Ignorability of BMI Non-Response in the Healthy Nevada Project Social Determinants of Health Survey

Kelly Wilkin

10/9/2022

Introduction

The Healthy Nevada Project (HNP) is a population health study. joint aims of providing free genetic screening to participants while providing researchers with a rich data set to analyze the social, environmental, and genetic causes of disease.

One unique feature of the study is its ability to re-contact consenting particpants with requests for additional information. The HNP uses follow-up surveys. Surveys can reveal important information about the social and economic conditions facing participants that would typically be unavailable through genetic testing results or the electronic medical record (EMR). survey has been used to study how participants with findings of positive results are shared with their doctors [**?**]. used to identify alleles associated with negative reactions to the covid vaccine [**?**]. used to show interactions between aces and a genetic variant on bmi [**?**].

Participation in the HNP and its subsequent follow-up surveys is voluntary. We know that participants skew toward a certain profile (white and female). However, not much else is known about selection. in the case of bmi and aces, it was noted that non-response to aces questions was low in the hnp survey. however, survey participation (unit response) was not universal among hnp participants. if participation in the survey is related to bmi, then estimated coefficients in an equation examining bmi *outcomes* will be biased; that is, non-response is not ignorable [**?**].

this paper examines whether hnp survey non-response is ignorable. link bmi measures from the emr to hnp participants. relate bmi to indicator of survey response. permits a direct test of non-response: regression of . if ignorable, expect to find no systematic relationship between bmi level and the probability of survey response. similar to test of earnings non-response bias in household surveys.

test by plotting survey response probabilities by quantiles of the bmi distribution. visually check for relationship. estimate model of individual response against measures of bmi while controlling for other observables.

find evidence of positive selection into survey response: those with higher bmi are more likely to respond. holds across sexes. women systematically more likely to respond than men at all bmi levels. however, general pattern is the same.

contribution and implications.

organization of paper.

Data

The data used for this study come primarily from two sources: the HNP SDH survey roster and a matched set of administrative records from the Renown electronic health record (EHR). Participant sex and ethnicity are derived from their sequenced exome, which is used to impute for missing values of these items. CGM staff conducts data linkage and assigns each HNP participant a "participant unique ID," or PUID, which links their records across the survey, genetic sequencing, and Renown EHR databases while maintaining compliance with the Health Insurance Portability and Accountability Act (HIPAA) of 1996.

Social Determinants of Health Survey Roster

The HNP SDH survey is administered by CGM staff using the Survey Monkey (SM) service. The SDH survey is sent to all HNP participants who consent to follow-up survey requests. It asks a total of X questions about demographics (e.g., age, sex, race and ethnicity), educational attainment, household income, as wells as a host of other items related to family health history, impulsive behaviors, adverse childhood experiences (ACEs), and smoking and drug histories. SDH survey requests are made via email, where participants are given a hyperlink to the SM survey. Participants who respond to the survey are entered into a raffle for a $250 gift card.

Survey requests remain open until participants either complete the survey or expressly opt out of it. HNP participants who have done neither may be receive additional requests (reminders) at the discretion of CGM staff. SM keeps a record, which we call the "roster," of those who were sent the survey and their current response status. CGM staffs retrieves, or "collects," new responses in batches. The roster used for this paper was collected as of August 21, 2022.[1]

Renown EHR

Summary Statistics

The following table shows summary statistics for the SDH survey roster by response status.

Results

Response rates vary by BMI. Response increase with BMI and this this holds across sex. Response rates for women are higher than that for men at all obesity classes and across quantiles of the BMI distribution.

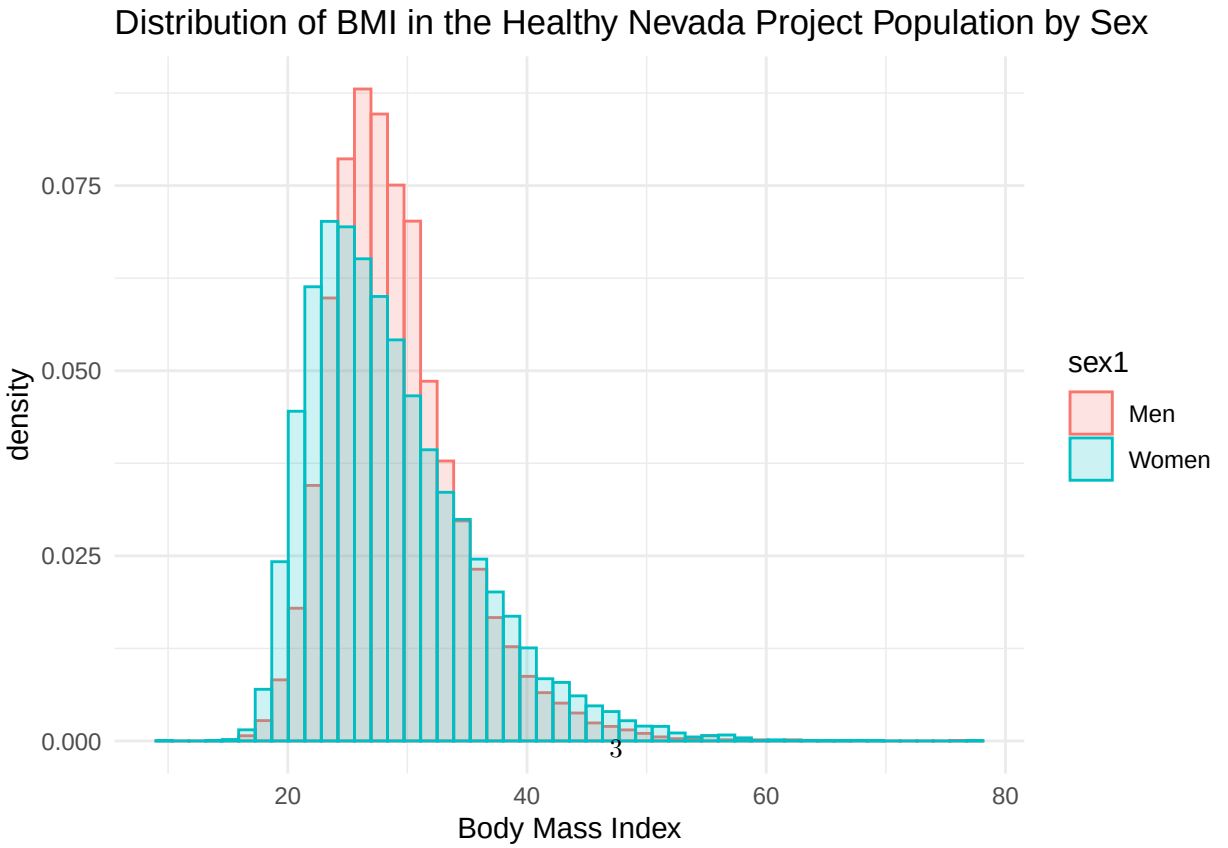
Start with an examination of the BMI distribution by sex. Figure shows a histogram of BMI by sex independent of response status.

[1]The SDH is currently undergoing a redesign and was last sent out on X date. Since then, X particpants have joined the HNP. These participants will receive the updated SDH following its approval by the DRI IRB.

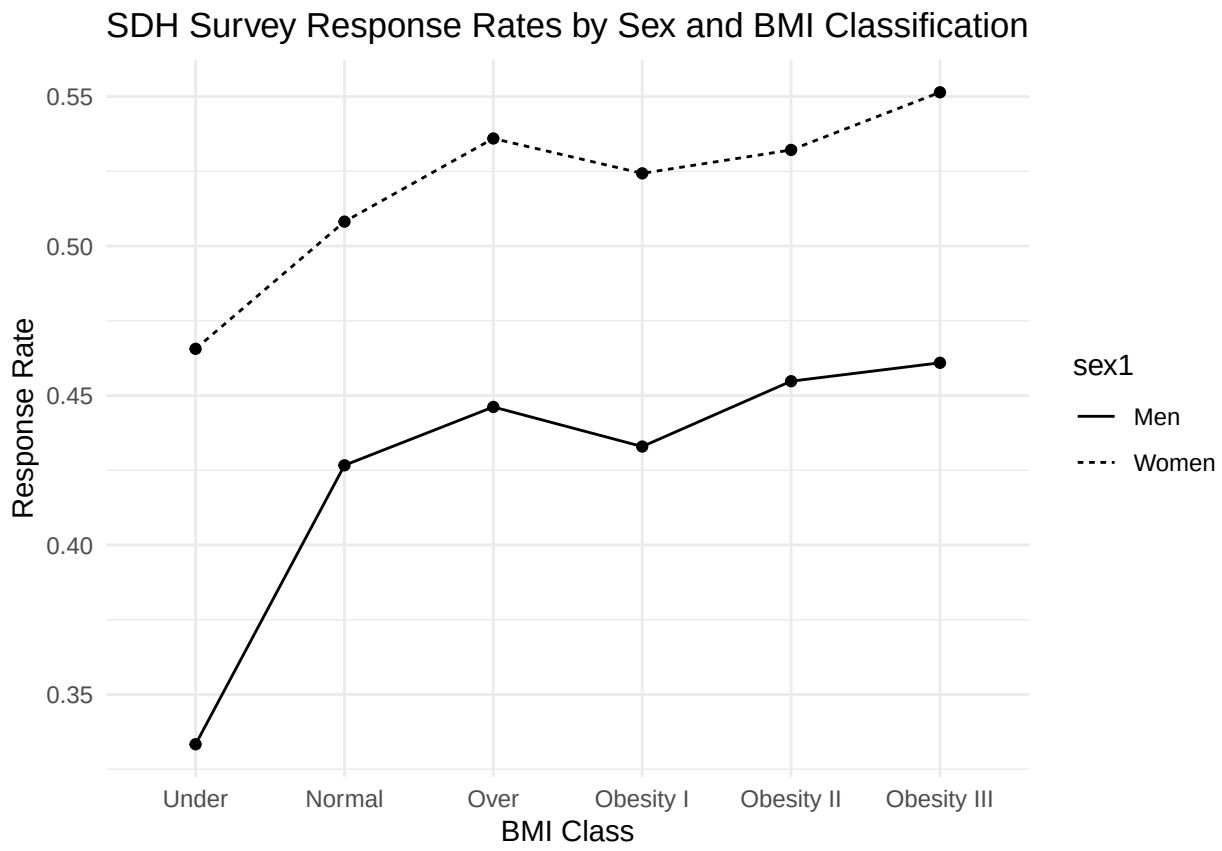
Table 1: Healthy Nevada Project Participant Characteristics by SDH Survey Response Status

		Response Status			
Variable	Overall	Responders	Non-Responders	p_value	numstar
Body Mass Index (kg/m2)	28.8	28.9	28.7	0.0010263	***
Underweight	0.008	0.007	0.009	0.1046277	
Normal weight	0.301	0.295	0.306	0.0594629	*
Overweight	0.337	0.339	0.335	0.4202914	
Obese	0.354	0.358	0.351	0.1885442	
Obesity I	0.202	0.199	0.205	0.2639397	
Obesity II	0.093	0.096	0.091	0.1593840	
Obesity III	0.059	0.063	0.055	0.0045900	***
Age (years)	49.6	51.3	48.1	0.0000000	***
Women	0.661	0.701	0.625	0.0000000	***
Men	0.339	0.299	0.375	0.0000000	***
White	0.876	0.898	0.854	0.0000000	***
Hispanic or Latino	0.021	0.015	0.028	0.0000000	***
Black	0.017	0.013	0.020	0.0000008	***
Asian	0.030	0.024	0.036	0.0000000	***
American Indian or Alaskan Native	0.009	0.008	0.009	0.5780340	
Hawaiian or other Pacific Islander	0.004	0.004	0.004	0.5732024	
RNO	0.868	0.888	0.849	0.0000000	***
LAS	0.132	0.112	0.151	0.0000000	***
Has Renown EHR	0.755	0.784	0.729	0.0000000	***
Observations	37,022	17,701	19,321	NA	NA

Note:
Author's calculations from the Social Determinants of Health (SDH) survey. Sample is Healthy Nevada Project participants who consented to and received a solicitation for the SDH survey. Survey information current as of August 21, 2022. p-values are from a two-sided t-test of equality of means between responders and non-responders. ***p<= 0.01, **p<=0.05, *p<=0.1.

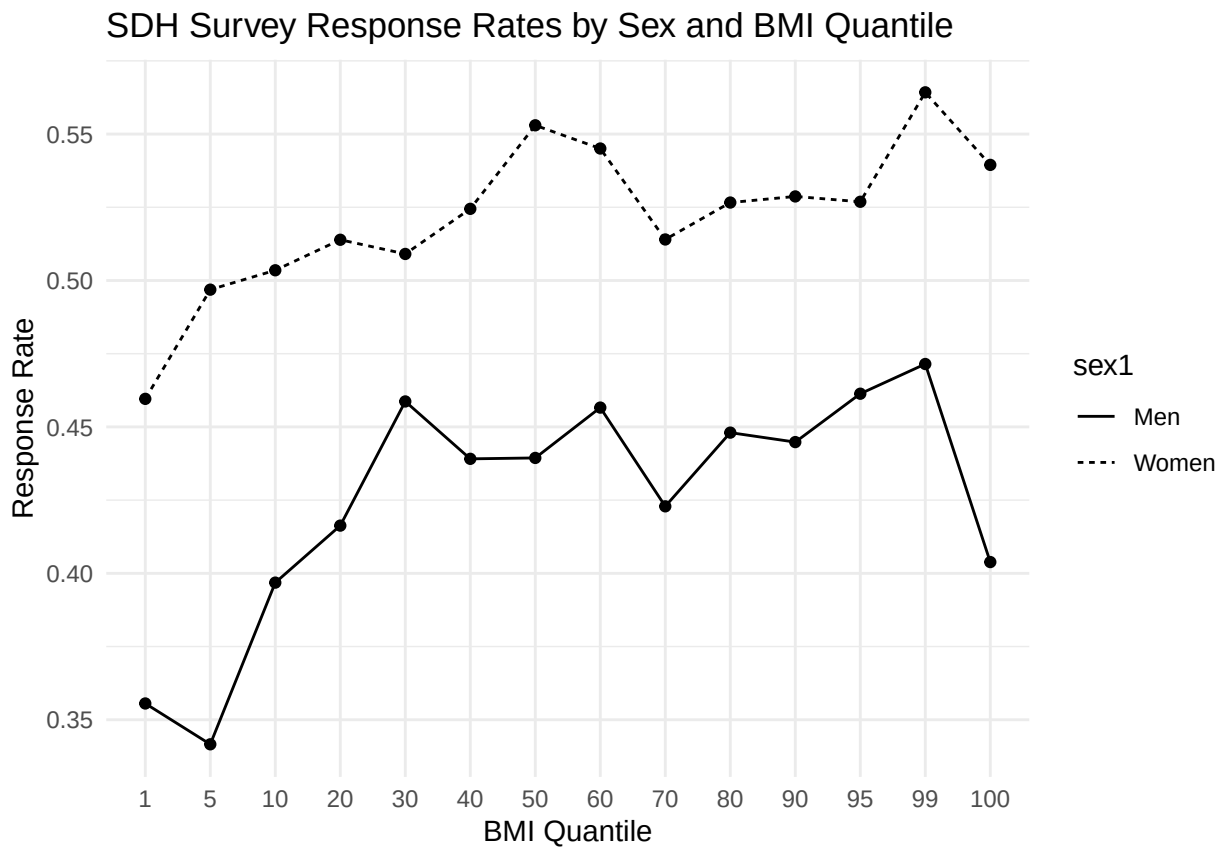


Next, examine response rates by sex across the standard classes commonly used to measure obesity. In addition to underweight, normal weight, and overweight, we further subdivide the obesity classes into obesity I, obesity II, and obesity III. Survey response rates increase with BMI, though at a decreasing rate, and the same pattern holds for both men and women.



We next examine non-response rates across more detailed quantiles of the BMI distribution, again by sex. The same general pattern holds: across all quantiles, response rates for women exceed those for men. Both relationships show significant non-linearity. For men, the survey response rate steadily increases from below 35 percent at the 5th percentile to over 45 percent at the 30th percentile. Their response rates stays relatively flat around 45 percent through the middle part of the distribution before rising again at the 90th through 99th percentiles, before it falls to about 40 percent for the highest BMI levels.

The survey response rate for women is about 47 percent in the first percentile and rises steadily to over 55 percent at the median. Like with men, the response rate dips between the 60th and 70th percentiles (though for women it never falls below 50 percent after the 5th percentile), and it also peaks in the 99th percentile (56 percent) before falling in the right tail of the BMI distribution.



Graphical analysis is informative for highlighting systematic differences in non-response patterns by BMI between men and women. However, it may mask other important differences that could also be related to response and BMI. We next specify a model to estimate the marginal effect of BMI on survey response propensity that allows us to control for observable differences in HNP participants.

We specify a binary outcome model of survey response r_i according to

$$r_i = \alpha + \gamma BMI_i + x_i'\beta + \epsilon_i,$$

where r_i is a binary indicator of survey response ($r_i = 1$ if respond, 0 otherwise), BMI_i is body mass index, x_i is a vector of observable participant characteristics, and ϵ_i is the error (assumed to be independent of x_i and the outcome). The parameter γ measures the relationship between BMI and the probability response.

We test for non-ignorable non-response bias by testing the hypothesis $H_0 : \gamma = 0$. If we reject the hypothesis, then we can conclude that there is non-ignorable response bias in BMI for the sample of SDH respondents, while failing to reject would lead us to conclude there is no relationship. We test this hypothesis separately for men and women.

Conclusion

High survey non-response rates create a condition where non-response bias may be a concern. In this paper we examined the incidence of non-ignorable survey response to the SDH survey, which is provides critical information about the social and economic condition of participants in the Healthy Nevada Project (HNP). Non-response bias may be a particular concern of the HNP since participation is purely voluntary.

A key feature of the HNP study design is the linkage of participant information across three critical data sources: the SDH survey, their sequenced exome, and administrative health records from a large regional

hospital. The EHR allows us to link non-self-reported measures of height and weight (the inputs to BMI calculations) for both survey respondents and non-respondents. This permits a direct test of response rates at different levels of BMI.

We find evidence of positive selection into survey response by BMI. Selection is non-linear (quadratic). This pattern of selection holds across sex. Women are more likely to respond to the survey than men across all levels of the BMI distribution.

Discussion

BMI is measured as patient averages. Ideally, would have participant's BMI at the time they took the survey. Instead, we use patient-level averages of non-pregnancy, non-hospice, and non-telemedicine encounters at any time in the EHR. While we could identify such encounters in the same year that they registered their Helix DNA sample kit for almost half of the sample, the remainder had encounters in either years prior to the kit registration date, after, or both. We tried a specification that used BMI levels for the same year when avaialable and imputed patient averages for those that did not. We found no difference in the results. I can think of two alternative approaches that might be worth a try: (a) estimate a model of BMI based on patient characteristics using the full EHR and imputing for the time period where it is missing in the same year for some, and (b) implement a weighting scheme that uses, say, an exponential decay function to apply less weight to more distant BMI measures.

About 25 perent of the survey roster does not have an EHR record. (For now, only use Renown. Will add UMC, which could pick up some LAS region respondents.) Some participants may not have had no contact with Renown or UMC, or perhaps that use another provide (Renown has a large market share but it's not complete). We need to use predictors from the SDH and other variables that can be constructed from it (and the exome) to model the probability of having an EHR record. Can use to construct inverse probability weights (IPWs). Re-estimate our models to see if the results still hold.

The HNP is an all-volunteer sample. While we have addressed issues of self-selection based on BMI, it is possible that the HNP sample does not generalize to the region or the rest of the state. More research is needed to examine the incidence of volunteer bias in the sample.