2021 Minnesota Health Access Survey Technical Report:
Study Design and Data Processing Methodology
SHADAC, June 2021

This report describes the Minnesota Health Access Survey (MNHA) data collection process and methodology, emphasizing the most recent administration of the survey completed in 2021. The 2021 MNHA represents the first time using a single address based (ABS) frame. This report describes this decision and other aspects of the methodology and is organized as follows:

Section 1. Overview of the MNHA
Section 2. Sampling Goals and Methodology
Section 3. Survey Content
Section 4. Survey Administration
Section 5. Response and Sample Coverage
Section 6. Data Editing and Variable Construction
Section 7. Survey Weights
Section 8. Data Analysis
Section 9. Availability of Research Findings

1. Overview of the MNHA

The Minnesota Health Access Survey (MNHA) is a biennial survey of non-institutionalized Minnesota residents. The survey collects detailed information on health insurance coverage options, access to coverage and health care services, and basic demographic data. The goal of the survey is to document trends in health insurance coverage, and access to insurance and health care at the state and regional level, as well as for select subpopulations (e.g., rural, low-income families, households with children, populations of color and American Indians). The MNHA represents a partnership between the Minnesota Department of Health (MDH) Health Economics Program and the University of Minnesota's State Health Access Data Assistance Center (SHADAC).

The MNHA data play an important role in monitoring trends in health insurance coverage, evaluating and informing health policy development in Minnesota on topics such as affordability of coverage, access to healthcare, and redesign of public program coverage. The MNHA provides precise and timely estimates on a range of coverage and access relevant questions, is adaptable and responsive to state health policy concerns, and ensures the availability of micro-data for time sensitive research and policy analysis.

The MNHA has been conducted a number of times over the years: in 1990, 1995, 1999, 2001, 2004, and every two years beginning in 2007. Beginning in 2007, MNHA funding is from a legislative appropriation to the Minnesota Department of Health and additional support from the Minnesota Department of Human Services since 2011. Each cycle MNHA survey data are collected through a subcontract with a survey vendor. SSRS, a full-service social science research firm based in Pennsylvania, has been the

vendor of choice since 2009. This report focuses primarily on the 2021 MNHA methodology and provides some historical data in table form.[1]

2. Sampling Goals and Methodology

Each year the sample is designed to ensure that estimates derived from the survey are representative of the overall population and inferences are largely unbiased. The sampling goals are to generate reliable health insurance coverage estimates for the state overall, the state's 13 Economic Development Regions (target of 300 completes each), the more populous counties (e.g., Hennepin and Ramsey), populations of color (African Americans, Hispanics, and Asians) and American Indians in Minnesota, low-income households, and households with children. In addition, the goal is to under-sample Minnesotans who are age 65 and over, due to the prominence of Medicare insurance among this segment of the population.

The 2021 MNHA marked a full transition to an address-based sample (ABS) frame. This decision was based on experience with dual sample frames in 2019, as well as rising costs and challenges representing the Minnesota population using a random digit dial (RDD) landline and cell telephone sample.[2] The survey was administered in two modes: computer-assisted web interviews (CAWI) and computer-assisted telephone interviews (CATI) among those unable or unwilling to complete the survey on the web.

ABS Frame
The ABS sample was generated from the United States Postal Service (USPS) Computerized Delivery Sequence File (CDSF) that contains information on all delivery addresses serviced by the USPS except general delivery addresses; it is updated weekly. Sample was drawn from all residential records with the following exceptions: addresses coded as vacant, seasonal (vacation), and PO boxes (other than those defined as "only way to get mail"); this avoids selecting duplicate addresses into the sample.

Predictive modeling was used to target demographics. This advance in survey sampling appends to the ABS frame preexisting commercially available data on household attributes to meet sample goals. Using random forest prediction, models are built using these data to score future samples to predict the desired attributes. SSRS appended all available data from voter registration databases, consumer databases, Marketing Systems Group (MSG) database information (all ranges of surnames), and Census Planning Database (PDB) data to the sample. All the appended data served as independent variables (features) in random forest models, while the self-reported attributes (demographics, etc.) from the 2019 MNHA served as the test data for dependent variables.

The 2021 MNHA used the following strata:
       1. Hispanic or Spanish language
       2. High density African American
       3. Low-income (less than 200 FPL)
       4. Asian
       5. Households with children
       6. Over 65+
       7. Residual non-match
       8. Residual match

---

[1] For information about earlier versions of the MNHA Technical Report contact Kathleen Thiede Call at callx001@umn.edu or the Health Economics Program at health.mnha@state.mn.us.

[2] Transitions from telephone surveys to self-administered and mixed-mode surveys. AAPOR Task Force Report, October 2019. Available from: https://www.aapor.org/Education-Resources/Reports/Transitions-from-Telephone-Surveys-to-Self-Adminis.aspx

Household Screening

To be eligible for the MNHA the household must be the primary residence and must be located in Minnesota. Within each household that consents to participate, an adult (age 18 and older) knowledgeable about household members' health insurance was asked to complete the survey

Target of Survey

The knowledgeable adult enumerates the household – providing gender, age, and relationship information for all household members. In households with more than one person, the CAI software randomly selects a person as the target of the survey. To increase the likelihood of selecting a child target, children are weighted 50 percent more than other household members. The proportion of child targets (under age 18) overall was 14.8 percent.

Sample Release
ABS sample was intended to be released in two waves, with continuous monitoring of outcomes and productivity by frame and strata.[3] Wave 1 was comprised of one third of the sample. Results from Wave 1 allowed for model refinement in the Wave 2 (adaptive design) which comprised two thirds of the sample. Wave 3 was added to supplement the sample with the goal of increasing the number of completes among people lacking insurance. The yield for the Wave 1 was projected to be 7.5:1 and the actual yield was lower at 5.2:1. Based on the experience with Wave 1, the projected yield for Waves 2 and 3 was 5.8:1 which was close to the actual yield shown in Table 1.

Table 1. Sample and Yield by Wave				
Wave	**Sample size**	**Completes**	**Actual Yield**	**Release Date**
1	27,386	5,250	5.2:1	10/01/2021
2	48,287	8,940	5.4:1	11/30/2021
3	24,148	4,414	5.5:1	12/03/2021
Total	99,821	18,604	5.4:1	

3. **Survey Content**

Each year the majority of MNHA survey questions ask about health insurance coverage for the randomly selected person within the sampled household – the target. This is followed by questions about health insurance coverage for all other household members, and education and employment information for all adults in the household. Information is also collected concerning potential sources of insurance (e.g., through the target's own or a family member's employer). Those lacking insurance are asked why they (or, in the case of a child target, their parents) did not purchase coverage.

In addition, questions are included about the target's health status, access to health care, affordability of care, dental coverage, marital status (requested for primary caregiver or wage earner if the target is a minor), county of residence, race, ethnicity, nativity, citizenship, and length of time living in the US. Finally, information about family income is requested along with questions relevant to weighting the data (e.g., home ownership; and in the event we needed to add a RDD sample frame questions about presence of working cell phones and a count of prepaid cellphones were included– L1A, L1B, L2, L2A).

---

[3] Details about the 2021 sample design in the SSRS Methods Report and Sample Plan are available by contacting Kathleen Thiede Call at callx001@umn.edu and the Health Economics Program at health.mnha@state.mn.us.

Some survey content changes each year the MNHA is conducted. This flexibility to alter questions to meet policy needs is a major advantage of the MNHA over other federal sources of data. Each year key stakeholders are consulted to inform revisions. Below is a summary of changes to the 2021 survey.

Revisions

Revisions to the 2021 MNHA focused on streamlining sections of the survey to reduce respondent burden. The most significant changes focused on the section of the survey exploring access to health insurance coverage through one's own or a family members' employer (NEWCOV), a spouse (SPCOV) or a parent for targets under 26 years of age (PARCOV).

This streamlining led to repositioning, adapting, and renaming several questions related to prior coverage (PRIOR2/UNIN2 became H21b), most recent coverage (UNIN1 became H21a), gaps in coverage (INSD3 became H20b, EVERT became H21), and reasons for gaining current coverage (INSD2 became H20) or lacking coverage (OWNCOV/OWNCV2 became NOCOV).

Because Indian Health Status (IHS) is not comprehension health insurance coverage, the IHS question (H4a) was moved to the end of the health insurance series (H4a).

Web administration of the survey provides the opportunity to include grids for questions related to multiple people in the same household (e.g., household roster – age, sex, relationship, etc.) and question sets focused on the same content and response options (e.g., AFFRD, TGRID, VGRID used to assess use of telehealth). Finally, during this restructuring, the education question (EDUC) moved before the employment questions. The web survey was optimized for mobile phone administration that did not include grids.

Additions/Substitutions

Several additions in 2021 were focused on potential COVID-19 disruptions and adaptations to accessing health care services. Among people not reporting any health care use in the past year, a set of potential reasons were provided to understand COVID-19 related cancellations or delays (NOUSE) and a question directly asking about postponing or delaying visit due to COVID-19 (DELAY) and reasons (WHYDELAY). Several questions were added concerning use of telehealth in the past year (TELEH), choice of telehealth versus office visits (TCHOICE), type of providers visited by phone (TPROVTYPE) and/or video (VPROVTYPE), and questions about access to the internet that can support video calls were added to an existing question (BANKWEB). Finally, two questions asked about the household's difficulty paying for necessities (such as food, housing, utilities, and healthcare) before (COVIDa) and since (COVIDb) the coronavirus outbreak.

In addition, response options for the gender question were expanded to be more inclusive (S7#_SEX), a sexual orientation question (S7#_SEXOR) and question asking about experiences of unfair treatment by health care provider based on gender, sexual orientation, gender identity or gender expression (SEXDIS).

Finally, influenced by continued interest in disaggregating social and structural forces that differentially impact racial and ethnic groups,[4] versions of several questions from earlier versions of the MNHA were reinstated. Specifically, participants who identify as Black or Asian in the main race question (RACE) receive follow-up questions that allow for more specific cultural identification (BLACK, ASIAN).

Omissions

---

[4] Kauh TJ, Read JG, Scheitler AJ. The Critical Role of Racial/Ethnic Data Disaggregation for Health Equity. Population Research and Policy Review. 2021, 40:1-7. https://dio.org/10.1007/s11113-020-09631-6

A number of questions were deleted as a result of streamlining the health insurance access sections described above and removal of several confirmation questions (H18a, VCHK##, NEWCHKQ##, PATHI, PATHI1, PROBLEM1, PATHU, PATHU1).

The move to an ABS sample also made question about landline and cell phones obsolete (PHONE, PHONE2, PHONE6-8, CELL5, 6, 9, 10).

Finally, both sets of health insurance literacy questions were deleted (HINAVA-D, HILMA-D). This was justified following the publication of a paper exploring the value of these questions for understanding access and affordability.[5]

4. **Survey Administration.**

The 2021 MNHA was conducted by SSRS, an independent survey research company based in Pennsylvania. The 2021 MNHA field period was October 1, 2021 – January 10, 2022. Challenges with invoicing delayed the start of survey and the need to add more sample extended the field period into early 2022. Table 2 provides the field period for each MNHA beginning in 2001. Beginning in 2004 the goal has been to field the survey late summer and into the fall, ending before the November Thanksgiving holiday. A detailed report of data collection procedures and timeline authored by SSRS is available by request.[6]

Table 2. MNHA Field Period	
Survey Year	**Dates**
2001	November 2000 - May 2001
2004, 2007	July - December
2011	September - December
2009, 2013, 2015	August - November
2017	June - October
2019	September - December
2021	October 2021 – January 2022

The ABS frame allows adoption of a mixed-mode survey format, encouraging participants the opportunity to complete the survey by web and providing a telephone option for those with broadband constraints or for those who prefer this mode. The online self-administered Computer Assisted Web Interview (CAWI) and a Computer Assisted Telephone Interview (CATI) version are part of the same program. Small differences between modes include phrasing of question read by respondents versus interviewers (first person vs second person); removing don't know (except where lack of knowledge is of interest such as the amount of the annual deductible, another household member's insurance coverage, employment, etc.) and refusal options because respondents are allowed to skip questions they do not wish to answer.

The CATI/CAWI survey was programmed and thoroughly reviewed by all partners (SSRS, MDH, and SHADAC) prior to pretesting the instrument. The review consisted of multiple iterations of analyzing the accuracy of the skip pattern logic and interviewer directions for this complex instrument.

---

[5] Call KT, Conmy A, Alarcon G, Dorelian AM, Hagge S, Simon AB. Health Insurance Literacy: How Best to Measure and Does it Matter to Health Care Access and Affordability. *Research in Social and Administrative Pharmacy.* 2021, 17(6):1166-1173. PMID: 32952089 DOI: 10.1016/j.sapharm.2020.09.002
[6] Contact Kathleen Call at callx001@umn.edu or Health Economics Program at health.mnha@state.mn.us.

Pretesting

Before beginning the field period, pretests were completed in both web and telephone modes between August 24 and 30, 2021 using sample from SSRS's omnibus survey. For the CAWI survey, 18 web cognitive interviews were completed. Of these 15 were completed on a desktop or laptop and 3 were completed on a mobile device. The goal was to test comprehension of new items, the flow of the reordered survey, and usability of the web instrument. The full team identified probing questions in advance. CAWI pretest participants were paid $30 for this more lengthy and demanding process. For the CATI survey, 25 interviews were conducted. CATI pretest participants were paid $10.

All pretest interviews were recorded and made available on a secure FTP site for review by the MDH and SHADAC team. In general, the CAWI and CATI programs worked well. CAWI participants found the layout clean and easy to navigate; several CATI participants indicated the survey was too long. Feedback from MDH, SHADAC and pretest results were incorporated in the final version of the programs. Pretests were not included in the final sample (SSRS pretest memo is available by request[7]).

Interviewer Training

Prior to pretesting, standard SSRS interviewer training was provided to account for participants requesting telephone interviews in the ABS frame. In addition, interviewers receive project specific training to understand the overall objectives of the project, survey content and any specific procedures. As part of this training, the following materials provided to interviewers:
- A video sharing the history of the MNHA survey co-created and presented by SHADAC and MDH representatives (created June 2021)
- A briefing document that contained information about the goals of the study, potential obstacles to be overcome in getting good answers to questions, and respondent problems that could be anticipated ahead of time as well as strategies for addressing challenges
- An annotated questionnaire
- Contact information for project personnel
- A list of potentially difficult words and their phonetic pronunciations (e.g., Minnesota counties)
- Copies of letters and postcards sent to respondents so they would be familiar with the materials seen by those who called in to complete an interview

Interviewers were monitored by Call Center Supervisors and provided feedback, where appropriate, to improve interviewer technique and use of materials. The project team regularly communicated with the Call Center Managers to address any questions or concerns.

Mailings

Sampled households were contacted up to three times and followed strategies designed to increase response rates (one measure of quality):

First mailing: The advance letter addressed to "Minnesota Resident" includes a description of the study, the expected time to complete the survey, instructions to complete the survey online via a secure website (URL and secure access code), and a 1-800 number to request a telephone interview if preferred or needed.

---

[7] Contact Kathleen Call at callx001@umn.edu or Health Economics Program at health.mnha@state.mn.us.

In 2021, standard stamps (vs metered postage stamps) are used and custom envelopes with windows made the $2 bill visible. Past research show that the use of a pre-incentive consistently leads to higher response rates.[8]

Households in high-density Hispanic areas (based on Census data) receive bilingual letters (front and back Spanish versus English).

Letters and postcards have logos and letters have signatures from both organizations representing the MNHA team: University of Minnesota and the Minnesota Department of Health.[9]

Second mailing: 1-week later postcard reminders are mailed to households not responding to the first letter. This includes a reminder about the invitation letter, instructions for completing on the web, a 1-800 number to request a telephone interview if preferred or needed, and a statement in Spanish to call the 1-800 number to complete the survey by phone. Again, standard stamps are used.

Third mailing: 2 weeks later a final letter is mailed to all households not yet responding. The final letter states the survey is toward the end of the data collection period. Otherwise, the letter contains the same elements included in the initial letter, without the $2 pre-incentive. Households in highly concentrated Hispanic areas receive a bilingual letter.

The final letter was sent using different mailing methods dependent on the sample strata:
- Hard-to-reach sample strata (Hispanic or Spanish language, high density African American, and low-income) received letter by either USPS first class, USPS priority or FedEx (respectively $7.40 and $8.20 per mailing, envelope included).
- All other strata received letters via USPS first class and mailed in a standard envelope ($0.47 for postage and $0.53 for the envelope, or $1.00 per mailing)

Results from a mail experiment indicated sending the third mailing by FedEX yielded the highest response rate for three hard-to-reach strata: (1) Hispanic or Spanish Language, (2) High density African American, and (3) Low-income (less than 200% FPL). (See next section and Table 3). However, the third mailing for Waves 2 was set for December 17 and FedEx was not able to fully complete this task (due to holiday demand, staff shortages and weather). Therefore, of the hard-to-reach third mailings for Wave 2 and Wave 3, most were sent via USPS Priority (95%) versus FedEX (5%). For Waves 2 and 3, the third mailing with FedEX/Priority included one additional sample strata -- residual non-match -- which yielded more uninsured cases in Wave 1 relative to other strata.

Mail experiment
An experiment was conducted in Wave 1 of the sample to determine the most effective third mailing method for the Wave 2 sample. The experiment had three arms: USPS first class, USPS priority mail or FedEx. This experiment was restricted to the hard-to-reach sample strata. As shown in Table 3, FedEx performed better than USPS priority mail, which performed better than USPS first class mail. The greatest impact was seen for the high density African American and low-income strata. The increase in

[8] Singer E, Ye C. The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*. 2013, 645(1):112-141. DOI: 10.1177/0002716212458082

[9] A 2019 MNHA experiment in the first wave of the ABS sample release randomly assigned sample to letters with (1) UMN, (2) MDH or (3) cobranded logos. The resultsGood indicated the MDH logo performed the best, followed by the cobranded logo. The cobranded logo was selected to represent the study team.

completion rates per strata also helped meet incidence goals (i.e., raw counts within these subpopulations).[10]

Table 3. Completion Rate for Wave 1 Mail Experiment by Strata, 2021			
Strata	USPS First Class	USPS Priority	FedEx
Hispanic or Spanish language	10.8%	10.6%	11.3%
High density African American	12.4%	14.1%	16.5%
Low-income (less than 200% FPL)	11.7%	12.3%	16.0%
Overall	11.9%	12.9%	15.7%

Telephone Protocol

After each mailing, SSRS phone lines were staffed with live interviewers for several days. Otherwise, respondents would get a voicemail recording. These calls were returned by an interviewer. A maximum of six calls were made to each number scheduled on different days and different times, unless an appointment was made. Interviewers left voicemail on the first call and then again at their discretion for other calls.

All outbound phone calls had a programmed caller ID message reading "UMN HLTH STUDY" and a "612" area code. The text is only visible on landline phones with a visible caller ID identifier. The phone number appears on both landlines and cell phones. This was less relevant in 2021 since outbound calls were limited to returning calls from participants who left voicemails.

Monitoring
Data collection was monitored over the course of the study (live by SSRS supervisors). MDH and SHADAC were able to log into Confirmit to monitor ongoing progress between bi-monthly meetings with the SSRS team.

Completes by Mode and Language

As described above, participants can complete the survey by web or telephone to account for broadband constraints and preferences. Table 4 shows that a total of 1,001 participants (5 percent) called to complete the survey by telephone; this is a bit higher than the portion of the 2019 ABS frame completing the survey by telephone (2 percent), the year a paper copy was a third option for completing the survey. In 2021, those completing the survey by telephone are more likely to be age 65 and over (56 percent) than those completing the web survey (16 percent).

Table 4. Completed MNHA Surveys Mode, 2021			
Mode	ABS Frame		Total
	Completes	Partial Completes*	
Telephone	1,001	0	1,001
Web	16,833	770	17,603
Total	17,834	770	18,604

*Partial completes are defined as participants who did not complete the entire survey yet did complete the health insurance section, the access to coverage series, and up to the first question in the access section (SatisfyA).

---

[10] Harrell V, McPhee C, Goyle A, Trieu J, Loveridge C, Call K, Simon AB. Look, This Mail is Important! Can Priority Mail or FedEx be an Effective Adaptive Design Intervention in ABS surveys? May 2022, Presented at AAPOR, Chicago IL.

The MNHA is conducted in English and Spanish. Participants are allowed to skip questions they do not want to answer; explicit refusal options are only included for sensitive questions and questions a proxy respondent may not be able to answer about a target. The time it takes to conduct an interview varies by household size, the target's insurance status, mode, telephone status, and survey language. The average length of time it takes to complete the MNHA interview is relatively consistent; telephone interviews taking longer to complete than the web survey (see Table 5). The gap between telephone and web completion times shrunk with the ABS only frame. Generally, in both English and Spanish, having an interviewer read the questions and response options for the telephone phone requires more time than reading the survey on the web. Due to the complexity of translating health insurance and access terms, the surveys completed in Spanish require more time on average than English language interviews.

Table 5. MNHA Average Survey Length in Minutes by Mode and Language*						
Year	Telephone	Web	Telephone		Web	
			English	Spanish	English	Spanish
2019 (RDD+ABS)	26	18	26	36	18	46
2021 (ABS)	25	20	25	44	19	36

*A total of 30 and 74 interviews were conducted in Spanish in 2019 and 2021 respectively. In 2019, 21 Spanish surveys were completed in the RDD and 9 in the ABS sample. In 2021, 73 of the Spanish surveys were completed web and 1 by telephone.

IRB Approval

The study received IRB approval from MDH and the University of Minnesota. For an ABS frame, the letter serves as consent, inviting consenting participants to use a unique access code to complete the web survey or contact SSRS to complete the survey by telephone. Participants are provided telephone number and email addresses for (1) the University of Minnesota Research Subjects' Advocates Line should they have concerns about the study experience and (2) Kathleen Call should they have questions about the survey. Each year records of participant contacts with Kathleen Call are recorded and coded to monitor the frequency and nature of participant inquiries.

Contacts from MNHA respondents are a relatively rare occurrence (see Table 6) given the large sample size (frame of nearly 100,000 as shown in Table 1). Introducing the ABS frame with surveys predominantly completed by web results in different reasons for potential participants to reach out. For example, a dominant reason for making contact are challenges with the web survey or calls to indicate broadband issues (29 percent in 2019 and 24 percent in 2021). Each mailing prominently features the phone number available to complete the survey; it was provided again when the call is returned. Reports that the survey was already been completed is another dominant reason for making contact (20 percent in 2019 and 24 percent in 2021); the lead-time required to print and prepare the mailings was described as the cause for the additional contact. However, the use of expensive mailing methods (USPS Priority and FedEx) may have led to the increase in callers, some of whom voiced concerns about what they deemed as frivolous use of resources. The proportion calling to voice a lack of interest in participating or inability to complete the survey increased in 2021 (14 percent compared to 6 percent in 2019). Finally, in 2021 a new "other" code was added which includes mostly people who left voicemails without contact information, or left contact information but no reason for the call, or sharing a concern the SSRS phone number, or people who simply had a story to share, representing 16 percent of the contacts. Calls to verify the legitimacy of the survey decreased substantially with sole reliance on ABS, indicating the letter of introduction provides adequate information for many prospective participants (from 46 percent in the RDD only frame in 2017 to 6 percent in 2021).

Table 6. Log of Contacts from MNHA Participants		2017 RDD		2019 RDD & ABS		2021 ABS	
Code	Labels	Count	%	Count	%	Count	%
1	Concerns about health care/health policy	14	26%	3	3%	1	0%
2	Legitimacy of survey, questions about survey, use of data	25	46%	17	17%	17	6%
3	Question about applying for insurance	4	7%	1	1%	1	0%
4	Complaints/not interested/can't complete/do not call	5	9%	6	6%	41	14%
5	Comments or questions about survey (positive, neutral or negative)	6	11%	8	8%	21	7%
6	No computer, challenges with web survey			29	29%	69	24%
7	Out of sample, business or not a MN resident			4	4%	16	6%
8	Already completed survey			20	20%	70	24%
9	Called after close of data collection; other miscellaneous			12	12%	8	3%
10	Other (left voicemail but no contact info, share a story)					46	16%
Total comments coded		**54**	**100%**	**100**	**100%**	**290**	**100%**
Unique contacts			*43*		*98*		*281*
Initiated by IRB			*2*		*10*		*29*
Email contact (versus telephone) contacts			*3*		*2*		*41*

5. Response Rates and Sample Coverage

Response rates are dropping for all modes of data collection[11] with stable but consistently lower response rates for web compared to telephone surveys, which is of concern because access to the internet is not randomly distributed and may lead to nonresponse bias.[12] Using data from four national US telephone surveys between 1996 and 2015, weighted estimates indicate either no change or decreases in the bias associated with falling response rates.[13] This is consistent with earlier analyses showing that there is not necessarily a link between low response rates and nonresponse bias.[14] Carefully specified sampling design and adaptive design throughout data collection lower nonresponse bias.[10,15]

---

[11] Williams D, Brick JM. Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology.* 2018, 6 (2):186-211. DOI: 10.1093/jssam/smx019.

[12] Daikeler J, Bošnjak M, Lozar Manfreda, K. Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates, *Journal of Survey Statistics and Methodology*. 2020, 8(3):513-539. DOI: 10.1093/jssam/smz008.

[13] Dutwin D, Buskirk TD. Telephone Sample Surveys: Dearly Beloved or Nearly Departed? Trends in Survey Errors in the Era of Declining Response Rates. *Journal of Survey Statistics and Methodology*. 2020, 9(3):353-380. DOI: 10.1093/jssam/smz044.

[14] Groves R, Peytcheva E. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis. *Public Opinion Quarterly*, 2008, 72(2): 167-189.

[15] Coffey S, Reist B, Miller PV. Interventions On-Call: Dynamic Adaptive Design in the 2015 National Survey of College Graduates, *Journal of Survey Statistics and Methodology*. 2019, 8(4):726-747. DOI: 10.1093/jssam/smz026.

The *response rate* is the ratio of the number of completed interviews divided by the number of eligible reporting units in a sample. The response rates reported below refer to AAPOR Response Rate #3[16] The *cooperation rate* is the ratio of all interviewed cases to all eligible cases contacted. The *refusal rate* is the proportion of all cases in which a housing unit or the respondent refuses to be interviewed, or breaks-off an interview, of all potentially eligible cases.[17] As presented in Table 7 below, consistent with other surveys, the MNHA response and cooperation rates have somewhat diminished over time. By contrast, refusal rates have fluctuated over time and are lower in ABS frames as a function of greater ability to passively refuse a mailed invitation compared to active refusals to telephone recruitment with RDD samples.

Table 7. MNHA Count of Complete Surveys, Response, Cooperation and Refusal Rates				
Survey Year and Frame	Total Completes^	Response Rate*	Cooperation Rate*	Refusal Rate**
2001 RDD	27,315	67%	78%	19%
2004 RDD	13,802	59%	68%	28%
2007 RDD	9,728	43%	57%	32%
2009 RDD	12,031	45%	53%	39%
2011 RDD	11,355	44%	45%	39%
2013 RDD	11,778	48%	48%	23%
2015 RDD	11,178	35%	36%	30%
2017 RDD	12,042	29%	30%	32%
2019 combined	11,530	22%	22%	6%
2019 RDD	3,673	16%	17%	17%
2019 ABS	7,857	24%	25%	0.8%
2021 ABS	18,604	20%	20%	1.5%

MNHA 2001-2007 represent landline sample frames; MNHA 2009 forward represent dual landline and cell phone sample frames.
^ The total count includes partial interviews. Cases were designated as partial completes if the survey was completed through the health insurance coverage (H series) (2001-2015), roster coverage and demographics, and access to coverage (COV) series where applicable (2015 forward).
* Based on AAPOR RR4 response and cooperation rates from 2001-2007; Based on AAPOR RR3 response and cooperation rates from 2009 forward which excludes partials.
** Based on AAPOR refusal rate 2 (REF2); includes estimates of eligible cases among unknown cases. For comparability with prior MNHA surveys, refusal rate calculations from 2009 forward ignored screening that occurred (e.g., excluding minors both years and over sampling of cell only households). Note that calculating refusal rates for ABS frames is challenging in that mail is sent out and not returned without a feedback loop to establish an active refusal.

6. Data Editing and Key Variable Construction

SSRS monitors data quality for telephone and web surveys throughout the survey field period. Data were checked using multiple methods including: (1) a "data cleaning" procedure in which data

---

[16] The American Association for Public Opinion Research. 2011. Standard definitions: Final dispositions of case codes and outcome rates for surveys. 7th edition. Lenexa, Kansas: AAPOR. Available at: http://www.aapor.org/Content/aapor/AdvocacyandInitiatives/StandardsandEthics/StandardDefinitions/StandardDefinitions2011.pdf
[17] Formula: (Refusal + Break off) / (Eligible + Unknown eligibility * Estimated Eligibility)

processors recreated variables in order to ensure that all variables were created correctly and have the appropriate numbers of cases, and (2) an independent checking of all variables by the project director to confirm that all variables were created correctly, have the correct number of cases, and were coded according to specifications.

Ranges and logic checks were performed throughout the data collection process. After the first several days of data collection, all variables were checked to ensure that data were being collected according to designated skip patterns. Additional data checks were implemented as part of the data file development work, checking for consistency across variables and family members, and developing composite measures of family and household characteristics. At the conclusion of data collection, all variables were checked again to verify that the transfer of data file had been accomplished accurately. Constructed variables (such as whether a respondent has health insurance) were checked to ensure that data had been correctly pulled from individual items to create the composite variable.

On October 11, 2021, SSRS discovered an error in the programming logic that was resulting in the fill for the variable NEWTYPE to not display properly, potentially impacting the respondent's ability to understand the question accurately. Specifically, the question displayed as "What type of insurance MISSING covered by?" The issue was immediately fixed in the program and the study team at SSRS evaluated the impact on the collected data.

Out of the 571 cases that experienced this error, 330 cases were identified to be a 2-person household. Nothing was done about these cases: it was very likely that the respondent understood the question despite the missing fill. However, for the 241 cases with more than two people in their household data, retrieval was attempted since it was likely that the respondents didn't understand that NEWCOV was being asked in the same order as STAT. Out of those, 122 agreed to be recontacted but only 108 cases had a valid email address. Data retrieval was conducted for these 108 cases via a short web instrument asking the STAT and NEWTYPE questions. The STAT question was asked to provide context even though no issue occurred with this variable. Participants were invited and reminded about this effort via email. A total of 50 cases were retrieved.

Both the original and the retrieved data were kept in the data set and three flags were created, as follows:
- Most up to date values can be found in the variables STAT and TYPE while the original values were maintained in the variables orig_stat and orig_newtype.
- NEWTYPE_ERR=1 indicates cases that experienced this error.
- NEWTYPE_DR=1 indicates cases that agreed to be recontacted among those that experience the error.
- NEWTYPE_RETRIEVED=1 indicates cases for which the data were successfully retrieved.

For details see section 3.11 of the SSRS Methodology Report.[18] SSRS provides a raw data file and a cleaned data file at the midpoint and at the end of the field period.

SHADAC performed additional logical edits and cleaning functions in the process of creating analytic variables. For example, if individuals reported carrying health insurance through the Indian Health Service (IHS) and no other coverage, they are coded as uninsured, because IHS is typically not considered insurance coverage. Further, logical conflicts potentially created during the imputation process were corrected.

---

[18] Contact Kathleen Call at callx001@umn.edu or Health Economics Program at health.mnha@state.mn.us.

For all variables that included response options allowing text-based entry (i.e., "Other, specify") by the interviewer (RDD) or participant (ABS) (e.g., race, ethnicity, industry), respondent's answers were reviewed and data was back coded to available response options, new categories were created if appropriate, or responses were left as "other." For example, in 2019 a new category was created for "Share Plans," a prevalent response for participants providing an answer to the open-ended insurance type.

With the ABS sample, households receive multiple reminders and the option to complete in different modes (web, phone) of administration. Household members may not be aware that another resident has completed the survey. In 2021 a total of 5 people answered the survey twice. Duplicates were removed from the data file after review of the cases to determine where data were most complete.

Income Imputations

Consistent with other surveys, income has high item nonresponse (i.e., respondents choose not to answer the question). Income related measures are important to the MNHA because of their association with various dimensions of health and access. Excluding cases with missing income data could introduce bias to our survey estimates (non-responders may share certain income characteristics), therefore family income was imputed for these respondents. A second advantage of imputation is that it allows all respondents to be included in calculations involving income, such as uninsurance rates by poverty level and eligibility for public programs or Advanced Premium Tax Credits (APTC) in the individual insurance market.

Each year a majority of cases answer the open-ended question about household income or respond to a follow-up question providing a set of income ranges for those unwilling to state their income in the first question (Table 9). This strategy has been effective. In 2021 only 6 percent of respondents did not respond to the income items.

Table 8. Percentage Responding to Income Questions and Missing Data			
Survey Year	**Open-end Income**	**Income Range**	**Missing Data**
2011	77%	14%	9%
2013	77%	11%	12%
2015	76%	11%	13%
2017	77%	10%	13%
2019	69%	24%	9%
2021	71%	23%	6%

Income was imputed using a statistical procedure known as hotdeck and designed for Stata.[19] The hotdeck procedure searches for cases with complete income data (donors) based on whether they are demographically similar to cases with missing data (recipients); a donor is selected randomly from the possible set of donors.[20] Demographic variables used in this imputation include gender, age, race/ethnicity, insurance type, household size, geographic region, educational achievement of target (or primary wage earner if target is a child) and use of government financial assistance programs, such as

---

[19] The software module was designed by Adrian Mander and David Clayton at the MRC Biostatics Unit of the Institute of Public Health in the University of Cambridge, UK.

[20] A hotdeck procedure was used for imputing other missing information needed for the income imputation: gender, age, homeownership, education, employment, race, language, country of birth, and internet access.

WIC, among those responding only to the categorical income question. In addition, we used the mode of survey implementation in this process.

Age Imputations
Respondents who were not comfortable providing age data were asked a categorical age question, placing target in one of four possible age groups: 0-17, 18-25, 26-64 and 65 years or older. In 2021 age was missing for only 35 targets (less than 1 percent, compared to 7.4 percent that did not provide age data in 2019). Age was imputed using the categorical age question, sex, marriage status, and household relationships – specifically, if the target was listed as a parent or a child.

Geographic Assignment
Geographic information for the ABS sample is precise by design, yet respondents provide their county and zip code in the survey and some discrepancies were found (n=735). To determine this assignment self-reported county was used first if available. If unavailable, zip code from the abs sample was used.

Computing the Primary Source of Health Insurance Coverage
The survey asks for all sources of health insurance coverage held by the target, given it is possible to have primary and secondary coverage. The hierarchy below assigns the *primary* source of coverage for targets with reports of multiple sources of insurance coverage:

1. Public: Includes all state and federal public coverage and military (Medicare, Medical Assistance, MinnesotaCare, VA, and Military health care, TRICARE, or CHAMPUS are asked about separately in the health insurance section of the survey; H-series).
2. Employer: Includes insurance through work or union and COBRA for employees and their dependents.
3. Individual: Includes all direct purchased coverage for individuals and families. MNSure (Minnesota's state-based exchange) coverage or Healthcare.gov is specifically named in follow-up to the H-series.
4. Uninsured: Includes not reporting any coverage and those who only report sources such as Indian Health Service or health sharing plans that are not considered comprehensive health insurance coverage.

The order of the hierarchy is based on researchers' understanding of which coverage likely acts as the primary payer of health care services. For example, if an individual reports Medicare coverage and retiree coverage through an employer, then public Medicare coverage was assigned as the primary source of coverage. Beginning in 2015, we use additional questions to assign coverage. Specifically, participants who answered "yes' to MNsure and "no" to paying a monthly premium were coded as having public coverage. Participants who answered "yes" to MNsure and "yes" to paying a monthly premium were coded as having individual coverage. The different types of public coverage are not separated out in the hierarchy because respondents often experience difficulties in differentiating among the different state and federal programs.[21]

Calculation of Public Program Eligibility and Access to Employer Coverage
Questions on the MNHA related to prior year income, household composition, age, and access to employer coverage were used to determine whether the currently uninsured were potentially eligible for public health insurance programs. Eligibility for the public programs are based on factors such as pay stubs from past two months, household size, household composition, age of household members,

---

[21] Pascale J, Fertig A, Call KT. 2019. Assessing the Accuracy of Survey Reports of Health Insurance Coverage Using Enrollment Data. *Health Services Research 54(5):1099-1109*; Call KT, Davern ME, Klerman JA, Lynch V. 2012. Comparing errors in Medicaid reporting across surveys: Evidence to date. *Health Services Research,* Apr;48(2 Pt 1):652-64.

pregnancy status, disability status, length of residence in Minnesota, immigration status, access to employer coverage, and level of employer contribution. Because the survey does not ask questions income questions that mimic eligibility and does not ask question related to respondents' level of employer contribution, pregnancy, disability, or immigration status, those factors are not considered in the process of determining potential public program eligibility.

The distribution of potential access to insurance for the point-in-time or currently insured can add to more than 100 percent because some people can have access to employer coverage and still be eligible for public health insurance programs in Minnesota. Respondents with incomes low enough to qualify for Medical Assistance (MA – Minnesota's name for Medicaid), MinnesotaCare or a MNSure subsidy who also reported having employer coverage were coded as being potentially eligible for public programs and having access to employer coverage.

Measuring Race, Ethnicity, and Country of Origin
The MNHA survey contains a series of questions that are used to allow respondents to self-report race and ethnicity. Collection and aggregation of this data has changed slightly over time to maintain consistency with guidelines established by the U.S. Office of Management and Budget.[22] To determine ethnicity, each respondent is first asked, *"Is the target person Mexican, Puerto Rican, Cuban, or from another Hispanic or Latino group?"* This is followed by a question about race, asking the respondent to choose one or more races that they consider the target person to be. With the exception of the weighting process, which required a mutually exclusive race variable, race and ethnic groups are generally defined using the Census Bureau's "any race" construct.[23] An individual is coded as belonging to a specific racial or ethnic group if that race or ethnicity is reported either alone or in combination with another race or ethnicity. Individuals for whom more than one race or ethnic identity is reported are included in all reported groups.[24] For this reason, counts obtained from the "any race" construct will total more than the population total for the state and percentages will sum to more than 100 percent.

As described in Section 3 (Survey Content), participants who identify as Black or Asian in the main race question (RACE) received follow-up questions that allow for more specific cultural identification. Disaggregated data are important for understanding nuanced differences in coverage and access than previously possible. This is dependent on sample sizes and disclosure rules and may require combining multiple years of data.

Country of origin, length of time in the United States (for non-U.S. born individuals) is collected for all targets. Beginning in 2015, citizenship was also collected for non-U.S. born targets. This information was also gathered for the parents of targets under 18.

Measuring Employment
The structure of the employment questions has been consistent since 2011. Information about employment status, employment at more than one job, and total hours worked per week at all jobs is collected for all adults in the household. Additional questions are asked about the adult's primary place of employment. In addition, a separate student status question is asked of all adults under 65. This design allows respondents to more appropriately identify themselves as employed, as well as students, when that is the case.

---

[22] Office of Management and Budget, 2003, *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.* Available at: http://minorityhealth.hhs.gov/templates/browse.aspx?lvlID=172
[23] Source: U.S. Census Bureau, 2003. *U.S. Census Bureau Guidance on the Presentation and Comparison of Race and Hispanic Origin Data.* Available at: http://www.census.gov/population/www/socdemo/compraceho.html
[24] In 2019 4.2 percent of target persons were reported to have more than one race. In 2017 and 2015 respectively, 4.6 and 2.3 percent of target persons were reported to have more than one race as compared to 1.8 and 1.5 percent in 2009 and 2007 respectively.

7. Survey Weights

The goal of weighting survey data is to adjust the results to account for sample coverage problems (the difference between respondents and non-respondents) and reduce potential bias associated with differential participation in the survey. Accounting for varying probabilities of selection and response rates through the application of weights enables the survey responses drawn from statistical samples to be representative of the entire population. The 2021 MNHA used only one sample frame: an ABS frame which significantly reduced the complexity of weighting the data compared to 2019.

Two types of weights were generated: 1) base weights and 2) post-stratification weights. The base weight takes into consideration that each target's probability of selection varies by sampling stratum, and the number of people living in the household. The post-stratification weights adjust the base weights to account for key characteristics of the state's population. Specifically, to more accurately reflect the population, sample weights were post-stratified by region, age, education, race, ethnicity, nativity (US versus foreign born), home ownership, household count, web access and public coverage. Details are provided below.

Base Weights

We used the following formula to calculate the basic base weights that estimate the probability of selection for each observation for this calculation:

$$Base\ weight_{ij} = \left(\frac{Universe_j}{Completes_j}\right)\left(\frac{Number\ of\ people\ living\ in\ the\ household_i}{Number\ of\ sampling\ units\ in\ the\ household_i}\right)$$

where *i* refers to the individual observation and *j* refers to the sampling strata. This formula provides the initial estimate of weights, where each base weight indicates how many people in Minnesota are represented by each sampling unit (or target, since we only have one target per household). For the number of sampling units in the household, we used 1 for all observations as we assume that most households have only one address where they receive mail.

Sampling Adjustments

These adjustments correct for the sampling strategy that over- and under-samples specific sub-groups of the population. This generally remains consistent with prior MNHA adjustments:
1. Over-sampling children: corrects for the higher probability of selection assigned to children in households where they are present, done during the interview
2. Under-sampling elderly: use of 65+ flag in the sample frame, done prior to the interview process

Post-stratification Weights

While the base weights adjust for the known unequal probability of selection, post-stratification weights adjust for ways in which the sample's demographics and the resulting completed interviews differ from what is known about the population from which the sample was drawn. Typically, post-stratification of the MNHA surveys rely on the most current data available from the US Census Bureau's American Community Survey (ACS), which should have been to 2020 ACS. However, because data collection of the 2020 ACS was interrupted by the pandemic, the data was deemed unstable for such use by the US Census Bureau. This resulted is the use of 2019 ACS data for control totals.

Generally, the 2021 post-stratification strategy is consistent with previous MNHA surveys including slight modifications to education, age and US or foreign-born status categories in 2019 that account for small sample sizes. Some categories in the race/ethnicity variable had small sample sizes (i.e., below 100 observations), however, we decided against collapsing these categories due to the high relevance of reporting estimates for these race/ethnicity groups. In addition, we did not collapse one category in the

education variable – no High School diploma -- that had a small sample size. We decided against this mainly because people who reported not having a High School diploma had uninsurance rates that were significantly different from the next category, High School diploma or GED. These decisions can be revisited in future rounds of the MNHA as some of these sub-groups or frames decrease in the sample.

After careful consideration in 2019 a new binary variable was added to the post-stratification process: public coverage. This continued in 2021. Using administrative data from all public programs: Medicare, Medicaid, MNCare, and TriCare (includes VA), we estimated the number of people with public coverage. This estimate is the control total used for our post-stratification process. Despite the correlation of this variable with our main outcome, we introduced it as other state-level surveys include this adjustment and this provides our estimates of the uninsurance rate and the estimated distribution of health insurance coverage with some time-trending stability.

In total, we used nine variables in the post-stratification process: (1) age, (2) education, (3) race/ethnicity, (4) US or foreign-born status, (5) home ownership, (6) household size, (7) area of residence, (8) a binary variable indicating access to the internet in the household, and (9) percent enrolled in public coverage.

8. Data Analysis

Given the complex sampling design for MNHA (i.e., dual frame, age screening, selection of individuals within households), data are analyzed using Stata statistical software.[25] This software calculates unbiased estimates of standard errors and confidence intervals in the face of the MNHA's complex and multistage sampling design. The analysis specifies survey weights and strata, and, in the case when information on members of given households are used, primary sampling units (PSUs). Differences between groups and changes over time are considered statistically significant when the p-value is less than 0.05. Generally, comparisons of estimates are made between years or within year with the population total serving as the reference group.

9. Availability of Research Findings

Research results from the MNHA are made available in multiple formats including:
- Short issue briefs on a variety of topics, including an overview of key results;[26]
- Presentation slides;[27] and
- An interactive data reporting system that allows users to query survey results with great flexibility. [28]

---

[25] StataCorp. 2015. *Stata Statistical Software: Release 14.* College Station, TX: StatCorp LP; StataCorp. 2013. *Stata Statistical Software: Release 13.* College Station, TX: StatCorp LP.
[26] Issue briefs are available online at the Health Economics Program's (HEP) home page: https://www.health.state.mn.us/healtheconomics.
[27] Presentation slides can be obtained from the Health Economics Program's Chartbook series: https://www.health.state.mn.us/data/economics/chartbook/index.html.
[28] The data reporting system can be accessed at: https://mnha.web.health.state.mn.us/Welcome.action.