COVID-19 Vaccination Coverage in the U.S.: Comparison among National Probability-Based Surveys – National Immunization Survey Adult COVID Module, Household Pulse Survey, and the AmeriSpeak® Panel

Vicki J. Pineau[1], Jason Fields[2], James A. Singleton[3], Elizabeth Allen[1], David Dutwin, Sarah Kornylo[1], Xiuli Tang[1], Michael Chen[3], and Benjamin Fredua[3]
[1]NORC at the University of Chicago, 55 East Monroe Street, Chicago IL 60603
[2]U.S. Bureau of the Census, 4600 Silver Hill Rd, Suitland-Silver Hill, MD 20746
[3]National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329

Abstract
This paper presents results from exploratory research into differences in survey estimates of the proportion of adults reporting receipt of ≥1 dose of a COVID-19 vaccine in the U.S. reported weekly or biweekly by three national probability-based surveys during 2021: National Immunization Survey Adult COVID Module (NIS-ACM) since April 22, 2021, the Household Pulse Survey (HPS) since January 2021, and questions included in the AmeriSpeak Panel biweekly Omnibus sample since February 2021. Though each survey is probability-based in design, each has unique features (e.g., sample frames, survey modes) that are in some respects expected to produce somewhat different estimates of COVID-19 vaccination coverage in the U.S; while other features, such as question wording, would be expected to produce similar estimates to each other and to official vaccination administration data available in the CDC COVID-19 Data Tracker[1]. We first review the sample coverage, sample design, survey mode, fielding procedures, question wording/instructions, and sample weighting across the three surveys. We then explore estimated vaccination coverage for subpopulations across the three surveys to CDC's administrative data for receipt of ≥1 dose of a COVID-19 vaccine for subpopulations. Comparisons are made for population subgroups (age, Hispanic ethnicity/race, sex) and at the state level when possible. Some findings as to the potential impact of differing sample designs, unit nonresponse, survey mode, and estimation methods on the differences observed between the reported vaccination coverage estimates from the three surveys and the COVID-19 vaccine administration data are discussed.

Key Words: COVID-19 Vaccination Coverage; CDC COVID Data Tracker; Nonsampling Error

The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

[1] https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-people-additional-dose-totalpop

1. Introduction

In December 2019, a cluster of patients, including healthcare workers, in Wuhan, Hubei Province, China experiencing symptoms of an atypical pneumonia-like illness were reported. (Centers for Disease Control and Prevention, 2022) (A Cluster of Health Care Workers with COVID-19 Pneumonia Caused by SARS-CoV-2, 2021) In January 2020, the U.S. declared COVID a public health emergency and allocated resources to develop countermeasures, including the development of vaccines. (Administration for Strategic Preparedness & Response, July) At the end of 2020 and into the first half of 2021, COVID-19 vaccines were approved for emergency use and made available to all adults aged 18 years and older.

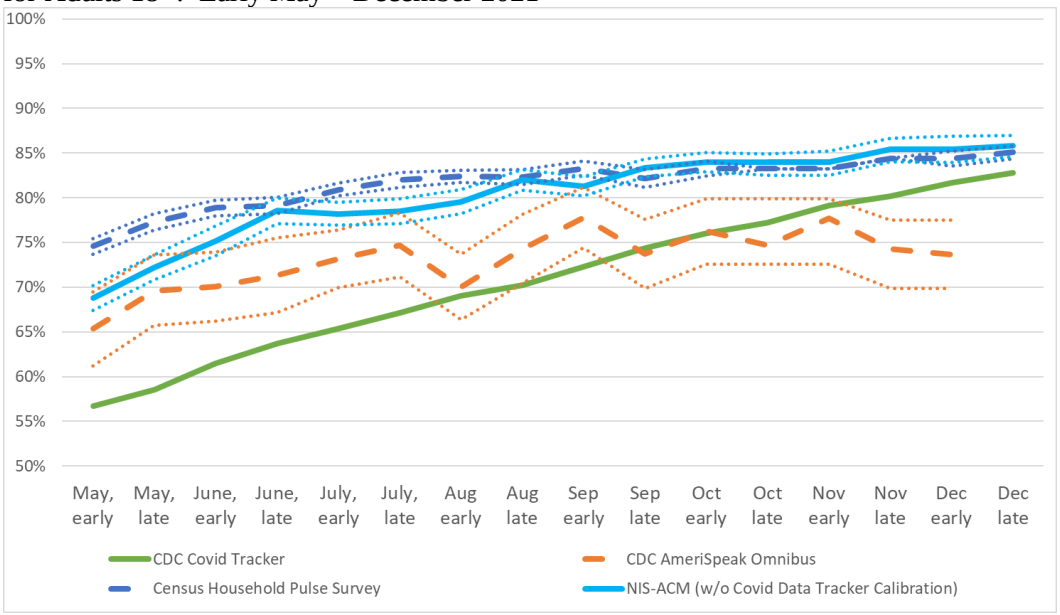
Beginning April 2020, CDC sponsored the fielding of various probability-based surveys to gauge various aspects of the pandemic on the U.S. population, including the socio-economic impact of the pandemic, intent of the population to get vaccinated, and to conduct public health monitoring of vaccination coverage across socio-demographic subgroups commensurate with the approval of COVID-19 vaccinations for the adult population. CDC first chose to include health related questions in the U.S. Census Bureau's on-going Household Pulse Survey (HPS) beginning April 23, 2020, for regular weekly/monthly assessment of the impact of the pandemic on employment status, consumer spending, food security, housing, education disruptions, and physical and mental wellness. (Fields, et al., Forthcoming) (U.S. Bureau of the Census, 2022) CDC additionally chose to include questions about perceptions toward COVID-19 vaccines and intentions to be vaccinated in the Omnibus Survey product of NORC at the University of Chicago's probability based survey panel, AmeriSpeak, for bi-monthly monitoring beginning December 2020. (Nguyen, et al., 2021) (Baack, et al., 2021) In the spring of 2021, the Centers for Disease Control and Prevention (CDC) leveraged the on-going National Immunization Survey (NIS) sample and its household screening methods to conduct an Adult COVID Module (NIS-ACM) (Centers for Disease Control and Prevention, 2022) survey for weekly monitoring of vaccination coverage, barriers to vaccination, vaccine hesitancy, and social attitudes and behaviors during the COVID-19 Pandemic.

The timing and the purposes of the three surveys sponsored by CDC to monitor the impact of the COVID-19 Pandemic were all different. Notably, while there were differences in design features across the studies, they were considered to all reasonably support the purposes/goals of each survey. All three survey samples are probability based, though each is likely subject to different types and levels of nonsampling error. As such, the goal of the work reported in this paper was to understand the differences among the three surveys by socio-demographics and investigate the possible survey features such as coverage, nonresponse, and reporting error that may be linked to any differences in sample makeup and survey point estimates.

Graph 1 presents weighted estimates of vaccination coverage for receipt of 1 or more doses of COVID-19 vaccination (1+COVID-19) from early May 2021 through late December 2021 for the three surveys as well as 1+COVID-19 vaccination coverage based on vaccination administration data reported in the CDC COVID Data Tracker. Though subject to its on limitations as described below in section 2.1, CDC's vaccine administrative data is utilized as the reference estimate in comparison to coverage estimates from the three surveys. The weighted vaccination coverage estimates for each survey were constructed using weights that include traditional weighting methods accounting for each survey's sample design and differential nonresponse. The weights for each survey also included a

post-stratification weighting adjustment to known demographic benchmarks to correct for known or expected differences between the sample and the target population. For NIS-ACM, the estimates in the graph are based on these traditionally weighted methods for comparability, but official NIS-ACM estimates on COVIDVaxView[2] utilize weights that include a calibration to vaccination totals from vaccine administration data. In early May 2021, all three surveys overestimated 1+COVID-19 vaccination coverage relative to CDC's vaccine administration data using traditional post-stratification weights. The absolute differences of the HPS and NIS-ACM survey estimates to the vaccine administration data diminished over the period, with estimates converging with each other and with the vaccination administration data. AmeriSpeak Omnibus estimates were closer to the vaccine administration data for most of the time period, overestimating vaccination coverage prior to early September 2021, and underestimating vaccination coverage relative to the vaccine administrative data after September 2021.

Graph 1. 1+COVID-19 Vaccination Coverage and Associated 95% Confidence Intervals for Adults 18+: Early May – December 2021*



*Weighted Estimates using demographically post-stratified weights for each survey. Estimates reported in the presentation may differ somewhat from officially reported estimates for various reasons. For the NIS-ACM, estimates published by CDC also include a weighting calibration to the COVID-19 vaccine administration data included as population control totals.

Differences in reported COVID-19 vaccination coverage among several national surveys that include both probability-based and nonprobability-based surveys have been documented by others (Bradley, et al., 2021), with a focus on use of the data defect correlation metric, the correlation between the outcome of interest and selection into the sample, to quantify the effects of sample bias on survey estimates. However, as noted in a response to Bradley, et al. by Reinhart and Tibshirani (Reinhart & Tibshirani, 2021), use of the data defect correlation metric to assess the quality of a survey estimate does not take into account any measurement error (e.g., response error, interviewer error, question wording) which may be a contributing factor to differences observed between survey estimates of COVID-19 vaccination coverage and vaccination coverage as reported in CDC's vaccine administration data. Thus, our approach to investigating the differences in

[2] https://www.cdc.gov/vaccines/imz-managers/coverage/covidvaxview/index.html

vaccination coverage estimates among the HPS, NIS-ACM, and AmeriSpeak Omnibus is one that follows the tenets of separating out the components of sampling and nonsampling errors towards an understanding and eventual aggregation of the errors in the near future. One area of concern in estimating COVID-19 vaccination receipt in the three surveys sponsored by CDC and the Census Bureau is whether the pandemic environment precipitated inflated self-reporting of vaccination receipt due to respondents' possible tendency to provide perceived desirable responses.

In this paper, we present some preliminary findings from exploratory analyses of the differences in reported 2021 1+COVID-19 vaccination coverage estimates observed among the three surveys for the U.S. adult population. It is important to note that estimated vaccination coverage reported here for the HPS, NIS-ACM, and AmeriSpeak Omnibus may differ somewhat from officially reported estimates for various reasons.

2. Methods

For each of the three surveys, we first catalogued the survey design features to better understand the differences that could lead to different types and levels of sampling and nonsampling errors. We examined COVID-19 vaccination coverage estimates from the HPS, the AmeriSpeak Omnibus and the NIS-ACM from May through December 2021. In any of the comparisons of weighted estimates presented below, NIS-ACM estimates were not constructed using weights that include calibration to vaccine administrative data. We compared vaccination coverage estimates across the three surveys and to published estimates by age and sex from COVID-19 vaccination administration data reported to CDC and shown on CDC's COVID Data Tracker[3]; these data are considered the best source of COVID-19 vaccination coverage available. The final source of data used in our analysis is the American Community Survey (ACS) from which we estimated benchmark distributions for various socio-demographic groups to assess sample representativeness of each of the three surveys. The data sources are described below in section 2.1.

We employed a z-test for proportions at the α=0.05 significance level to identify all statistically significant differences between survey estimates from each of the three COVID-19 surveys (HPS, NIS-ACM and Amerispeak Omnibus) to each other; or when comparisons of survey estimates from each of the three surveys are compared to ACS estimates. No statistical testing of differences was conducted for comparisons of survey estimates from each of the three COVID-19 surveys to vaccination coverage from COVID Data Tracker as it is administrative records data.

Survey estimates and variances were constructed using SAS® Version 9.4 software for the HPS, NIS-ACM, AmeriSpeak Omnibus, American Community Survey estimates.

2.1 Data Sources

2.1.1 Household Pulse Survey (HPS)

The HPS is a U.S. Bureau of the Census authorized collection under Title 13 United States Code, Sections 8(b), 182, and 193. The HPS is an address-based sample using the Census Bureau's Master Address File (MAF) as a sampling frame and is designed for estimation

[3] COVID Data Tracker. Atlanta, GA: US Department of Health and Human Services, CDC; 2022, July 26. https://covid.cdc.gov/covid-data-tracker

at the state, D.C. and top 15 MSA levels. [4] Table 1 lists the main sample design features of the 2021 HPS. Emails and phone numbers are matched to the selected sample resulting in approximately 86% of the MAF sample with a matched email, cell phone or both. The mode of interview is web only. Sampled households were contacted by both email if an email was available, and by text if a phone number was available. Multiple follow up contact attempts (using up to five associated email addresses and up to five cell phone numbers) are made to nonrespondents by email and SMS over the approximate 12-day data collection period for a maximum of 18 contact attempts. The sample size is approximately 1 million addresses on average for each bi-weekly reporting period in 2021. The average weighted response rate during all cycles of data collection was 5.9%. The question asking the respondent about whether they received a COVID-19 vaccination is worded as follows:

- Have you received a COVID-19 vaccine? Yes, No

2.1.2 AmeriSpeak Omnibus COVID-19 Module

The AmeriSpeak Omnibus used for the COVID-19 Module is an address-based sample panel selected from NORC's 2010 national frame and is designed for national level estimation. Table 1 lists the main sample design features of the 2021 AmeriSpeak Omnibus. The mode of interview is primarily Web, with an optional phone mode. Sampled panelists are contacted via email or by telephone or both, depending on their stated preferences as AmeriSpeak panel members. Panelists are not typically reminded to complete the Omnibus survey via email or phone since the fielding time is over a one-week period. The sample size is approximately 1,000 twice monthly. The average AAPOR response rate for the 2021 Omnibus samples was 2.6%. The question asking the respondent about whether they received a COVID-19 vaccination is worded as follows:

- (April 2021 and later) Have you received at least one dose of a COVID-19 vaccine? Yes, No, Don't Know, Skipped

2.1.3 National Immunization Survey Adult COVID Module (NIS-ACM)

The NIS-ACM is an RDD Cell Telephone sample designed for reporting vaccination rates for states, D.C., Puerto Rico, USVI, Guam, and selected local areas. Table 1 lists out the main sample design features of the 2021 NIS-ACM. It is a cell telephone mode only survey and the sample size is approximately 15,000 weekly. This activity was reviewed by CDC and was conducted consistent with applicable federal law and CDC policy (e.g., 45 C.F.R. part 46.102(l)(2), 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq.). The American Association for Public Opinion Research Response Rate 3 (AAPOR RR3) household response rate was 22.8% as of December 29, 2021.

The question asking the respondent about whether they received a COVID-19 vaccination is worded as follows:

- Have you received at least one dose of a COVID-19 vaccine? Yes, No, Don't Know, Refused

2.1.4 CDC COVID Data Tracker Vaccination Administration Data

CDC's COVID Data Tracker includes vaccination reports from all vaccine partners including jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities. CDC estimates the number of

[4] https://www2.census.gov/programs-surveys/demo/technical-documentation/hhp/Phase3-2_Source_and_Accuracy_Week%2036.pdf

people with at least one dose, fully vaccinated, and with a booster dose in the COVID Data Tracker. CDC estimates are based on data that includes a dose number. Some important limitations of the tracker data as described by CDC[5] are listed below:

- As of August 9, 2021, all vaccine reporting entities have the ability to update or delete their previously submitted records.
- The dose number reported to CDC may be incorrect because the data reported to CDC by vaccine partners does not have personally identifying information (PII).
- CDC may not be able to link multiple unique person identifiers for different jurisdictions or providers to a single person. For example, a person's booster dose may appear to be a person's first dose when reported.
- CDC has capped the percent of vaccination coverage metrics at 95%. This cap helps address potential overestimates of vaccination coverage due to 1st, 2nd and booster doses that were not linked for the same person.
- Vaccination reporting to CDC may suffer from administrative delays and CDC may incur some delay in synthesizing the data from reporting entities.

2.1.5 American Community Survey (ACS)

Socio-demographic distributions were constructed using the 2015-2019 five- year ACS Public Use Microdata Sample (PUMS) data for the adult population age 18 years and older and used as benchmarks in this paper because of the high level of accuracy and reliability in ACS socio-demographic survey estimates.[6] Distributions for the following variables were constructed for comparisons against the sample demographic distributions of the HPS, AmeriSpeak Omnibus and NIS-ACM.

2.2 Differences in Survey Design Features

Table 1 presents the survey design features of each of the three surveys. There are several notable differences among the survey features that may influence the differences observed in 1+COVID-19 vaccination coverage reported by each survey. We focused the exploratory analysis presented in this paper into differences in three survey design feature areas across the three surveys that may influence the differences observed in the vaccination coverage: (1) target population coverage, (2) differential response, and (3) differences in wording of the question that asks the respondent if they were vaccinated with at least one dose of a COVID-19 vaccine.

In regard to differences in target population coverage, the NIS-ACM sampling frame is a cellular phone Random Digit Dial (RDD) frame and thus excludes the population not reachable by cellular phone, namely those only accessible via landlines telephones and/or no phones. The loss in population coverage for adults 18+ living in landline only or phoneless households is estimated to be on the order of 2.3% during July-December 2021 (Blumberg & Luke, 2022). The HPS sample is built on the Census Master Address; contact information (cell phone and email address) is available for approximately 86% of the sample and this subsample with contact information comprises the fielded sample. The AmeriSpeak sample frame is NORC's National Frame for which coverage of the U.S. population has been estimated to be approximately 97%.[7]. For the NIS-ACM and the AmeriSpeak Omnibus panel, the degree of estimated noncoverage error is unlikely to be a

[5] https://www.cdc.gov/coronavirus/2019-ncov/vaccines/reporting-vaccinations.html#:~:text=CDC%20may%20not%20be%20able,and%20under%2Destimate%20subsequent%20doses.

[6] https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/

[7] https://amerispeak.norc.org/us/en/amerispeak/about-amerispeak/panel-design.html

large source of error affecting estimates of self-reported COVID-19 vaccination; noncoverage error has the potential to impact HPS estimates to a larger degree.

Response rates for the late December 2021 NIS-ACM, late December 2021 HPS, and Oct-Dec AmeriSpeak Omnibus surveys were 22.8%, 6.4%, and 2.6%. Thus, there is nonignorable potential for nonresponse bias in estimates of vaccination coverage for each of the three surveys to the extent nonrespondents were different in getting and/or reporting their COVID-19 vaccination status than respondents. For insight into potential error in COVID-19 vaccination status due to nonresponse in the NIS-ACM, we can look to a very similar NIS sister survey -- the NIS-FLU for the 2019-2020 influenza season -- which captured 2019-2020 seasonal influenza vaccination status for children and teens. The 2021 NIS-ACM and 2020 NIS-Flu both utilize the NIS-Child RDD sampling frame and are fielded using the same methods and thus have similar levels of nonresponse. It is also assumed that both have similar potential for nonresponse bias for "seasonal" vaccination coverage estimates. Nonresponse bias for flu vaccination coverage in the 2019-2020 NIS-Flu was estimated to be on the order of a 1.36% overestimate with a confidence interval of [-7.14%, 10.08%] (NORC at the University of Chicago, unpublished data, 2020). Nonresponse bias in 1+COVID-19 vaccination coverage in the NIS-ACM may be of the same order but requires further investigation.

The third area of focus for exploratory analysis is whether question wording for receipt of COVID-19 vaccinations results in misreporting vaccination status by respondents. The question wording is the same for the NIS-ACM and AmeriSpeak Omnibus for the May-December 2021 time period: "Have you received at least one dose of a COVID-19 vaccine?". However, the HPS question is somewhat different: "Have you received a COVID-19 vaccine? ". We hypothesize that response error, due to question wording or to respondents' desire to provide socially/politically acceptable responses to COVID-19 vaccination status questions, could be a significant contributing factor in differences observed between the HPS and NIS-ACM vaccination coverage and the COVID-19 vaccination administration data on COVID Data Tracker.

Table 1. Survey Design Features for NIS-ACM, HPS, and the AmeriSpeak Omnibus COVID Module, 2021

SURVEY DESIGN FEATURE	NATIONAL IMMUNIZATION SURVEY ADULT COVID MODULE (NIS-ACM)	CENSUS HOUSEHOLD PULSE SURVEY (HPS)	AMERISPEAK OMNIBUS
SPONSOR	**CDC**	**U.S. Bureau of the Census**	**CDC**
TARGET POPULATION	18+ Adults living in the U.S. PR, VI and Guam. Reporting at the national, state, D.C., and selected local levels	18+ Adult household population in the U.S. Reporting at the national, state, DC, and 15 largest MSAs	18+ Adult U.S. household population. Reporting at the national level

SAMPLING FRAME	Cell-phone Random Digit Dialing (RDD) using Marketing Systems Group (MSG) sampling frames	Census Master Address File (MAF) and Census Contact Frame	2010 NORC National Sample Frame
OVERSAMPLING	Immunization Information Systems (IIS) telephone numbers associated with age-eligible children in the state or local IIS are linked to RDD telephone numbers indicating presence of child in HH and used to stratify the sample and oversample HHs expected to have children or teens. *Note: No personal or confidential data is shared by the IIS under this oversampling methodology.*	Sixty-Six independent sampling areas were defined, and sampling rates determined by a desired CV of 3% on an estimate of 40%. The 11 smallest states had a higher CV threshold	No oversampling in Omnibus (higher percent of young adults and minorities oversampled in panel)
SAMPLING METHOD	SRS within estimation areas; vaccination administration-based stratification and oversampling of households with children in ~28 states.	SRS within estimation areas	SRS with strata defined by crossing Race/Hispanic ethnicity, age, education and gender
RECRUITMENT MODE	Telephone	Email and SMS text invitation to online survey	Mailings, phone, incentives, nonresponse follow-up with enhanced incentives and in-person follow-up for AmeriSpeak Panel; Web/Phone for Omnibus survey.

WITHIN-HOUSEHOLD SELECTION OF ADULT	Parent/Guardian if responding adult completed an earlier NIS module for an age-eligible child ≥ 6 months – 17 years in the household; Responding adult otherwise	Contact frame information is person level. The responding adult is the selected representative	Sample one panelist per household for each Omnibus wave
INTERVIEW MODE	Telephone	Self-administered online interview through Qualtrics	Self-administered web (CAWI) and phone (CATI).
FIELDING INTERVAL (WAVE)	Continuous; Weekly Estimates	Varied - initially weekly, then bi-weekly, and now once per month. Collection periods have gone from 6-day periods to 13-day collection windows	Twice per month
AVERAGE COMPLETED SAMPLE PER WAVE	~15,000	Varied from 55,000 to 130,000 with an average of ~79,600 over 42 cycles	~1000
RECEIPT OF 1+COVID-19 VACCINATION QUESTION	Have you received at least one dose of a COVID-19 vaccine? Yes, No, Don't Know, Refused	Have you received a COVID-19 vaccine? Yes, No	April 2021 and later: Have you received at least one dose of a COVID-19 vaccine? Yes, No, Don't Know, Refused, Skipped
POST-STRATIFICATION AND/OR CALIBRATION WEIGHTING VARIABLES	**ACS:** Census Region, State and Selected Local Areas, Sex, Metro Status, Education, Race/Ethnicity, Age; **NHIS**: HH cell-phone status; **CDC**: Vaccine Administration Totals for 1+COVID-19 vaccination status for gender by Age group	U.S. Census Population Controls based on the latest Vintage 2019 population estimates updated each year in HPS: U.S. Total, State, Hispanic origin and Race, Sex ACS: Education	Age, Gender, Census Division, Race/Hispanic Ethnicity, Education, Age x Gender, Age x Race/Ethnicity, Race/Ethnicity x Gender

EXTERNAL DATA SOURCES FOR WEIGHTING	Census 2019 1-year ACS; NCHS NHIS; CDC	Census Population Controls from the Vintage 2019 series	Current Population Survey (CPS) socio-demographic benchmarks developed by the U.S. Bureau of the Census
GEOGRAPHIC LOCATION	Respondent report of zip code	Sampled Master Address File identification number (MAFID), updated by respondents to current state for movers in weighting.	ABS address, updated when respondents move
RESPONSE RATE	Council of American Survey Research Organizations (CASRO) Response Rate (12/29/22): 22.85%	Average Weighted Response weight: 5.9% for all cycles; 6.4% since shift to 13-day collection cycles. Full and Sufficient Partial interviews divided by sampled households.	Average Weighted AAPOR RR3 Household Cumulative Response Rate (Oct – Dec 2022): 2.6%
QUESTIONNAIRE/ MODULE LENGTH	~70 questions; ~9.6 minutes	~78 questions; ~20-22 minutes for phase 3.6	20 questions in module
TIME INTO INTERVIEW WHEN 1+COVID-19 QUESTION IS ASKED	First question in module	Item 17; first item following demographics	Within first 5 questions of module
CALCULATION OF 1+COVID-19 VACCINATION COVERAGE RATE (NUMERATOR/DENOMINATOR)	**Numerator:** Have you received at least one dose of a COVID-19 vaccine? All completes with a response of Yes or No (Excludes DK and Refused); **Denominator**: All completes	**Numerator:** Have you received at least one dose of a COVID-19 vaccine? All completes with a response of Yes or No (Excludes DK and Refused); **Denominator**: All completes	**Numerator:** Have you received at least one dose of a COVID-19 vaccine? All completes with a response of Yes or No (Excludes DK and Refused, Skipped); **Denominator**: All completes

Abbreviations: ***Puerto Rico (PR), Virgin Islands (VI), District of Columbia (DC), Metropolitan Statistical Area (MSA), Random Digit Dialing (RDD), Marketing Systems Group (MSG), Census Master Address File (MAF), Immunization Information Systems (IIS), Coefficient of Variation (CV), Simple Random Sampling (SRS), Short Message Service (SMS), Computer Assisted Web Interviewing (CAWI), Computer Assisted Telephone Interviewing (CATI), Household (HH), Current Population Survey (CPS), Address Based Sampling (ABS), Master Address File Identification Number (MAFID), Council on American Survey Research Organizations (CASRO), American Association Public Opinion Research Response Rate 3 (AAPOR RR3), Don't Know (DK).***

3. Results

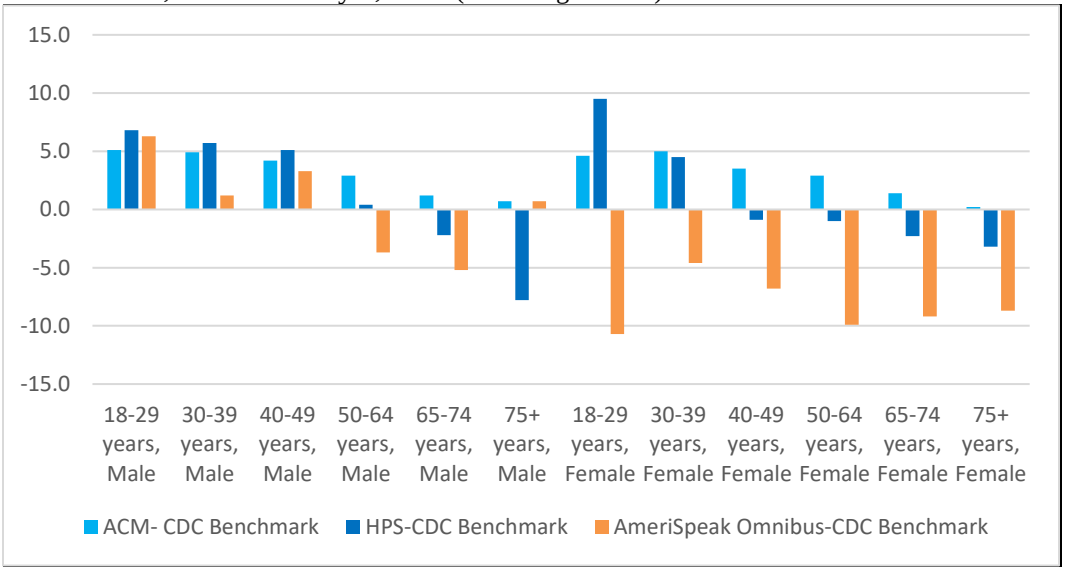
In section 3.1 below, we present differences in 1+COVID-19 vaccination coverage by age and gender for the NIS-ACM, HPS, and AmeriSpeak Omnibus compared to the CDC COVID Data Tracker vaccine administration data to investigate whether differences were greater for certain age by gender groups. In section 3.2 we explore differences in vaccination coverage estimates among the three surveys for various socio-demographic groups and by state to better understand whether there was consistency or not across the surveys for subgroups to be more or less vaccinated. In section 3.3 we present differences in several socio-demographic sample distributions for each survey to ACS benchmark distributions to investigate potential population coverage error. And finally in section 3.4, we present results from an experiment conducted by NORC to investigate potential reporting error associated with alternative wording of the question asking respondents if they received at least one dose of a COVID-19 vaccine in the NIS-ACM and the AmeriSpeak Omnibus surveys.

Analyses presented in sections 3.1 – 3.3 are based on survey data collected from November 28, 2021, through January 1, 2022, for the NIS-ACM; from December 29, 2021 – January 10, 2022, for the HPS; and from October Week 1 – December Week 1, 2021 for the AmeriSpeak Omnibus.

3.1 Comparison of Vaccination Coverage Estimates for Demographic Subgroups

Compared with 1+COVID-19 vaccination coverage by age and gender from CDC COVID Data Tracker in Graph 2, we observe that differences range from 0.02 to 5 percentage points (ppts) higher across males and females, with the somewhat larger overestimates of vaccination coverage for younger age groups for the NIS-ACM. HPS vaccination coverage estimates were most different from COVID Tracker for Males 75+ years of age and for females 18-29 years of age with an underestimate of 7.8 ppts and overestimate of 9.5 ppts respectively. AmeriSpeak Omnibus vaccination coverage estimates differed from COVID Tracker ranging from a 6.3 ppt overestimate for males 18-29 to an underestimate for females of 10.7 ppts.

Graph 2. Differences in 1+COVID-19 Vaccination Coverage for NIS-ACM, HPS, and AmeriSpeak Omnibus by Age and Gender compared with COVID Data Tracker, November 28, 2021 - January 1, 2022 (Percentage Points)

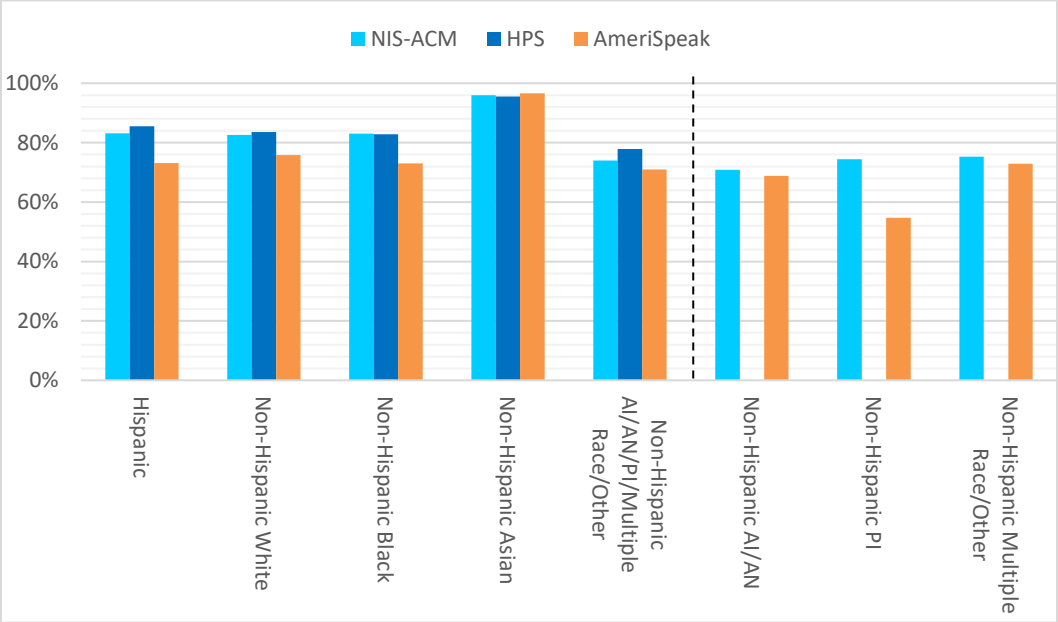


3.2 Vaccination Coverage Estimates for Socio-demographic Subgroups: NIS-ACM, HPS, and AmeriSpeak Omnibus

In this section, we estimated differences in 1+COVID-19 vaccination coverage between the three surveys among several socio-demographic subgroups.

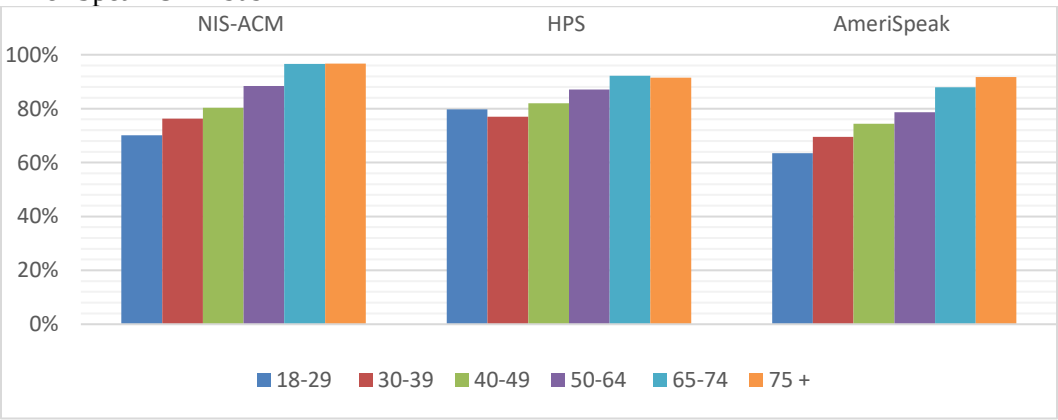
In Graph 3 to the left of the dashed vertical line, we observed that HPS and NIS-ACM vaccination coverage estimates were mostly comparable by race and Hispanic ethnicity for all listed race/Hispanic ethnicity subgroups. AmeriSpeak Omnibus vaccination coverage estimates tended to be lower than the NIS-ACM and HPS estimates – statistically lower than NIS-ACM and HPS estimates for Hispanic, Non-Hispanic White, and non-Hispanic Black. To the right of the vertical dashed line, we only report NIS-ACM and AmeriSpeak Omnibus estimates for individual race/Hispanic ethnicity subgroups of American Indian/Alaskan Native (AI/AN) and Pacific Islander and Multiple Race/Other. Estimates for individual race/Hispanic ethnicity subgroups of AI/AN and Pacific Islander and Multiple Race/Other are not reported here for HPS for reasons related to disclosure protection. NIS-ACM and AmeriSpeak Omnibus vaccination coverage estimates were not statistically different from each other using a z-test for the difference in two proportions at the .05 significance level.

Graph 3. 1+COVID-19 Vaccination Coverage by Race/Hispanic Ethnicity: NIS-ACM, HPS, AmeriSpeak Omnibus



Comparisons of vaccination coverage by age group across the three surveys is presented in Graph 4. We conducted z-tests for the difference in proportions and found that HPS estimates were significantly higher than NIS-ACM for 18-29 year olds, and significantly lower than NIS-ACM for 65-74 and 75+ year olds. AmeriSpeak vaccination coverage estimates were significantly lower than NIS-ACM and HPS estimates for all age groups (except for HPS age 75+), with differences ranging from 5 – 9 ppts lower than NIS-ACM estimates and 4 – 17 ppts lower than HPS estimates (all p-values < 0.05).

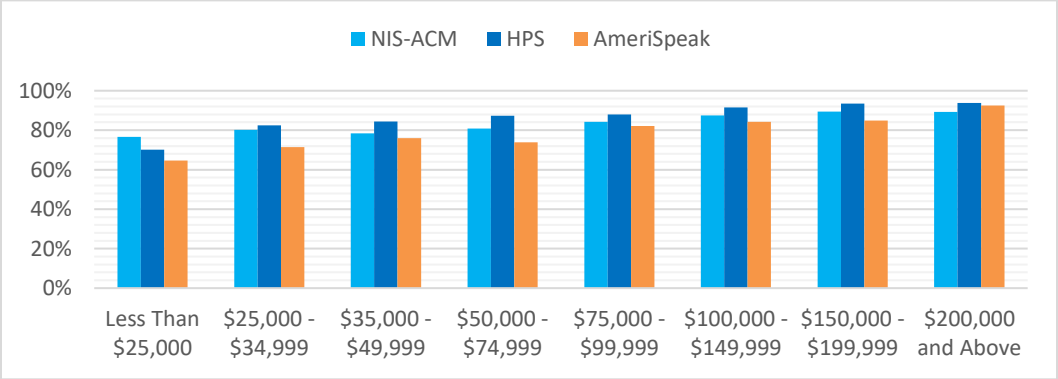
Graph 4. 1+COVID-19 Vaccination Coverage Estimates by Age Group: NIS-ACM, HPS, AmeriSpeak Omnibus



Comparisons of vaccination coverage for household income across the three surveys is presented in graph 5. We conducted z-tests for the difference in proportions and found that HPS vaccination coverage estimates were statistically significantly higher than NIS-ACM coverage estimates for income > 35K and AmeriSpeak vaccination coverage estimates

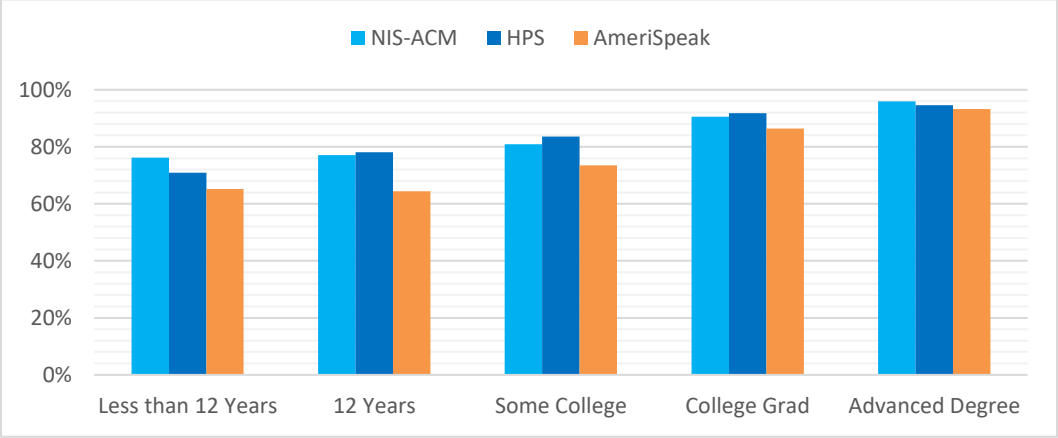
were significantly lower than NIS-ACM and HPS for all income groups (all p-values < 0.05).

Graph 5. 1+COVID-19 Vaccination Coverage Estimates by Income Group: NIS-ACM, HPS, AmeriSpeak Omnibus



Comparisons of vaccination coverage by education level are presented in Graph 6. We found that the HPS vaccination coverage rate was significantly higher than the NIS-ACM rate for those respondents with some college education. AmeriSpeak vaccination coverage estimates were statistically significantly lower compared to the NIS-ACM and HPS coverage estimates for respondents with 12 years, some college, and college graduate levels of education (all p-values < 0.05)

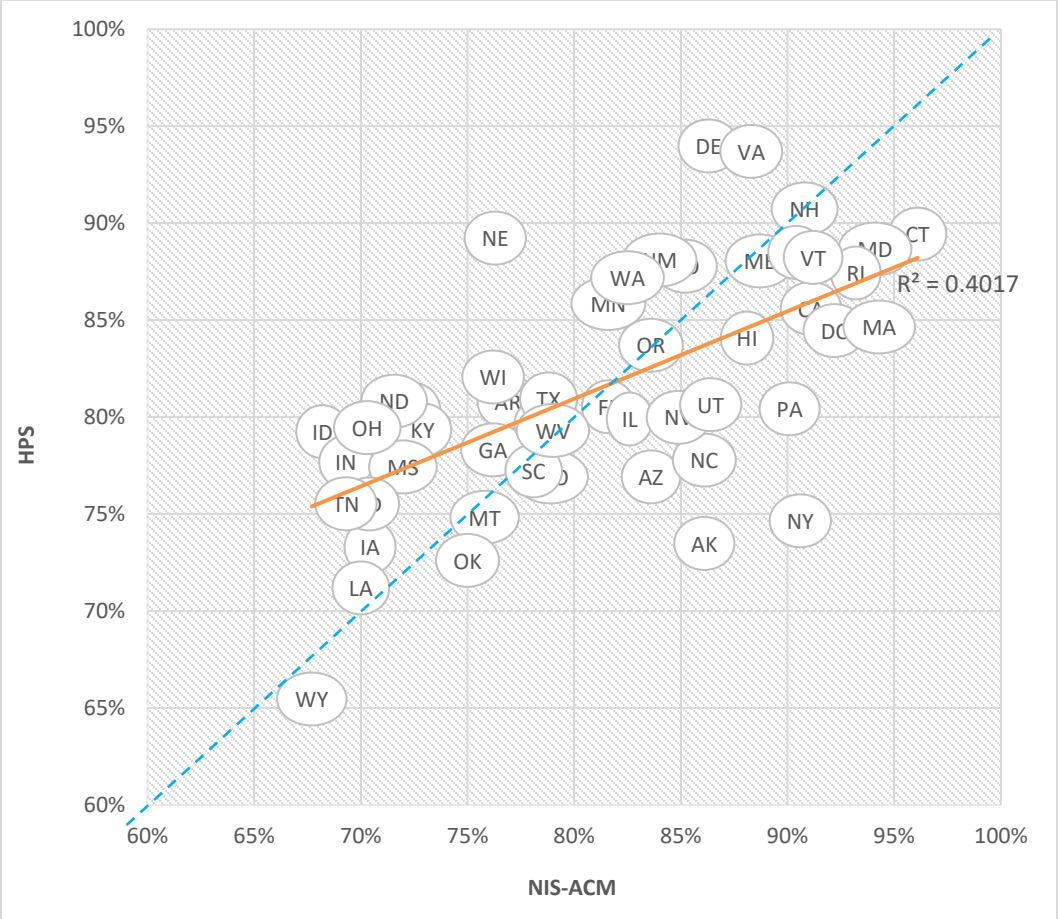
Graph 6. 1+COVID-19 Vaccination Coverage by Education Group: NIS-ACM, HPS, AmeriSpeak Omnibus



For state level vaccination coverage, sufficient sample sizes were available to report for the NIS-ACM and HPS and these vaccination coverage estimates are plotted against each other in Graph 7. Estimates of vaccination coverage significantly differed using a z-test for the difference in proportions at the p<0.05 significance level between the NIS-ACM and HPS for AL (70.5% vs. 80.1%), CA (91.1% vs. 85.6%), CT (96.1% vs. 89.4%), DC (92.2% vs. 85.5%), ID (68.2% vs. 79.2%), MA (94.3% vs. 84.6%), MS (70.2% vs. 78.7%), NY (90.6% vs. 74.7%), and PA (90.1% vs. 80.4%). The blue dashed line represents the identity line, and the orange line represents the fitted least squares line between the NIS-ACM and

HPS vaccination coverage across the states. An R^2=0.4017 suggested some moderate correlation in state level estimates between the NIS-ACM and HPS. States with vaccination coverage with less than 80% tended to have less variability between the NIS-ACM and HPS than states with vaccination coverage higher than 80%.

Graph 7. 1+COVID-19 Vaccination Coverage by State: NIS-ACM versus HPS

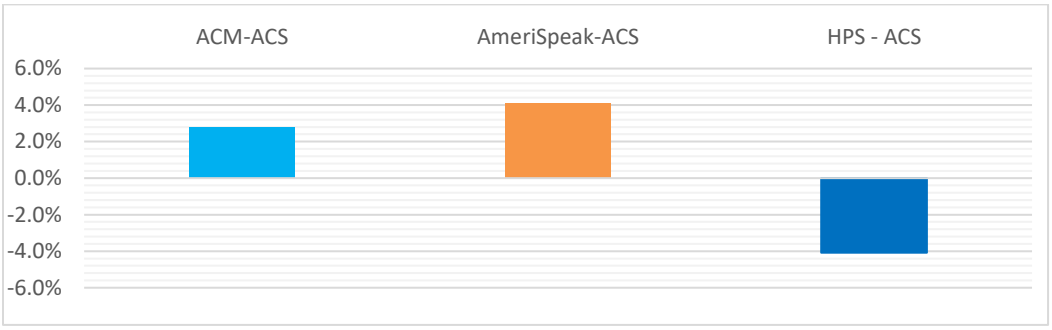


3.3 Comparison of Sample Demographics to ACS Benchmarks

In this section, we present an examination of several variables that are not used in post-stratification weighting for any of the three surveys to assess whether when compared to reliable benchmarks from the ACS, do we observe any sample bias due to coverage or other nonsampling error. In Graph 8 we present differences of each survey's design-weighted estimate of health insurance coverage compared to the ACS estimate. We observed that the HPS estimate of health insurance was lower than the ACS estimate. Both NIS-ACM and AmeriSpeak estimates of the population with health insurance were higher than ACS estimates. Though the difference in estimates of health insurance for each survey were statistically different than the ACS estimate using a z-test for the differences in proportions at the α=0.05 significance level, the absolute difference in the estimate of the

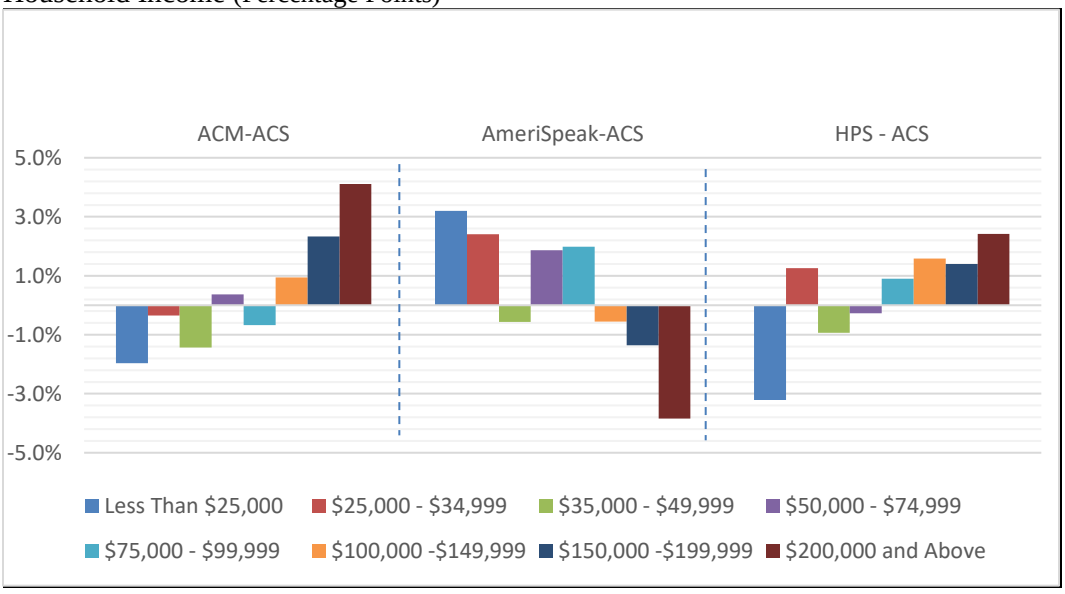
population with health insurance from the ACS estimate were small, ranging from 2.8 to 4.1 ppts.

Graph 8. Difference of Survey Design Weighted Estimate to ACS Estimate for "Have any Kind of Health Insurance=Yes" (Percentage Points)



Household income is also not used as post-stratification variables for any of the three surveys. Across the three surveys differences in survey estimates for each household income level from the ACS estimates ranged from -3.8 to 4.1 ppts In Graph 9, we find that NIS-ACM overestimated persons with >= $150K relative to the ACS, overestimated by 2.1 ppts for those with household income $150K to $199,999 household income and 4.1 ppts for those with household income > $200K. AmeriSpeak Omnibus household income estimates were higher than the ACS for persons with low income, and lower than the ACS for persons with household income higher than $200K. HPS underestimated persons with < 25K income relative to the ACS. In general, Though all differences noted above are statistically significantly different using a z-test for proportions at the α=0.05 significance level, the differences from the ACS observed for household income were modest.

Graph 9. Differences of Survey Design Weighted Estimates to ACS Estimates for Household Income (Percentage Points)



As discussed earlier for the NIS-ACM, population coverage error is not expected to be a substantive contributor to any potential bias in vaccination coverage; thus, the modest differences in socio-demographics of the NIS-ACM sample to ACS benchmarks may be attributable to nonresponse bias.

3.4 Experiment on Alternative Question Wording for Receiving at Least One Dose of a COVID-19 Vaccine

Past total survey error analysis of seasonal influenza vaccination coverage captured by the 2019-2020 NIS-Flu (NORC at the University of Chicago, 2020) provides insight into potential bias in self-reported COVID-19 vaccination status in the 2020 NIS-ACM. Net bias due to household level over-reporting and under-reporting of influenza vaccination status by parents/guardians for children and teens in the 2020 NIS-Flu was estimated to be an 8.26% overreporting with a confidence interval of [2.24%, 14.13%]. As such, experimentation with alternative question wording for receipt of a COVID-19 vaccination appeared warranted.

In August 2021, NORC conducted an experiment to investigate whether some level of the differences in NIS-ACM vaccination coverage estimates compared to the CDC COVID vaccine administration data may be attributed to misreporting error by NIS-ACM respondents due to hurried or dishonest responses, social desirability reasons, confusion, or other reporting error. Using a non-probability internet sample (web mode only), we tested two alternative versions of the receipt of a COVID-19 vaccination question along with the NIS-ACM version of the question, called VAX2. The VAX2_H1 alternative has a "not yet" response option and the VAX2_H2 question is designed after the Axios-Ipsos Poll question that splits out the yes responses into "yes I have received the vaccine," "Yes, a member of my immediate family," and "Yes, someone else.". From the experiment, we found one alternative version resulted in lower vaccination coverage relative to the VAX2 question the VAX2_H2 version. The three versions of the question were tested using a nonprobability sample in a split sample design with 1,206, 1,120, and 1227 completes achieved for the VAX2, VAX2_H1, and VAX2_H2 versions respectively. The three version of the question tested are listed below along with a graph illustrating the differences in vaccination coverage estimates among the three, with VAX2_H2 resulting in the largest change in vaccination rate relative to the VAX2 question:

- VAX2 (NIS-ACM)
  Have you received at least one dose of a COVID-19 vaccine? *Yes, No, Don't Know, Skipped*

- VAX2_H1
  Have you received at least one dose of a COVID-19 vaccine? *Yes, Not Yet, No, Don't Know, Skipped*

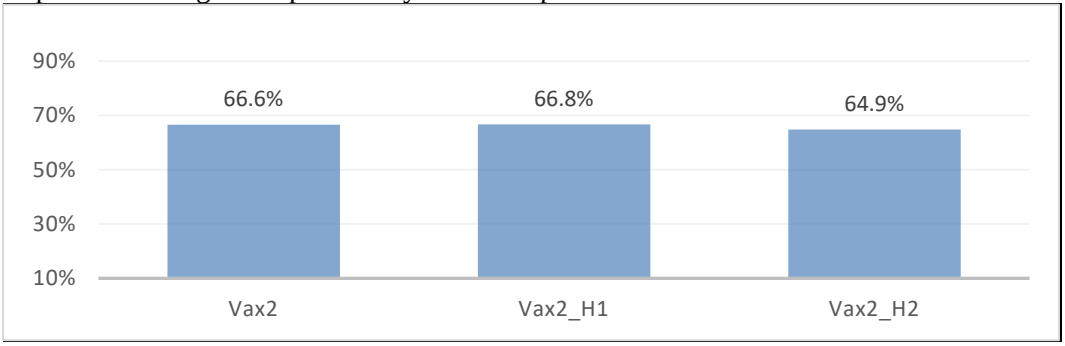
- VAX2_H2 (Axios-Ipsos Poll Question Wording[8])

[8] https://www.ipsos.com/sites/default/files/ct/news/documents/2022-04/Topline_Axios_Ipsos%20W65_withheld_0.pdf

Do you personally know anyone who has already received the COVID-19 vaccine? [Select all that apply] *Yes, I have received the vaccine; Yes, a member of my immediate family; Yes, someone else, No, Skipped*

As shown in Graph 10, the vaccination coverage estimate was 1.7 percentage points lower using the VAX2_H2 question compared to using the VAX2 question. Additionally, the VAX2_H2 vaccination coverage estimates were lower than VAX2 estimates for most age groups (18-29, 40-49, 50-64, 65-74, 65+ years), NonHispanic White, NonHispanic Asian, and those with either some college or an advanced degree. Though directionally VAX2_H2 provided results in the expected direction – lower vaccination coverage than VAX2 – the difference is small. Even under an assumption that the samples are probability based (which they are not), the differences among the 3 versions would not be statistically significant at the .05 significance level using a z-test for the difference in proportions.

Graph 10. 1+COVID-19 Vaccination Coverage: August 2021 NORC Question Wording Experiment using a Nonprobability Web Sample



4. Summary and Next Steps

The goal of the exploratory data analyses presented in this paper was to assess whether or not the three probability-based surveys sponsored by CDC that report 1+COVID-19 vaccination coverage were consistent with each other across socio-demographic groups since early in the vaccination roll-out period each survey reported estimates of vaccination coverage at the national level that were higher relative to CDC's administratively collected vaccination data on COVID Data Tracker. We identified some differences in vaccination coverage across the surveys as summarized below. We also investigated whether sample bias due to noncoverage or differential nonresponse may be a source contributing to the differences observed in vaccination coverage estimates by comparing sample estimates of variables not used in post-stratification weighting to ACS benchmarks. We found minor evidence of sample bias for health insurance coverage and household income in the three surveys. Finally, we fielded an experiment to begin to better understand whether response error could be reduced with alternative question wording. Our findings are summarized below:

- **Differences in 1+COVID-19 vaccination coverage estimates from the three surveys relative to CDC's COVID Data Tracker vaccination coverage were not consistent by age and gender:**

  - NIS-ACM estimated higher vaccination coverage for persons 18-64 years old, both sexes;
  - HPS estimated higher vaccination coverage for persons 18-49 years old both sexes; estimated lower vaccination coverage for those age 65+ years, both sexes;
  - AmeriSpeak Omnibus estimated higher vaccination coverage for males 18-49 years; estimated lower vaccination coverage for all ages of females and for males age 50+ years.
- **Some differences in 1+COVID-19 vaccination coverage among the three surveys were observed by age group, education level, household income and race/Hispanic ethnicity, and by state (HPS and NIS-ACM only):**
  - The HPS vaccination coverage estimate was significantly higher than the NIS-ACM estimate for 18-29 year-olds and significantly higher than NIS-ACM for those with some college at $p<0.05$;
  - The NIS-ACM vaccination coverage estimate was significantly higher than HPS for those with less than $25K annual household income at $p<0.05$;
  - AmeriSpeak vaccination coverage estimates were lower than NIS-ACM and HPS by age group, race/ethnicity (except NonHispanic Asian), education level, and household income level;
  - Estimates of vaccination coverage for states were significantly different at $p<0.05$ between the HPS and NIS-ACM for nine states and the correlation of state estimates between the two surveys was moderate, indicating substantive variability in reported state level vaccination coverage between the two surveys.
- **Modest differences of sample socio-demographic distributions of each survey to ACS benchmarks:**
  - Significant differences in estimates of health insurance coverage and household income from the ACS were found for each survey at $p<.05$, but the differences were modest -- on the order of 2-4 ppts.
- **The vaccination coverage estimate for an alternative wording of the vaccination question was lower than for the NIS-ACM question wording in the experiment:**
  - The vaccination coverage estimate for one of two experimental versions of the vaccination question was 1.7 percentage points lower than the control.
  - Although not conclusive, results from the experiment directionally demonstrate the potential impact of question wording on reporting error.

Among the sources of error that are the focus of the exploratory analyses presented in this paper, we put forward that population coverage error, nonresponse bias and response error may be substantive contributing factors to the differences observed between the HPS and the CDC vaccination administration data. For the NIS-ACM, likely primary contributing factors to differences in vaccination coverage estimates relative to the COVID-19 vaccination administration data are nonresponse error and response error. For the AmeriSpeak Omnibus, likely primary contributing factors are nonresponse bias, followed by response error.

Further research is underway by CDC, NORC, and the U.S. Census Bureau to better understand and quantify the sources of error in each of the three surveys that impact estimates of 1+COVID-19 vaccination coverage estimates. CDC is funding a study through the University of Colorado that will investigate the validity of self-reported vaccination status by matching the self-reports of vaccination to state and local vaccination registries (immunization information systems) .

NORC is working with CDC in several relevant research efforts for the NIS-ACM including (1) Fielding a questionnaire experiment in which a new question is added into the comparable COVID-19 survey for children, the NIS Child COVID Module (NIS-CCM), that asks respondents if they have a COVID-19 shot card for their child and if so to use the shot card during the interview; (2) Measuring response propensities by using the COVID Data Tracker estimates for sex, age as reference estimates; (3) Compare household-level and provider-level reported vaccinations in the NIS-CCM to help inform about potential misreporting error in the NIS-ACM, made possible as provider vaccination data was collected for NIS-CCM children age 13-17 in 2021; (4) Develop Total Survey Error (TSE) estimates of COVID-19 vaccination coverage using a model for the NIS-ACM similar to the TSE model developed for the NIS-Child. (Wolter, et al., 2017)

The U.S. Census Bureau is working with CDC in research efforts as well for the HPS including (1) evaluating Bayesian adjustments to weighting to reduce non-response bias that aligns with differential vaccination status; and (2) implementing new methods to incorporate sample frame address inaccessible by contact frame strategies (email and cell-phone SMS).

5. Limitations

Analyses presented in this paper primarily focused on vaccination data as reported in the three surveys during the October through December 2021 time period. Comparative results to the COVID-19 vaccination administration data from COVID Data Tracker and among the three surveys may be different for May through September 2021 since differences in national vaccination coverage for each of the surveys from COVID Data Tracker were typically larger.

Vaccination administration data reported on the CDC COVID Data Tracker is considered the gold standard for vaccination coverage in the U.S. Even so, it is known to have limitations as described in section 2.1.4 that are likely to lead to over estimating vaccination coverage. In development of a Total Survey Error model for the **NIS-**ACM, one goal is to incorporate estimates of error in the CDC COVID Data Tracker data, at least in terms of a sensitivity analysis and how they may impact estimates of total survey error.

References

A Cluster of Health Care Workers with COVID-19 Pneumonia Caused by SARS-CoV-2. (2021, February). *Journal of icrobioloby Immunology an Infection, 54*(1), 54-60.

Administration for Strategic Preparedness & Response. (July, 15 2022). *RENEWAL OF DETERMINATION THAT A PUBLIC HEALTH EMERGENCY EXISTS*. Retrieved from ASPR.hhs.gov: https://aspr.hhs.gov/legal/PHE/Pages/covid19-15jul2022.aspx

Baack, B. N., Abad, N., Yankey, D., Kahn, K. E., Razzaghi, H., Brookmeyer, K., . . . Singleton, J. A. (2021, June 25). COVID-19 Vaccination Coverage and Intent Among Adults Aged 18-39 Years - United States, March-May 2021. *MMWR Morbidity and Mortality Weekly report, 70*(25), pp. 929-933. doi:10.15585/mmwr.mm7025e2. PMID: 34166337; PMCID: PMC8224866.

Blumberg, S. J., & Luke, J. V. (2022, May). Wireless Substitution: Early release of estimates from the National Health Interview Survey, July-December 2021. doi:https://doi.org/10.15620/cdc:116630

Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X.-L., & Flaxman, S. (2021). Unrepresentative Big Surveys Significantly Overestimate US Vaccine Uptake. *Nature*, 695-700. Retrieved from https://doi.org/10.1038/s41586-021-04198-4

Centers for Disease Control and Prevention. (2022, August 16). *CDC Museum COVID-19 Timeline*. Retrieved October 3, 2022, from Centers for Disease Control and Prevention: https://www.cdc.gov/museum/timeline/covid19.html

Centers for Disease Control and Prevention. (2022, July 26). *National Immunization Survey Adult COVID Module (NIS-ACM): COVIDVaxViews|Data|Center for Disearse Control and Prevention (CDC.gov)*. Retrieved from Data.CDC.gov: https://data.cdc.gov/Vaccinations/National-Immunization-Survey-Adult-COVID-Module-NI/udsf-9v7b

Fields, J. F., Hunter-Childs, J., Tersine, A., Sisson, J., Parker, E., Velkoff, V., . . . Shin, H. (Forthcoming). *Desing and Operation of the 2020 Household Pulse Survey, 2020*. U.S. Census Bureau. Retrieved from https://www2.census.gov/programs-surveys/demo/technical-documentation/hhp/2020_HPS_Background.pdf

National Center for Health Statistics. (2022, July 20). *Health Care Access, Telemedicine, and Mental Health*. Retrieved from CDC NCHS COVID-19 Data from NCHS: https://www.cdc.gov/nchs/covid19/health-care-access-and-mental-health.htm

Nguyen, K. H., Srivastav, A., Razzaghi, H., Williams, W., Lindley, M. C., Jorgensen, C., . . . Singleton, J. A. (2021, February 12). COVID-19 Vaccination Intent, Perceptions, and Reasons for Not Vaccinating Among Groups Prioritized for Early Vaccination -- United States, September and December 2020. *Morbidity and ortality Weekly Report, 70*(6), pp. 217-222.

Reinhart, A., & Tibshirani, R. (2021, September 2). Big data, big problems: Responding to "Are we there yet?". *arXiv*. Retrieved from https://arxiv.org/abs/2109.00680

U.S. Bureau of the Census. (2022, August 31). *Household Pulse Survey Technical Documentation*. Retrieved from United States Census Bureau: https://www.census.gov/programs-surveys/household-pulse-survey/technical-documentation.html

Wolter, K. M., Pineau, V. J., Skalland, B., Zeng, W., Singleton, J. A., Khare, M., & et al. (2017). Total Survey Error Assessment for Sociodemographic Subgroups in the 2012 U.S. National Immunization Survey. In P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, & et al., editors, *Total Survey Error in Practice* (pp. 433-453). Hoboken, N.J.: John Wiley & Sons, Inc.