income: 1. In the bottom third of the population 2. In the middle third of the population 3. In the top third of the population	Bottom third Middle third Top third
EMPLOYED	Based on your observation, would you say at least one adult resident of the sample unit is employed? Yes or no.	Yes No
HHLANG	Based on your observation, would you say that the residents of the sample unit speak a language other than English?	Yes No
OVER65	How old would you estimate the residents of the sample unit to be? 1. All occupants under the age of 30; 2. All occupants over the age of 65; or 3. Other age composition.	The responses were recoded to all occupants over the age of 65 versus other.

polychoric/tetrachoric correlations were performed in SAS. To evaluate the relative strength of correlations, absolute values were taken. As a summary measure, the average absolute correlation between a CPD/NOI variable and the health variables was produced. For example, the ACS measure of uninsured from the CPD was correlated with each of 19 health variables from the person file. The absolute value of each of 19 correlations was taken and then summed. The total was divided by 19 to produce an average absolute correlation between the ACS uninsured measure and the 19 person-level health variables.

With regard to the NOI variables, past analyses of select observations revealed that the outcome of the attempt on which the observations were recorded was a strong predictor of measurement error [22]. Not surprisingly, if the observations were recorded after contact was made with a household, agreement between the observation and survey data increased. To guard against inflating the magnitude of the correlations between the NOI observations and health outcomes, the correlational analysis involving the NOI variables was limited to cases where the observations were recorded on an attempt coded as a noncontact. This most closely approximates the ideal data collection protocol in which the observations would be recorded prior to ever making a contact attempt on the household.

All correlations were produced using SAS PROC FREQ with the POLYCHORIC option. Note that the analyses used base weights but ignored the complex sampling design of the NHIS.

Results

Table A4 summarizes the results of all the correlations. Again, since we were more interested in summarizing the magnitude as opposed to the direction of the correlations, the table presents absolute correlations. As shown in Table A4, five NOI measures had the strongest average absolute correlation with the 84 health variables: whether or not the interviewer observed evidence that one or more residents are disabled, handicapped, or has a chronic health condition (WHEELCHAIR); interviewer assessment of whether a household's income falls in the bottom, middle, or top third of household incomes in the larger area (HHINC); whether or not the interviewer observed evidence that all residents of the household are over the age of 65 (OVER65); the interviewer's assessment of the physical condition of the address (ADDR_COND); and whether or not the interviewer assessed that one or more

adults at the residence are employed (EMPLOYED). CPD variables with the highest average absolute correlations with the 84 health measures included the percent of persons in the block group/tract with a college degree (PCT_COLLEGE_ACS); average aggregated household income of the block group/tract (AVG_HHINC_ACS); the average aggregated value of occupied housing units in the block group/tract (AVG_HOUSEVALUE_ACS); and the percent of adults 25 or older with less than a high school education (PCT_LTHS_ACS).

Given that the average absolute correlation from all correlations was .0670 (bottom row of Table B4), we selected variables with average absolute correlations of .0700 or greater as possible inputs to a balancing propensity model. This reduced the original list of 50 NOI/CPD variables to 19.

Table B4. Summary of Polychoric/Tetrachoric Correlations between 35 CPD Variables, 15 NOI Variables (noncontacts only), and 84 NHIS Person/Family, Sample Child, and Sample Adult Variables: NHIS, 2013 (base weights)

CPD/NOI Variable	Average Corr.	.1000- .1999	.2000- .2999	.3000- .3999	.4000 +	Strongest Corr.
wheelchair	0.1425	28	11	4	5	0.4919
hhinc	0.1202	29	13	2	0	0.3888
over65	0.1086	24	9	4	0	0.3838
addr_cond	0.1029	26	8	1	0	0.3245
employed	0.1025	24	7	3	0	0.3545
pct_college_acs	0.0983	25	9	0	0	0.2830
avg_hhinc_acs	0.0978	29	8	0	0	0.2988
avg_housevalue_acs	0.0908	28	6	0	0	0.2895
pct_lths_acs	0.0888	25	4	0	0	0.2652
hlang	0.0857	24	4	0	0	0.2461
pct_femnohusbhhd	0.0831	25	3	0	0	0.2908
yards	0.0824	19	5	0	0	0.2493
walls	0.0805	27	3	0	0	0.2372
pct_poverty_acs	0.0799	24	3	0	0	0.2742
smoker	0.0755	17	3	0	1	0.4189
pct_uninsured_acs	0.0736	22	2	0	0	0.2680
pct_vacant	0.0709	22	0	0	0	0.1911
pct_unemployed_acs	0.0705	18	1	0	0	0.2213
pct_mrdcplhhld	0.0704	17	2	0	0	0.2266
pct_nhasian	0.0681	19	1	0	0	0.2526
pct_ownerocc	0.0660	16	2	0	0	0.2264
pct_hispblack	0.0656	21	1	0	0	0.2048
pct_nhwhite	0.0650	20	1	0	0	0.2005
pct_65plus	0.0634	18	0	0	0	0.1781
pct_famwchildu6	0.0622	17	1	0	0	0.2140
graffiti	0.0607	12	0	0	0	0.1923
children	0.0603	18	0	0	0	0.1775
bars	0.0590	14	1	0	0	0.2061
pct_forborn_acs	0.0566	9	1	0	0	0.2124

pct_hispanic	0.0565	14	0	0	0	0.1911
pct_crowding_acs	0.0553	13	0	0	0	0.1946
pct_othlang_acs	0.0545	11	0	0	0	0.1598
pct_under5	0.0541	15	0	0	0	0.1808
pct_noncitizen_acs	0.0535	11	0	0	0	0.1570
pct_pubassist_acs	0.0532	9	0	0	0	0.1514
pct_spanish_acs	0.0498	10	0	0	0	0.1877
access	0.0494	10	0	0	0	0.1959
avg_personsinhhld	0.0487	3	0	0	0	0.1362
pct_nonfamhhld	0.0452	5	0	0	0	0.1375
bicycle	0.0439	6	0	0	0	0.1273
pct_nhblack	0.0431	6	0	0	0	0.1591
pct_children	0.0428	3	0	0	0	0.1177
pct_multiunit_acs	0.0425	6	0	0	0	0.1527
pct_singperhhld	0.0408	3	0	0	0	0.1302
locks	0.0402	5	0	0	0	0.1754
pct_hhldmovedin_acs	0.0377	4	0	0	0	0.1525
pct_hhldwchildren	0.0370	1	0	0	0	0.1033
pct_nophone_acs	0.0364	4	0	0	0	0.1124
pct_nhaian	0.0337	0	0	0	0	0.0838
TOTAL	0.0670	779	113	14	6	0.4919

Bold = NOI variable. Note that only the correlations with health variables were considered when the NOI observations were recorded during an attempt resulting in a noncontact.

Response Rates for Groups Defined by the 19 NOI/CPD Variables

Table B5 presents response rates by the NOI observations and the CPD variables selected for the balancing propensity model. For the NOI observations, response rates were available for 2013-2015. For the CPD measures, we present response rates for 2013 and 2014. As noted earlier, a goal of the experiment is to balance response across groups defined by the NOI observations and CPD variables included in the balancing propensity model. By that, we are trying to achieve similar response rates, not achieve similar numbers of cases. Take, for example, the smoking observation. For each household, interviewers are asked to assess whether they observed any evidence that household residents may be smokers. In 2013, 7.3% of in-scope households were observed to include smokers. The response rate for these households was 83.4%, while the response rate for the households where no evidence of smokers was observed was 75.6%. Hence, smoking households, as assessed by the interviewers, were over-represented in the sample. Knowing that this observation is also correlated with several health outcomes on the NHIS (see Appendix B), the potential for nonresponse bias increases. For example, from the correlation analysis, we know that the smoking observation is positively associated with the sample adult not receiving specific health care services due to cost. Since we are over-representing smoker households in our sample, our estimate of the percentage of adults who did not receive specific health care services due to cost in the past 12 months may be too high. Using this simple example, our goal via adaptive design would be to achieve similar **response rates** for the smoking and non-smoking (as defined by the interviewers) households to help protect against nonresponse bias.

There are clear differences in response rates for the majority of the 19 variables, with chi-square analysis revealing only two non-significant differences in response rate distributions: whether or not the interviewer observed a well-tended yard or garden in 2013, and whether or not the interviewer observed peeling paint or damaged walls in 2015. Some of the observed differences in response rates were sizeable. For example, the response rate for households where interviewers observed evidence that a resident may be disabled, handicapped, or have a chronic health condition was 80.1% in 2015. The response rate for households where this was not observed was 70.5%, representing a nearly 10 percentage point difference between the two groups. When looking at all three years of data, the gap in response rates between these two groups of households has been growing.

Focusing on other NOI observations, households where the interviewer observed evidence of a language other than English being spoken had higher response rates for all three years compared to households where this was not observed. Also, households assessed to have one or more working adults had lower response rates for all three years compared to households where this was not observed.

Among the CPD variables, it is clear that higher response rates are achieved among lower educated and less affluent households. For both 2013 and 2014, response rates were higher for households in the top decile, compared to households in the bottom decile, of the ACS measure of the percentage of adults aged 25 and older that are not high school graduates. In addition, households in the bottom decile of the ACS measure of average aggregated household income had higher response rates for both years compared to households in the top decile.

In sum, we observe consistent response rate differences across the set of variables (available for both responding and nonresponding households) found to be more highly-related to several NHIS health outcomes. Again, a primary goal of the adaptive design experiment will be to reduce the response rate differences across these variables in order to reduce the potential for or magnitude of nonresponse bias in critical health estimates.

Table B5. Response Rates (standard error) by Select NOI, Census 2010 (block group/tract), and ACS 2009-13 (block group/tract) Measures: NHIS, 2013-2015 (base weighted)

	2013	2014	2015
<i>NOI Measures</i>			
Evidence that household speaks a language other than English			
Yes	79.6 (0.25)	78.5 (0.25)	75.4 (0.34)
No	75.7 (0.18)	74.1 (0.22)	69.8 (0.32)
All household residents over the age of 65			
Yes	81.1 (0.35)	81.0 (0.39)	77.9 (0.39)
No	75.5 (0.18)	73.8 (0.21)	69.6 (0.31)
Household income relative to general population			
Bottom third	79.9 (0.30)	77.3 (0.34)	73.1 (0.48)
Middle third	75.0 (0.17)	73.8 (0.20)	69.5 (0.30)
Top third	73.7 (0.31)	73.0 (0.38)	69.5 (0.41)
At least one adult resident of the household is employed			
Yes	75.2 (0.18)	73.5 (0.22)	69.4 (0.32)
No	80.0 (0.24)	78.9 (0.28)	74.4 (0.37)
Sample unit has a wheel chair ramp or other indicators that residents are handicapped or disabled			
Yes	82.1 (0.56)	83.6 (0.56)	80.1 (0.73)
No/Unable to observe the sample unit	76.0 (0.17)	74.3 (0.20)	70.5 (0.30)
Any indication that the residents are smokers			
Yes	83.4 (0.43)	81.2 (0.55)	77.9 (0.52)
No/Unable to observe the sample unit	75.6 (0.17)	74.2 (0.19)	70.3 (0.31)
Sample unit has a well-tended yard or garden			
Yes	76.1 (0.19)	75.1 (0.22)	71.3 (0.27)
No/Not applicable	76.4 (0.24)	73.8 (0.29)	69.0 (0.45)
Damaged walls or peeling paint			
Yes	78.8 (0.33)	76.1 (0.40)	71.0 (0.61)
No	75.9 (0.17)	74.4 (0.20)	70.4 (0.29)

Table B5. continued

	2013	2014	2015
Physical condition of the sample unit			
Very poor/poor	78.8 (0.54)	76.7 (0.57)	70.6 (0.63)
Fair	77.1 (0.30)	74.8 (0.26)	69.7 (0.49)
Good	75.8 (0.20)	74.5 (0.25)	71.2 (0.30)
Very good	75.3 (0.28)	73.9 (0.32)	69.9 (0.34)
Census 2010 (block group/tract), and ACS 2009-13 (block group/tract) Measures			
% of ACS population 25+ that has a college degree (ACS 2009-13)			
Bottom decile	79.0 (0.54)	77.7 (0.46)	N/A
Top decile	72.8 (0.27)	72.7 (0.24)	N/A
% of 2010 Census occupied housing units with a female householder and no husband (Census 2010)			
Bottom quartile	73.8 (0.30)	73.5 (0.27)	N/A
Top quartile	77.3 (0.26)	75.2 (0.24)	N/A
Average aggregated household income of ACS occupied housing units (ACS 2009-13)			
Bottom decile	79.6 (0.47)	75.8 (0.48)	N/A
Top decile	71.0 (0.21)	71.9 (0.31)	N/A
% of 2010 Census housing units with no registered occupants on Census day (Census 2010)			
Bottom quartile	74.1 (0.20)	73.1 (0.25)	N/A
Top quartile	78.4 (0.31)	75.2 (0.37)	N/A
% of 2010 Census occupied housing units where householder and spouse in same household (Census 2010)			
Bottom decile	75.7 (0.37)	71.6 (0.32)	N/A
Top decile	73.9 (0.34)	73.1 (0.39)	N/A
Average aggregated value for ACS occupied housing units (ACS 2009-13)			
Bottom decile	78.5 (0.49)	75.4 (0.35)	N/A
Top decile	71.7 (0.25)	72.1 (0.28)	N/A

Table B5. continued

	2013	2014	2015
% of persons uninsured (ACS 2009-13)			
Bottom quintile	74.0 (0.25)	73.5 (0.26)	N/A
Top quintile	78.0 (0.31)	75.9 (0.41)	N/A
% of ACS population 25+ that are not high school grads (ACS 2009-13)			
Bottom decile	73.4 (0.28)	72.7 (0.29)	N/A
Top decile	78.9 (0.41)	77.6 (0.53)	N/A
% of ACS civilians 16+ that are unemployed (ACS 2009-13)			
Bottom quintile	74.3 (0.33)	73.0 (0.30)	N/A
Top quintile	77.8 (0.33)	75.7 (0.46)	N/A
% of persons in poverty (ACS 2009-13)			
Bottom quintile	73.8 (0.26)	73.2 (0.26)	N/A
Top quintile	78.2 (0.34)	75.7 (0.34)	N/A