2019 Minnesota Health Access Survey Technical Report:
Study Design and Data Processing Methodology
SHADAC, June 2021

This report describes the data collection process and methodology behind the Minnesota Health Access Survey (MNHA), emphasizing the most recent administration of the survey completed in 2019. The 2019 MNHA represents a break in series, involving for the first time the blending of two sample frames: the introduction of an address based frame and retention of a smaller the dual landline/cell telephone frame used in all past MNHA surveys. This report describes this decision and the upshot for other aspects of the methodology. The report 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, 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 developing state health policy issues, 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.[1] This technical report focuses primarily on the 2019 MNHA, providing some cumulative data in table form.[2]

2. Sampling Goals and Methodology

Each year the sample is designed to ensure that estimates derived from a 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, for the state's 13 Economic Development Regions (target of 300 completes each) and the more populous counties (e.g., Hennepin and Ramsey), and populations of color (African Americans, Hispanics, and Asians) and American Indians in Minnesota.

In 2019, due to rising costs, falling response rates and challenges representing the Minnesota population using a random digit dial (RDD) landline and cell telephone sample[3], the decision was made to transition primarily to an address based sample (ABS) frame (target of 8,000 completes). A smaller RDD frame (target of 3,000 completes) was retained to assess comparability and reliability with earlier waves of MNHA data. Results from the dual RDD and ABS frame in 2019 will inform the viability of moving to an ABS frame for future MNHAs.

Dual RDD Frame
Using data available through SSRS[4], the RDD sample incorporated stratification by region, racial and ethnic populations. As in past years, techniques were applied to remove inactive numbers from the RDD sample.[5]

Consistent with MNHA RDD frames beginning in 2009, the 2019 MNHA included landline and cell phone numbers. Including persons who can be potentially reached through both phone frames may increase the representativeness of the sample. There is evidence that those who live in dual telephone usage households and use their cell phone for all or most of their calls (termed "cell mostly") may be different than landline-only and cell phone-only populations.[6] However, as the prevalence of cell phone-only households rises, moving to a single cell-phone only frame may be viable.[7]

In 2019 we continued to allocate a higher proportion of the total sample to the cell frame (73%, or 2,200 completes) as compared to the landline frame (27%, or 800 completes). The cell frame target increased to 2,700 during the field period to increase the likelihood of meeting sampling goals for African Americans. In the cell frame numbers associated with a prepaid cell phone flag (consistent with the sample design since 2015) were oversampled at twice the rate of their prevalence in the cell phone sample (i.e., 12% of cell

---

[1] 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.

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

[3] 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

[4] GENESYS Sampling Systems, Marketing Systems Group (MSG)

[5] The landline sample draws from banks of telephone exchanges that contained at least three listed household phone numbers (versus numbers assigned non-residential households). This increases the efficiency of the sample by increasing the likelihood of reaching an eligible household thereby reducing study costs. Both the landline and cell frames are "scrubbed" or prescreened to remove inactive numbers; interviewer time focuses on active phone lines resulting in cost savings.

[6] Lee S, Brick JM, Brown ER, Grant D. Growing cell-phone population and noncoverage bias in traditional random digit dial telephone surveys. Health Services Research. 2010; 45(4):1121-39.

[7] Ganesh N, Khare M, Ormson EA, Zeng W, Jeyarajah J, Yankey D, Zhao Z Wolter KM. Noncoverage adjustments in a single-frame cell-phone survey: Weighting approach to adjust for phoneless and landline-only households. 2014, JSM Survey Research Methods Section. Available from: https://www.amstat.org. Accessed October 2018

phone owners are prepaid owners; released 24% of sample as prepaid). This is based on evidence that numbers with the prepaid flag are more than twice as likely to be uninsured, more likely to be low income and from a population of color or American Indian community as compared to cell phone owners paying a monthly fee (Dutwin and Malarek, 2014).[8]

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. Sample was pulled from all residential records with the exception of addresses coded as vacant, seasonal (vacation), and PO boxes other than those defined as "only way to get mail," which avoids duplication of Minnesota residents in the sample selection.

The ABS used a model-based (random forest) sampling design to achieve sampling goals. In addition to stratifying the sample based on the 13 EDRs, substrata targeted households likely to have the following attributes:
1. Household member age 65+
2. Household member age 18-19
3. Hispanic Household Member
4. Low income
5. African American household member
6. Child in household
7. Spanish-speaking
8. High density American Indian census block groups (15% or more of residents)
9. High density African American census block groups (40% or more of residents)
10. Residual (auxiliary data appended; none of target groups predicted)
11. No match (none of auxiliary data could be appended)

Variables representing the information above are from the following databases: voter registration, consumer, MSG and the Census Planning Database. ABS and RDD sample are released in waves, with continuous monitoring of outcomes and productivity by frame and strata.[9] The ABS was released in two waves; wave one results allowed for model refinement in the second wave. Wave one yield was better than expected (nearly 60% of the target sample total). Details about the 2019 MNHA RDD and ABS samples are included in Appendix A.

Screening by Age
Age screening in both sample frames is necessary given the disproportionately high participation in surveys among those 65 years and older and the desire to oversample children. In the ABS and the landline portion of the RDD we continued to use a list-assisted elderly screen, removing 75 percent of cases identified with this flag. In addition, in the landline telephone (outgoing RDD and incoming ABS) and web (ABS) we screened out 66 percent of households comprised of only adults age 65 or older; in the cell frame 50 percent of households comprised of only adults age 65 or older were terminated.

In order to produce reliable estimates for Minnesota children, when selecting the target from the household (RDD and web), children under age 18 within the household are given a 50 percent higher probability of selection than adults in the household.
Age screening is not possible for those completing the paper version of the survey, yielding more 65 and older households and fewer children and young adults with this mode of data collection (see Appendix B for unweighted age distribution by survey frame and mode).

---

[8] Dutwin D, Malarek D. The use of recent activity flags to improve cellular telephone efficiency. Survey Practice, 2014; 7(1):1-10.
[9] Details about the 2019 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.

Selecting Target of Survey

Within each household that consents to participate, an adult (age 18 and older) knowledgeable about household members' health insurance is asked to complete the survey. After identifying that the household is eligible to participate in the study (e.g., Minnesota is primary place of residence), the household is enumerated and gender, age and relationship information is gathered for all household members. In the telephone and web survey, one person is then selected at random to be the target of the survey. In the paper survey, the person with the most recent birthday is the target and enumeration of the household and demographics are collected at the end of the instrument.

Screening Outcomes in 2019

Across RDD and ABS frames 598 households were terminated for not living in Minnesota or for being contacted in a vacation home, and 2,021 were terminated for being a 65+ only household. A total of 394 respondents were terminated for saying that they were unable to provide health insurance information for others in the household. Contacting businesses resulted in 2,689 terminations. In the cell frame, 956 respondents were disqualified for not being 18 years of age or older.

The proportion of child (under age 18) targets overall was 17.3 percent. This varied by sample frame (18.4% for RDD; 16.8% for ABS) and survey mode (21.3% for web; 17.9% for phone; 8.0% for paper survey).

3. **Survey Content**

Each year the majority of MNHA survey questions ask about health insurance coverage for the randomly selected individual 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, the target's health status, access to health care, affordability of care, and dental coverage is assessed, along with details about 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., number of phone lines, home ownership).

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. Here we summarize changes to the 2019 survey. A "Survey Changes Workbook" is available upon request and a comprehensive record of change in survey content up through 2015 is available in the 2015 MNHA Technical Report.[10]

Revisions

The universe statement for the question asking if everyone living in the household has the same insurance plan (INSURE) was altered so that this question is skipped when the private policy covering the target is a single (vs a family) policy.

---

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

There is continued interest in monitoring the affordability of insurance and health care services, including how much people are paying out of pocket for health care before their insurance pays for health care. The questions about high deductibles health plans (HDHP IND/FAM) among those reported to have employer sponsored, Cobra, or non-group private insurance were modified to allow for reporting the exact amount for both family and individual policies, followed by a categorical question for those not reporting an exact amount (HDHP INDb/FAMb).

In addition, respondents and targets with private insurance were asked about their confidence in their ability to pay for medical care up to the annual deductible of your current health plan in the event of major medical expenses (DEDUCT).

Additions/Substitutions

Due to rising costs of insurance and challenges to the Affordable Care Act (ACA) we added a question asking the respondents/targets with insurance if they were concerned they may not have coverage at some point in the next 12 months (LOSS). Those responding "yes" are asked the main reasons they have this concern (LOSSWHY) – this is populated with five potential reasons and other, specify.

In the event resources become available for a follow-up survey, we included a question asking about willingness to be contacted again (RECONTACT).

Finally, for the web survey (ABS sample) space was provided for the respondent to share any additional thoughts.

Omissions

While cost and affordability remains of high interest several items were deleted from the 2019 survey:

- declaring bankruptcy due to medical bills was deleted as an option for the financial burden question (FINBUR2d) due to low prevalence
- reason for forgoing medical (AFFRD_REA) was dropped due to complexity
- uninsured who reported cost as the reason for lacking coverage – the question assessing what specific aspects of the cost were a problem in gaining insurance (COST) and the question asking for the top reason (MAINCOST)
- questions (HSA) establishing the presence of a Health Savings Accounts (HSAs), or Health Reimbursement Accounts (HRAs) and the question asking if the employer contributed to either the HSA or HRA (EMPCONT)

Paper Version

The paper survey sent with the final ABS mailing created several challenges including comparability with the telephone and web version. For the paper survey skip pattern had to be simplified, in some cases items changed order, and some items had to be removed. For example, full rostering of household members occurs at the end of the survey for the paper version as the person with the most recent birthday is selected as the survey Target rather than random selection from the household roster.

In addition, the paper version reordered some questions to avoid complicated skip patterns better suited to telephone and web modes. For example, marital and employment status was asked of respondents and adult targets but not household members over age 18 in the paper version. The English and Spanish paper survey was a 24-page printed booklet with a color front page and black and white interior.

4. Survey Administration

The ABS frame allows adoption of a mixed-mode survey format, providing participants the opportunity to complete the survey by web, telephone or a paper copy. Specifically, the ABS sample received an advance letter describing the study and offering the options of accessing the survey via a secure website or calling the survey center to request a telephone interview. Participants not responding to the first mailing or reminder postcard received a third mailing with a paper copy of the survey (a modified version of the MNHA). Households in high-density Hispanic areas (based on Census data) received a bilingual letter and the final mailing included a paper copy of the survey instrument in English and Spanish.

We did not complete telephone follow-up for ABS sample elements with a matched telephone numbers based on concerns that addresses matched with telephone numbers do not adequately capture under-resourced segments of the population.

The 2019 MNHA was conducted by SSRS, an independent survey research company based in Pennsylvania. The Computer Assisted Telephone Interview (CATI) program was adapted to create an online self-administered Computer Assisted Web Interview (CAWI) version for ABS participants who selected the web mode. The web and paper version required rephrasing of questions read by respondents (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. Testing was more involved in 2019 due to multiple modes of administration and the implementation of a new program, resulting in a later start to the field period (see Table 1).

Before beginning the field period pretest were completed in all three modes. For the web (n=8) and paper (n=9) modes, the SSRS research team conducted cognitive interviews involving follow-ups and probing with respondents through Zoom, an online conferencing system. Web and paper pretest participants were Minnesota residents age 18 or over identified by the MDH and SHADAC team, but not employed by either organizations. Pretesting the telephone mode (n=40) was less extensive (no probes or follow-up) as this was the sole mode of data collection for past MNHA surveys. Telephone pretest participants were identified from the SSRS omnibus survey.

ABS pretests indicated that the invitation letters and the paper survey, including skip patterns and instructions were clear and easy to understand. Probing specific questions identified, in advance, as potentially challenging resulted in revisions to the instrument to improve clarity. Web survey feedback was also generally positive. SHADAC and MDH team members reviewed recorded telephone interviews and provided feedback before the survey was finalized. Pretests were not included in the final sample (SSRS pretest memo is available by request[11]).

Interviewer Training and Data Collection Monitoring

Prior to pretesting, standard SSRS interviewer training was conducted for the RDD frame and incoming telephone calls for the ABS frame. Specialized project training included a question-by-question ("Q by

---

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

Q") manual that explains the motivation behind each question and provides responses to common or potential inquiries from respondents. SHADAC also provided a recording of the pronunciations of Minnesota counties to help interviewers correctly identify and record the respondent's county of residence.

Calls are monitored over the course of the study (live by SSRS supervisors, and via de-identified audio recording by MDH and SHADAC staff), with intermittent interviewer training provided as needed. MDH and SHADAC were able to log into Confirmit to monitor ongoing progress between bi-monthly meetings with the SSRS team. A detailed report of data collection procedures and timeline authored by SSRS is available by request.[12]

Table 1. 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

Strategies to Maximize Response Rates

While not the only measure of quality, best practices were used to achieve the highest possible response rates. The strategies described below by sample frame.

RDD frame:
- A maximum of 8 call attempts on no answers, answering machines, and busy signals
- Calling at different times of day and days of week
- Setting callback appointments
- Resting sample after 4 call attempts
- Use of a specialized caller ID flag
- Based on a 2017 MNHA experiment
  o on the fifth attempt the caller ID was switched between a 651 area code with a MDH ID and a 612 area code with a University of Minnesota (UMN) ID
  o voicemails are left on the first and fifth attempt[13]
- Multiple attempts to convert explicit refusals using specially trained interviewers
- $5 remuneration for the cell frame to offset any costs to participants, particularly those with prepaid minutes
- Availability of bilingual Spanish/English CATI program and interviewers

ABS frame:
- Initial letter with a $2 bill as a pre-incentive
- Initial attempt is followed by a postcard reminder and a third mailing (letter plus paper survey and postage paid return envelope)
- All mailings encourage completing the survey on the web, and the option to call in to complete by telephone

---

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

[13] Call KT, Hagge SL, Simon AB, Alarcón G, Turner K, Dutwin D. Won't you please pick up? How do voicemails and call number impact the likelihood of survey response? Presented at the Annual AAPOR Conference, May 17, 2018, Denver CO and the 2017 Technical report section 5. Available upon request from Kathleen Call at callx001@umn.edu.

- Application of letterhead experiment (below) from first sample release to second sample release
- Web and paper survey available in Spanish and English

Letterhead Experiment

The switch to ABS invited an experiment to assess which organizational letterhead (with logo) elicited the highest response rate: (1) University of Minnesota, (2) the Minnesota Department of Health, or (3) a cobranded letter. Addresses in the first wave of the ABS sample release (36,193 pieces of sample) were randomly assigned to one of the treatments. As shown in Table 2, the MDH logo resulted in the highest response rate followed by the cobranded letterhead. Later waves of sample received the letterhead representing both organizations because this is the most accurate representation of the team described in the body of the letter.

Table 2. Response Rates by Experimental Condition: Letterhead and Logo, 2019			
Experimental Condition	**Minnesota Department of Health**	**University of Minnesota**	**Both Logos**
Sample Size	7,100	7,100	21,993
Response Rate	29.9%	21.0%	23.6%

Completes by Frame and Mode

Table 3 presents the number of 2019 MNHA surveys completed by frame (RDD represents 32 percent of the total completes; ABS 68 percent) and survey mode. For the ABS frame, 134 participants (2 percent) called to complete the survey by phone, 5,209 completed the web interview (66 percent) and 2,514 completed the paper copy (32 percent) sent with the third survey reminder. The number of paper completes was higher than initially expected (based on SSRS experience with Oregon survey). (See Appendix B for unweighted distribution by survey frame and mode.)

Table 3. Completed MNHA Surveys by Frame and Mode, 2019			
Mode	**Frame**		**Total***
	RDD	**ABS**	
Telephone	3,673	134	3,807
Landline	*619*		
Cell (not prepaid)	*2,534*		
Prepaid cell	*520*		
Web		5,209	
Paper		2,514	
Total*	3,673	7,857	11,530

*Totals include partial interviews (n=355).

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 has been relatively consistent over the years (see Table 4). Variation in interview length between cell and landline samples has dropped over time. In 2019, average time to complete the interview took longer for the telephone than the web survey. 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. In 2019, the web survey took 10 more minutes on average to complete in Spanish than the telephone interview.

Table 4. MNHA Average Survey Length in Minutes by Frame, Mode, and Language					
Survey Year	RDD		ABS	RDD		ABS
	Landline	Cell	Web	English	Spanish*	Spanish*
2013	18	22		20	29	
2015	19	23		22	32	
2017	22	22		22	32	
2019	26		18	18	36	46

*A total of 97, 142, 123 and 30 interviews conducted in Spanish in 2013, 2015, 2017 and 2019 respectively. In 2019, the RDD sample yielded 21 Spanish completes (0.6% of the RDD sample) and the ABS sample yielded 9 Spanish completes (0.1% of the frame).
In 2001, surveys were also completed in Hmong (32) and in 2004 surveys were completed in Hmong (85) and Somali (38). Due to the high cost of translation and the low number of surveys completed in Hmong and Somali, these languages were discontinued in 2007.

Remuneration

MNHA respondents in the cell phone sample are offered $5 remuneration for completing the interview. This is particularly important due to the decision to oversample prepaid cell phone numbers in the cell phone frame. Contact information is acquired at the end of the interview and stored in a database separate from survey responses. Although compensation is offered, not all respondents provide the contact information necessary to receive compensation, and take-up of this offer dropped between 2015 and 2017, especially for the non-prepaid sample (Table 5). Although cell phone completes represent a smaller portion of the total sample in 2019, take-up of compensation increased, especially for the non-prepaid sample.

Table 5. Percent of MNHA Cell Phone Sample Accepting Remuneration					
Cell phone sample:	2015	2017	Difference (2017-2015)	2019	Difference (2019-2017)
Prepaid	62%	59%	-3%	71%	12%
Non-prepaid	54%	46%	-8%	62%	16%
Total	55%	48%	-7%	64%	16%

IRB Approval

The study received IRB approval from MDH and the University of Minnesota. For the RDD frame consent to participate is obtained as part of the survey protocol read by trained interviewers; for the 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. RDD and ABS participants are provided telephone numbers for the University of Minnesota Research Subjects' Advocates Line should they have concerns about the study experience and Kathleen Call should they have questions about the survey. Each year records of participant contacts are recorded and coded to monitor the frequency and nature of participant inquiries (see Table 6).

Contacts from respondents is a relatively rare occurrence given the large sample size of the MNHA each year. Introducing the ABS sample in 2019 resulted in different sources of calls from potential participants. Of those making contact, 29 percent indicated they did not have a computer or had

difficulty accessing the web survey. The phone number to complete the survey by phone was provided. Another 20 percent called to say they had completed the survey; we explained that reminders were likely posted before they completed the web survey. While 17 percent of calls were to verify the legitimacy of the survey, this decreased compared to 2015 and 2017 (70 and 53 percent respectively). This may have declined as the letter and web survey direct participants to web information about the MNHA. In 2019, the portion of calls relaying concerns about their health, the cost and adequacy of their insurance coverage, and opinions about federal and state health policy more generally fell back to 2015 levels. Very few voice complaints about the survey or request we not contact them. Finally, 12 percent called indicating their web access code did not work; we explained that data collection was closed.

Table 6. Log of Contacts from MNHA Participants		2015		2017		2019	
Code	Labels	Count	%	Count	%	Count	%
1	Concerns about health care/health policy	1	4%	14	26%	3	3%
2	Legitimacy of survey, questions about survey, use of data	19	70%	25	46%	17	17%
3	Question about applying for insurance	1	4%	4	7%	1	1%
4	Complaints/do not call	3	11%	5	9%	6	6%
5	Comments about survey (+/-)	3	11%	6	11%	8	8%
6	No computer, challenges with web survey					29	29%
7	Out of sample, business or no MN resident					4	4%
8	Already completed survey					20	20%
9	Called after close of data collection; other miscellaneous					12	12%
Total comments coded		**27**	**100%**	**54**	**100%**	**100**	**100%**
Unique contacts		*27*		*43*		*98*	
Initiated by IRB		*2*		*2*		*10*	
Email versus telephone contacts		*6*		*3*		*2*	

5. Response Rates and Sample Coverage

Over time response rates have dropped for all surveys, with some evidence of having plateaued in recent years.[14] For telephone based surveys this general trend is attributable to growth in the non-contact rate (e.g., fewer people answering their phone as a result of telephone screening devices) and small growth in refusal rates (e.g., households/individuals declining to participate perhaps due to frustration with marketing, bot calls, and survey research in general).[15] Falling response rates and the

---

[14] Pew Research Center. May, 2017. What low response rates mean for telephone surveys: Telephone polls still provide accurate data on a range of social, demographic and political variables, but some weaknesses persist. Available at: http://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/. Accessed October 2018.

[15] Lavrakas PJ et al., The future of U.S. general population telephone survey research. Report from the AAPOR task force, 2017. Available at: http://www.aapor.org/getattachment/Education-Resources/Reports/Future-of-Telephone-Survey-Research-Report.pdf.aspx. Accessed June 2017.

implications for data quality is the subject of intense attention and scrutiny.[16,17] Response rates are a commonly used indicator of the quality of a survey. However, research indicates that lower response rates are not necessarily associated with greater response bias because surveys with high and low response rates demonstrate similar levels of absolute bias.[18,19,20,21]

In general terms, the response rate is the ratio of the number of completed interviews divided by the number of eligible reporting units in a sample; the cooperation rate is the ratio of all interviewed cases to all eligible cases contacted. The response rates reported below refer to AAPOR Response Rate #3[22] from 2009 to present for the blended sample (cell and landline combined), which is the equivalent of the number of completed interviews divided by the total number of eligible phone numbers.[23] 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.

Table 7. MNHA Count of Complete Surveys, Response, Cooperation and Refusal Rates				
Survey Year	Total Completes^	Response Rate*	Cooperation Rate*	Refusal Rate**
2001	27,315	67%	78%	19%
2004	13,802	59%	68%	28%
2007	9,728	43%	57%	32%
2009	12,031	45%	53%	39%
2011	11,355	44%	45%	39%
2013	11,778	48%	48%	23%
2015	11,178	35%	36%	30%
2017	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%

MNHA 2001-2007 represent landline sample frames; MNHA 2009 forward represent dual landline and cell phone sample frames.

[16] Czajka JL, Beyler A. Declining response rates in federal surveys: Trends and Implications. Mathematica Policy Research, June 15, 2016. Available at: https://aspe.hhs.gov/system/files/pdf/255531/Decliningresponserates.pdf. Accessed June 2017;

[17] Lavrakas PJ et al., The future of U.S. general population telephone survey research. Report from the AAPOR task force, 2017. Available at: http://www.aapor.org/getattachment/Education-Resources/Reports/Future-of-Telephone-Survey-Research-Report.pdf.aspx. Accessed June 2017.

[18] Groves R. Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70(5): 646-675, 2006; Groves R, Peytcheva E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly, 72(2): 167-189, 2008.

[19] Keeter S, Kennedy C, Dimock M, Best J, Craighill P. Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. Public Opinion Quarterly, 70(5): 759-799, 2006.

[20] Davern M, McAlpine DD, Beebe TJ, Ziegenfuss J, Rockwood T, Call KT. Are lower response rates hazardous to your health survey? An analysis of three state telephone health surveys. Health Services Research 45(5):1324-1344, 2010.

[21] Groves R, Peytcheva E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly, 72(2): 167-189, 2008.

[22] 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

[23] To estimate the number of eligible phone numbers among numbers with *unknown* eligibility (e.g., no answer), this rate applies the ratio of eligible to ineligible numbers among the numbers with *known* eligibility to the *unknown* numbers and includes the resultant number within the denominator of the response rate calculation.

Table 8 presents the counts of completed surveys and dispositions separately for the landline and cell phone frames from 2009 forward, when cell phone sampling was introduced. As shown, response and cooperation rates are consistently higher in the landline compared to the cell phone frame. As the cell phone frame becomes a larger portion of the total sample, the response and cooperation rates for the total sample are blended and skew toward the cell phone rates.

Table 8. MNHA Count of Complete Landline (LL) and Cell Phone Surveys, Response, and Cooperation Rates								
Survey Year	**Interview counts**			**Response Rate***			**Cooperation Rate***		
	Total^	**LL**	**Cell**	**Total**	**LL**	**Cell**	**Total**	**LL**	**Cell**
2009	12,031	9,811	2,220	45%	50%	31%	53%	58%	40%
2011	11,355	7,028	4,327	44%	48%	39%	45%	49%	40%
2013	11,778	4,952	6,421	48%	50%	40%	48%	51%	41%
2015	11,178	3,139	8,039	35%	41%	29%	36%	36%	30%
2017	12,042	3,291	8,751	29%	34%	25%	30%	35%	26%
2019 RDD	3,673	619	3,054	16%	23%	14%^	17%	24%	15%

^ 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 series where applicable in 2015/17.
LL = Landline
* Based on AAPOR RR3 response and cooperation rates, which excludes partials.
^ Prepaid cell response rate is 13.3%; Non-prepaid cell was 14.2%

6. Data Editing and Key Variable Construction

SSRS monitors data quality throughout the survey field period using range and logic checks for telephone and web surveys. SSRS provides a raw data file and a cleaned data file at the midpoint and at the end of the field period. Data were checked using multiple methods including: (1) a procedure in which data processors recreate the process of CATI/CAWI variable creation (derived from skip patterns, definitions of codes and ranges specified in the instrument) to ensure that all variables were created correctly and had appropriate numbers of cases, and (2) the project director independently checked off all SPSS variables to confirm they were created correctly, had the correct number of cases, and were coded according to specifications.

This ongoing monitoring led to the early discovery of an error in the CATI/CAWI program in October (one month in to the field period): a subset of persons (n=651) were not asked the education question. After consulting with MDH and SHADAC, SSRS attempted to obtain the information calling RDD participants or mailing the request for information to ABS participants. Information was obtained for 187 households.

Consistent with earlier MNHA surveys, SSRS performed additional checks on household composition variables for the telephone and web surveys. In general, household data remained as reported by the respondent. Cases with illogical household relationships were flagged for review by the research directors. If there was a clear and logical way to correct seemingly illogical household relationships, a change would be made to the data (example: A two-person household where the child is age 40 and the parent is 6 years of age).[24]

Cleaning the 2019 paper survey data was more complicated because participants did not always follow instructions (e.g., entering responses not required, misunderstanding skip pattern, providing extraneous information, entering data in the wrong column for household grids, etc.). As a result, SSRS delivered the paper survey data in its original form. MDH and SSRS cleaned the data to the extent possible. This experience with data cleaning revealed that the MNHA is too complex to translate to a user-friendly paper survey. Combined with evidence of duplication across interview modes, future waves of data collection should carefully consider the value of the paper mode and entertain other strategies to encourage participation in the third reminder. (See Appendix B for unweighted distribution by survey frame and mode.)

MDH and 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.

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, paper survey) of administration. Household members may not be aware that another resident has completed the survey, or may not recall having completed the web version when the paper version arrives. A total of 327 people answered the survey twice, in two of the three available data collection modes. Duplicates were removed from the data file after review of the cases to determine where data were most complete (with a preference for omitting duplicate paper survey data).

Income Imputations

Consistent with other surveys, income has the highest item nonresponse (i.e., respondents choose not to answer the question) of any of the survey items. Income related measures are important to the MNHA because of their association with various dimensions of health and our interest in estimating the proportion of the population that is uninsured but appears to be eligible for public health insurance or Advanced Premium Tax Credits (APTC) in the individual insurance market. 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

---

[24] Details about the 2019 household composition checks in the SSRS Methods Report are available by contacting Kathleen Thiede Call at callx001@umn.edu and the Health Economics Program at health.mnha@state.mn.us.

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 among the uninsured.

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 2019 only 9 percent of respondents did not respond to the income items which is consistent with the past MNHA surveys.

Table 9. 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%

Income was imputed using a statistical procedure known as hotdeck and designed for Stata.[25] 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.[26] Demographic variables used in this imputation include gender, age, race/ethnicity, insurance type, household size, geographic region, telephone interruption, educational achievement of target (or primary wage earner if target is a child) and use of government financial assistance programs, such as 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, allowing the target to be identified as a member in one of four possible age groups: a 0-17 year old child, an 18-25 year old young adult, a 26-64 year old adult, or an adult 65 years or older. For the 7.4 percent of cases that refused the initial age question in 2019 (6 percentage points higher than in 2017), 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
For the RDD frame, respondents' geographic location (county FIPS codes) is provided by SSRS's sister organizations, Marketing Systems Group (MSG). In addition, respondents provide their county and zip code in the survey. For those cases in which the GENESYS FIPS did not match the respondent provided county or zip code, the respondent provided data are used.[27] By design, geographic information for the ABS sample is precise, yet respondents provide their county and zip code in the survey and some

---

[25] 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.
[26] A hotdeck procedure was used for imputing other missing information needed for the income imputation: gender, age, homeownership, education, employment, race, country of birth, length living in the US, and phone status.
[27] For the Landline sample frame: If respondent zip code and county matched but differed from Genesys, we used respondent provided county; if respondent zip code and county did not match, we used the variable that matched Genesys; if respondent county, zip code and Genesys county did not match, we used respondent county. For the cell phone sample, respondent provided county was used.

discrepancies were found (n=279); when both web and paper versions of the survey were returned we accepted the web data.

Computing the Primary Source of Health Insurance Coverage
MNHA are asked about all sources of health insurance coverage available to them, given it is possible to have primary and secondary coverage. We calculate Minnesotans' primary source of health insurance coverage, meaning individuals reporting insurance are assigned only one type of coverage. The following hierarchy is used for determining the *primary* source of coverage for people who report access to multiple sources:

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 included in follow-up questions following the H-series.
4. Uninsured: Includes those without any coverage and those who only have sources such as Indian Health Service that is 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.[28]

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, 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

---

[28] 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.

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.[29] 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.[30] 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 of these reported groups.[31] 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.

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.

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 2019 MNHA, represents overlapping RDD and ABS sample frames and the new ABS frame introduced two new interviewing modes: the Computer Assisted Web Interview and a paper survey sent to households that did not respond to the first two invitations to participate. This transition from telephone RDD to RDD and ABS constitutes a break-in-series for the MNHA, yet we made every effort to reduce potential bias to our estimates. Our main strategy for minimizing bias was through a new weighting approach.

---

[29] 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

[30] 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

[31] 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.

In consultation with a sampling statistician (Trent Buskrik, Bowling Green State University), we tried eleven different approaches to weighting with each completed sequentially and most attempts improving on previous versions. Those that combined the RDD and ABS frames before raking the whole sample together were rejected due to the high design effects for some estimates of the main outcome: the Minnesota uninsurance rate (ranging from 4 to 12). Below we summarize the main characteristics of the weighting adopted in 2019. See Appendix C for an overview of the evaluation.

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 weight 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, nativity (US versus foreign born), home ownership, household count, and telephone usage. Finally, after the RDD and ABS frames were post-stratified independently, they were appended. Details are provided below.

Base Weights
Having introduced a new frame, we need to estimate base weights that soundly approximate the probability of selection of the sampling unit, while producing a harmonic representation of the state population. SSRS provided the ABS estimates of the base weights per strata but these were not equivalent to those used for the RDD. Thus, we needed to estimate these base weights using a similar formula. We used the following formula to calculate the basic base weights that estimates 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).

The number of sampling units in the household depends on the specific frame. For example, for the landline and cell samples this is the number of answerable landlines and cell phones in the household. For the ABS, we used 1 for all observations as we assume that most households have only one address where they receive mail. In addition, we updated the universe for the landline frame in 2019, replacing the population estimated living in the specific strata for the number of landlines in the strata (provided by SSRS).

Sampling Adjustments
These adjustments correct for the sampling strategy that over- and under-samples specific sub-groups of the population. This remains consistent with prior MNHA adjustments:
1. Elderly screener: corrects for screening out of households with only 65+ people, screener done during the interview
2. Over-sampling children: corrects for the higher probability of selection assigned to children in households where they are present, done during the interview
3. Listed sample: corrects for under-sampling of 65+ listed sample, done prior to the interview process
4. Prepaid oversampling: corrects for assigning a higher probability of selection to pre-paid cell phones, done prior to the interview process

While the cell phone base weights use all of these adjustments, the base weights for the landline and ABS frames only include the first three.

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. Post-stratification of the MNHA surveys rely on the most current data available from the US Census Bureau's American Community Survey (ACS); typically from the year just prior to the MNHA data collection year.

The RDD and ABS frames were post-stratified independently and appended after post-stratification. Generally, the 2019 post-stratification strategy is consistent with previous MNHA surveys. We updated several variables due to small sample sizes in specific categories, which causes strain on the post-stratification process, and we eliminated the age and education interaction used in the past:
- Education: this variable has now 7 categories: target is 18 or younger, no High School diploma, High School graduate or GED, some college, Associate's degree, Bachelor's degree, Post-graduate degree or studies, instead of the 3 categories used in previous years: No High School diploma, High School Diploma, and Some college or higher education.
- Age: targets who were 18 at the time of the interview were moved from their original group (18-24) and added to the younger age group (6-18) to be consistent with our education variable that reports the highest educational level of anyone 19 or older.
- US or foreign born status: the country of origin question is now asked of every target so we no longer restrict this variable to targets 3 or older, which was adjusted in the imputation process.

Some categories in the race/ethnicity variable had small sample sizes (i.e., below 100 observations); 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 for the RDD frame. 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.

In addition, after careful consideration we added a new binary variable to the post-stratification process: public coverage. 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: eight were used for both the RDD and ABS frames: age, education, race/ethnicity, US or foreign born status, home ownership, household size, and area of residence. The ninth variable differed by frame. The RDD frame used the type of phone usage in the household (i.e., whether it is landline only, cell phone only, mixed but mostly landline, or mixed but mostly cell phone[32]), whereas the ABS used a binary variable indicating access to the internet in the household.

---

[32] The control totals for telephone usage are derived from the National Health Interview Survey for Midwest; this analysis performed by JV Luke at the National Center for Health Statistics.

Pre-paid Cell Phone Post-stratification

Although this adjustment is applied before the main post-stratification process described above and also before appending the landline and cell phone frames, in 2019 we introduced a new correction to the cell frame: post-stratifying for pre-paid cell phones in Minnesota.

Because we use pre-paid cell phones as a strategy to achieve sample goals for specific, hard-to-reach sub-groups of the population, the contribution to the sample of this sub-sample can change year to year and we need to correct the weights for any potential bias that this sub-sample can introduce. The best approach is to use the rate of prepaid cell phones in the state ("frame"), however, we used the rate of prepaid phones in the cell sample as our best proxy. SSRS provided these estimates from the MNHA sample by strata. Using these estimates, we calculated the following adjustment for each strata:

$$Pre-paid\ adjustment = \frac{Prepaid\ rate\ in\ the\ state}{Prepaid\ rate\ in\ the\ sample}$$

where the prepaid rate in the sample is estimated by strata using the adjusted base weights that take into account all over- and under-sampling adjustments described above, including the prepaid over-sampling correction.

Appending Frames

The RDD and ABS frames are essentially overlapping frames of the Minnesotan population. To provide one set of 2019 MNHA estimates we both frames using a composite adjustment that corrects for the effective sample size of each frame. Initially, we used a simple weighted average based on the number of completes in the sample; the alternative used in the end was a lambda adjustment. Previous years of the MNHA have already used this composite to append the landline and cell frames. We apply that composite to append both frames into the RDD, prior to post-stratifying this frame, and then use it again to append the RDD and ABS frames after each has been post-stratified separately. This composite relies on estimating the coefficient of variation using the following formula:

$$CV_f = \frac{\sigma_f}{\mu_f}$$

where σ_f is the standard deviation of the set of weights for frame f, and μ_f is the mean of the weights for frame f.

Using this estimate, the effective sample size for each frame is calculated using the following formula:

$$ES_f = \frac{S_f}{1 + CV_f^2}$$

where S_f is the sample size for frame f and ES_f is the estimates effective sample.
Using these, we create the following composite factors:

$$CF_{RDD} = \frac{ES_{RDD}}{ES_{RDD} + ES_{ABS}}$$

$$CF_{ABS} = \frac{ES_{ABS}}{ES_{RDD} + ES_{ABS}}$$

We use these factors and multiply them times the post-stratified weights for each individual. The resulting multiplication are the final weights for 2019. These final weights are labeled as fnweight in our final dataset, which corresponds to our approach #8 (or fnweight8 in the archived datasets).

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.[33] 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;[34]
- Presentation slides;[35] and
- An interactive data reporting system that allows users to query survey results with great flexibility. [36]

---

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

Appendix A: Sample Frames

Table A-1. Landline Sample Design, 2019

Strata	Region	EDR	All Household	Sample Ratio	Target Completes	African Americans	Native Americans/ American Indians	Asians	Hispanic/ Latino	Residual (prim. Whites)
1	Northwest	1	23,662	0.16%	38	0	1	0	1	36
2	Headwaters	2	17,928	0.15%	26	0	1	0	2	23
3	Headwaters (Beltrami, Becker, Mahnomen, Cass)	2, 4, 5	9,562	0.53%	51	0	16	0	1	34
4	Arrowhead	3	78,554	0.07%	54	1	2	0	2	49
5	West Central, East Central, North Central	4, 5, 8	134,549	0.07%	95	1	1	1	3	89
7	Mid-Minnesota, Southwest	6, 10	59,457	0.12%	71	2	0	1	6	62
8	Upper MN Valley	7	12,680	0.25%	32	0	0	0	2	30
9	Central	9	83,468	0.07%	58	2	0	1	2	53
10	South Central	11	47,032	0.08%	38	1	0	0	3	34
11	Southeast	12	102,006	0.07%	72	2	0	3	5	62
12	Twin Cities	13	443,110	0.03%	149	11	1	12	11	114
13	Twin Cities (Hennepin & Ramsey)	13	115,047	0.10%	116	16	1	19	11	69
Total			1,127,055	0.07%	800	36	23	37	49	655

Table A-2. Cell Phone Sample Design, 2019

Strata	Region	All Households	Sample Ratio	Target Completes	African Americans	Native Americans/ American Indians	Asians	Hispanic/ Latino	Residual (prim. Whites)
14	Northwest	185,671	0.05%	102	1	4	1	5	91
16	Headwaters	169,869	0.11%	179	1	18	3	7	150
17	Arrowhead	468,040	0.03%	124	1	7	1	2	113
18	West Central	260,488	0.04%	97	1	4	0	4	88
19	North Central	234,547	0.05%	125	1	5	1	5	113
20	Mid-Minnesota	201,260	0.05%	93	0	2	0	7	84
21	Upper MN Valley	54,878	0.15%	85	0	0	0	4	81
22	East Central	151,470	0.06%	86	1	1	3	2	79
23	Central	418,279	0.02%	79	2	2	1	2	72
24	Southwest (Nobles)	39,736	0.10%	40	0	0	1	7	32
25	Southwest (other)	219,640	0.05%	101	1	2	2	2	94
26	South Central	359,719	0.04%	128	2	1	4	8	113
27	Southeast (Olmstead)	285,513	0.03%	81	2	1	0	6	72
28	Southeast (Other)	465,047	0.03%	124	2	2	3	5	112
29	Twin Cities	4,890,393	0.02%	756	55	10	37	55	599
Total		8,404,550	0.03%	2,200	70	59	57	121	1,893

Table A-3. ABS Sample Design, 2019

Strata	Region	EDR	HH member 65 or over	HH member 18 to 29	Hispanic HH member	Low Income	African American HH member	Child in HH	Spanish-Speaking	Residual	No Appendable Data	High density AIAN	High density African American
1	Northwest	1	74	240	14	130	13	312	26	296	91	0	0
2	Headwaters	2	149	424	51	216	30	536	26	510	131	1180	0
3	Headwaters (Beltrami, Becker, Mahnomen, Cass)	2, 4, 5	66	182	26	88	14	224	12	243	72	779	0
4	Arrowhead	3	104	326	38	138	17	446	27	389	120	323	0
5	West Central, East Central, North Central	4, 5, 8	150	392	38	209	25	542	34	505	158	446	0
7	Mid-Minnesota, Southwest	6, 10	43	139	10	48	3	208	34	168	42	0	0
8	Upper MN Valley	7	70	205	11	112	12	267	19	270	67	45	0
9	Central	9	53	187	16	67	10	256	13	230	40	312	0
10	South Central	11	50	245	21	43	40	381	26	254	48	0	0
11	Southeast	12	72	190	18	118	24	273	64	272	77	58	0
12	Twin Cities	13	48	217	26	62	14	233	30	214	58	0	0
13	Twin Cities (Hennepin & Ramsey)	13	101	342	55	123	59	475	74	416	109	104	15
Total			1,607	6,129	1,349	1,806	2,791	8,240	1,333	6,888	1,771	3,271	1,008

Appendix B: Selected Demographics by Frame and Survey Mode, Unweighted

	ACS	Frame			
		RDD		ABS*	
	Controls	LL	Cell	Web	Paper
Unweighted counts		619	3,054	5,209	2,514
EDUCATION					
N/A	23.6%	12.9%	19.6%	20.9%	7.9%
Less than HS grad	5.4%	3.4%	2.4%	1.1%	4.3%
HS grad or GED	18.4%	20.5%	19.3%	10.9%	20.5%
At least some college	52.6%	63.2%	58.8%	67.1%	67.3%
HOUSEHOLD SIZE					
1	11.5%	25.9%	22.9%	19.1%	28.7%
2	29.7%	40.2%	36.4%	35.7%	35.3%
3-5	49.5%	29.7%	36.2%	42.1%	33.0%
6+	9.3%	4.2%	4.5%	3.2%	3.1%
INTERNET ACCESS					
No internet access	6.2%	5.5%	2.4%	1.0%	10.6%
Internet access at home, on phone, or away from home	93.8%	94.5%	97.6%	99.0%	89.4%
RACE/ETHNICITY, mutually exclusive					
White	79.6%	91.3%	84.9%	88.9%	88.6%
Black	6.4%	1.8%	3.6%	2.4%	3.3%
Asian or Pacific Islander	4.8%	1.6%	2.0%	2.6%	2.2%
Other and two or more races	2.9%	1.9%	3.4%	2.7%	2.3%
American Indian	0.9%	1.3%	1.8%	1.0%	1.8%
Hispanic	5.4%	2.1%	4.3%	2.5%	1.8%
AGE					
0-5	7.8%	1.3%	0.3%	0.3%	2.7%
6-17	15.9%	6.8%	3.9%	2.5%	10.3%
18-24	8.1%	18.4%	10.8%	12.1%	19.2%
25-34	13.7%	100.0%	100.0%	100.0%	100.0%
35-54	25.3%	0.0%	0.0%	0.0%	0.0%
55-64	13.6%	0.0%	0.0%	0.0%	0.0%
65+	15.5%	1.3%	4.7%	6.3%	2.7%
HOUSEHOLD WITH 18-30 ONLY					
No		99.5%	93.1%	94.4%	97.5%
18-30 only		0.5%	6.9%	5.6%	2.5%
HOUSEHOLD WITH 65+ ONLY					
No		80.9%	88.9%	87.8%	71.5%
65+ only		19.1%	11.1%	12.2%	28.5%
EMPLOYMENT					
Not employed		45.2%	27.7%	25.0%	41.0%
Employed		54.8%	72.3%	75.0%	59.0%
INCOME					
0-99% FPG		10.8%	10.1%	5.7%	9.2%
100-199% FPG		13.3%	14.4%	8.8%	18.4%
200-299% FPG		16.8%	16.4%	13.3%	17.0%
300-399% FPG		12.1%	13.8%	15.0%	16.0%
400+% FPG		47.0%	45.4%	57.2%	39.5%

* 134 of the ABS sample completed the survey by telephone

Appendix C: Evaluation of Revisions to the MNHA Weighting Strategy, MNHA 2019

The 2019 Minnesota Health Access Survey introduced a substantial Address Based Sample. This novel frame, along with a reduced RDD sample, led to a shift of the overall sample we observed when comparing some unweighted demographic characteristics between the 2017 and 2019 MNHA. In addition, this new frame brought up two interviewing modes: the Computer Assisted Web Interview and a paper copy of the survey sent to households that did not respond to the first two requests from the ABS (initial letter and reminder postcard).

Although, these updates to the sample already constitute a break-in-series for the MNHA, we did every effort to reduce potential bias to our estimates. Thus, we required a new weighting approach that would help us minimize these potential biases. We worked in consultation with Trent Buskirk, PhD, Novak Family Professor of Data Science and Chair of Applied Statistics and Operations, at Bowling Green State University.

In total, we tried eleven different approaches to weighting. As these were done sequentially, most of these attempts were improvements on previous versions. Just a handful of these attempts, those that implied combining the RDD and ABS frames before raking the sample as a whole, were rejected and no longer pursued due to the high design effects for some estimates of the main outcome: the state's uninsurance rate (ranging from 4% to 12%).

As show in the table below, the resultant estimate of uninsurance and insurance types are consistent regardless of the approach. Following this evaluation, the team selected **fnweight8** for all analyses of the total MNHA data set.

[Note that separate weights have been created for analysis of the adult population where the Target is the Respondent - answering survey questions about their own insurance and experiences.]

Base weights

Having introduced a new frame, we needed to estimate base weights that would be a sound estimate of the probability of selection of the sampling unit, while producing a harmonic representation of the MN population. Since this was the first time we used the ABS frame, SSRS sent us estimates of the base weights per strata but these were not equivalent to those used for the RDD. Thus, we needed to estimate these base weights using a similar formula

Formula

Before any adjustment for our sampling strategy (i.e., over- or under-sampling certain sub-groups of the population), combining two frames (i.e., the effective sample size or lambda adjustment), or post-stratification (i.e., pre-paid cell phone for the Cell frame or raking), we needed to calculate basic base weights that estimated the probability of selection for each observation. We use the following formula for this calculation:

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

where *i* refers to the individual observation and *j* refers to the sampling strata. This formula gives us the initial estimate of weights, were each base weight indicate how many people in MN are represented by each sampling unit (or target, since we only take one target per household).

The number of sampling units in the household depends on the specific frame. For example, for Landlines this is the number of answerable landlines in the household. For the ABS, we used 1 for all observations as we assumed that most households would have only one address where they received mail.

In addition, we updated the Universe for the Landline frame, replacing the population estimated living in the specific strata for the number of landlines in the strata, which was provided by SSRS.

Sampling adjustments

These adjustments correct for the sampling strategy that over- and under-samples specific sub-groups of the population. This revision did not change any of the following adjustments:

- Elderly screener: corrects for the screening out of households with only 65+ people, screener done during the interview

- Kids over-sampling: corrects for the higher probability of selection assigned to children in households where they are present, done during the interview
- Listed sample: corrects for the under-sampling of 65+ listed sample, done prior to the interview process
- Prepaid oversampling: corrects for assigning a higher probability of selection to pre-paid cell phones, done prior to the interview process

While the Cell phone base weights use all of these adjustments, the base weights for the Landline and ABS frames only include the first three.

Post-stratification

As mentioned before, all weighting attempts that tried post-stratifying all frames together, RDD and ABS, produced estimates with high design effects and were discarded. Instead, we post-stratified both frames independently.
The post-stratification strategy did not change from previous years. However, we updated the set of variables because we had small sample sizes in some specific categories. As a result, we eliminated the age and education interaction from the post-stratification process and updated the following variables:
- Country of origin: since the question is now asked of every target, we no longer restrict this variable to targets 3 or older
- Education: this variable has now 7 categories: target is 18 or younger, no High School diploma, High School graduate or GED, some college, Associate's degree, Bachelor's degree, Post-graduate degree or studies
- Age: we took targets who were 18 at the time of the interview from their original group (18-24) and added them to the younger age group (6-18)

Some categories in the race/ethnicity variable still had small sample sizes (i.e., below 100 observations), but we decided against collapsing these categories due to the high relevance of these race groups.
In addition, this year we decided to add a new binary variable to the post-stratification process: public coverage. Using administrative data from all public programs, we estimated the number of people with this type of coverage. This estimate is the control total for our post-stratification process. Although this variable is highly correlated with our main outcome, many other state-level surveys use it. One of the main advantages of this update is that it offers some time-trending stability to the main outcome.
In total, we used nine variables in the post-stratification process. Eight were used for both frames, RDD and ABS: age, education, race/ethnicity. home ownership, household size, and area of residence. In addition, the RDD used the type of phone usage in the household (i.e., whether it is only landline, only cell phone, mixed but mostly landline, or mixed but mostly cell phone), whereas the ABS used a binary variable indicating access to the internet in the household.

Appending Frames
Since the 2019 MNHA does not provide two independent estimates, one from its RDD frame and the other from its ABS frame, we needed to append both frames using a composite adjustment that corrects for the effective sample size of each frame.
- We used the percent of completes by frame (# completes from RDD/ABS/# compltes) to append the RDD and ABS samples

fnweight3
- In addition, we took age and education out of the raking process as we already raked for age * education
- Instead of using the percent of completes to append the RDD and ABS frames, we decided to use a new lambda composite that reflects the effective sample size
- No further changes

fnweight4
- We added the % of the population that has public coverage (from MDH administrative data) as a raking variable
- No further changes

fnweight6

- We collapsed some categories of the age*education variable (no HS diploma and HS graduates collapsed) as we have a very small number of completes, mainly in the RDD, and this causes problems in the raking process
- No further changes

fnweight7

- We changed the set of raking variables and exclude age*education while adding age and education separately, mainly because the interaction left some categories with very few completes (mostly in the RDD frame)
- We used new categories for education in order to make it more granular, also we changed our NA category from 0-17 to 0-18 in order to avoid using the PWE's education for some cases of 18-year-olds (The variable we impute uses the PWE's education for anyone under 19)
- We had to change the age categories to be consistent with the previous change, taking 18 year olds from their original category of younger adults to the previous age group of minors
- We updated the USBRN variable to assign the country of origin for all, and not exclude 0-2 like we did in previous years
- No further changes

fnweight8

- We added a post-stratification step for the Cell sample that addresses the pre-paid sub sample. This was implemented to the final adjusted Cell base weights, and before appending the LL and Cell (before the RDD lambda)
- In addition, we found a mistake in the lambda code for RDD, where the code was calling in a base weight that wasn't final
- No further changes

Fnweight9

- Trent suggested collapsing the education categories 'No HS diploma' and 'HS diploma or GED' in order to ease the raking process from the small sample size in the 'No HS diploma' category, which is below 100 cases in the RDD frame.
- No further changes

Weights	fnweight3			fnweight4			fnweight5			fnweight7			fnweight8*			fnweight9		
Note	Raked RDD & ABS separately			Raked RDD & ABS separately, added PUBLIC to raking			Raked RDD & ABS together, added PUBLIC to raking			Raked RDD & ABS separately, added PUBLIC to raking and changed EDUC & US born (took Age*Educ out)			Raked RDD & ABS separately, poststratified for prepaid in the Cell sample and fixed an error in RDD's lambda			Raked RDD & ABS separately, <HS and HS/GED collapsed in educat6		
Frame	RDD	ABS	Total	RDD	ABS	Total	RDD	ABS	Total	RDD	ABS	Total	RDD	ABS	Total	RDD	ABS	Total
Public	36.3%	32.0%	33.3%	37.5%	37.5%	37.5%	48.1%	32.0%	37.5%	37.5%	37.5%	37.5%	37.5%	37.5%	37.5%	37.5%	37.5%	37.5%
SE	0.0134	0.0093	0.0077	0.0136	0.0099	0.008	0.015	0.0091	0.0079	0.0136	0.0097	0.0079	0.0132	0.0097	0.0078	0.0132	0.0097	0.0078
DE	2.719	3.227	3.041	2.835	3.336	3.179	3.503	2.88	3.046	2.792	3.21	3.082	2.728	3.161	3.023	2.758	3.162	3.033
Group	50.7%	59.8%	57.0%	49.9%	55.3%	53.6%	39.3%	60.5%	53.3%	49.5%	55.2%	53.4%	49.6%	55.2%	53.4%	50.2%	55.2%	53.6%
SE	0.0138	0.0094	0.0077	0.0138	0.0096	0.0078	0.0133	0.009	0.0078	0.0137	0.0094	0.0077	0.0133	0.0094	0.0077	0.0133	0.0094	0.0077
DE	2.65	2.928	2.803	2.729	2.949	2.846	2.901	2.609	2.798	2.652	2.852	2.762	2.609	2.808	2.714	2.621	2.808	2.718
Individual	4.9%	4.3%	4.5%	4.8%	3.9%	4.2%	3.9%	4.2%	4.1%	4.9%	4.1%	4.3%	4.9%	4.1%	4.3%	5.0%	4.1%	4.4%
SE	0.006	0.0032	0.0029	0.0059	0.0029	0.0027	0.0043	0.003	0.0025	0.006	0.003	0.0028	0.0058	0.003	0.0027	0.006	0.003	0.0028
DE	2.673	1.965	2.199	2.742	1.72	2.086	1.909	1.738	1.794	2.757	1.819	2.148	2.674	1.791	2.109	2.787	1.793	2.159
Uninsured	8.0%	3.9%	5.1%	7.8%	3.3%	4.7%	8.7%	3.3%	5.1%	8.1%	3.3%	4.8%	8.1%	3.3%	4.8%	7.4%	3.3%	4.6%
SE	0.0089	0.0041	0.004	0.0088	0.0034	0.0037	0.0131	0.0034	0.0051	0.0089	0.0034	0.0036	0.0085	0.0034	0.0036	0.0072	0.0033	0.0033
DE	3.782	3.704	3.749	3.873	2.915	3.428	8.37	2.738	6.091	3.763	2.896	3.362	3.552	2.851	3.233	2.838	2.73	2.783

* Weight selected for MNHA 2019 analyses.