METHODOLOGY REPORT

Massachusetts Health Survey, 2020-2021

6/18/2021

PRESENTED TO:
Blue Cross Blue Shield of Massachusetts Foundation
Kaitlyn Kenney Walsh, PhD
Jessica Gottsegen, MPH

PRESENTED BY:
NORC at the University of Chicago
Jenna T. Sirkin, PhD
David Dutwin, PhD
Michael Yang, PhD
Lindsay Liebert, MPP
Rebecca Rasky, BA
Jennifer Marek, MA

Table of Contents

Overview of the Massachusetts Health Survey

Since 2006, the Blue Cross Blue Shield of Massachusetts Foundation (the Foundation) has conducted a periodic survey, known as the Massachusetts Health Reform Survey (MHRS), to monitor key measures pertaining to health insurance coverage and health care access and affordability in Massachusetts. For the first time in 2018, in alignment with its focus on access to behavioral health services, the Foundation included several questions pertaining to access to care for mental health (MH) and substance use (SU) services in this survey.

In 2020, the rapidly increasing cost and declining response rates[1] of continuing the MHRS survey dual-frame telephone random digit dialing (RDD) approach led the Foundation to explore alternative survey approaches to gathering information on the health care needs and experiences of Massachusetts adults. As a result of the exploration, the Foundation funded a new survey with NORC at the University of Chicago that relies on multiple sample sources—including probability-based and nonprobability-based samples—and multiple modes of survey fielding (web-based and phone surveys). The survey aimed to gather information on the behavioral health care needs and experiences accessing care among Massachusetts residents, along with timely information on experiences with COVID-19. NORC designed the survey in collaboration with the Foundation and fielded the Massachusetts Health Survey (MHS) from December 2020 to March 2021.

The new survey design represents practical tradeoffs between expected survey accuracy and survey costs. Probability samples provide representative samples at relatively high cost. The inclusion of nonprobability samples lowers cost but likely increases estimation errors.[2] To keep estimation error to a minimum, NORC implemented their TrueNorth weighting method to combine the probability and nonprobability samples.[3] As discussed below, TrueNorth is a statistical procedure utilizing small area estimation methods that has been shown to reduce error from nonprobability sampling

Goals of the Sampling Strategy

In the field of survey research, costs have increased across all data collection paradigms, while response rates have been declining. This has led to an exploration of alternative methodologies and a "fit-for-purpose" paradigm of survey research to explore methods that maximize data quality for a given fixed budget. Given these considerations, NORC designed a blended-sample approach to combine area probability samples with an online convenience sample (i.e., nonprobability sample), as a way to balance a given cost with a sample that represented the overall Massachusetts adult population and provided representation within key regions and key subpopulations. The overall blended sampling strategy for this study consisted of respondents from three different sources.

- **NORC's AmeriSpeak® panel.** A national probability-based panel that is designed to be representative of the US household population. All active adult AmeriSpeak panelists (19 and over) from Massachusetts were selected for this study. **Exhibit 1** provides the distribution of the AmeriSpeak panel across counties in Massachusetts.
- **Supplemental address-based sample (ABS).** A random ABS of Massachusetts households, recruited via a postcard invitation, to provide a Massachusetts representative sample and to expand the sample size obtained from probability samples.
- **Supplemental nonprobability sample.** NORC engaged an online panel vendor to procure additional sample to expand the sample size for the study overall and for population subgroups

with high expected nonresponse rates (e.g., individuals who are racially or ethnically marginalized and individuals with lower income).

Exhibit 1. Massachusetts County Distribution in AmeriSpeak Panel

COUNTY	PERCENT
Barnstable	1.6%
Berkshire	0.9%
Bristol	7.4%
Dukes	0.0%
Essex	10.7%
Franklin	0.4%
Hampden	13.8%
Hampshire	4.3%
Middlesex	22.6%
Nantucket	0.0%
Norfolk	7.2%
Plymouth	5.8%
Suffolk	13.7%
Worcester	11.5%

The sampling strategy for the MHS combined samples from three sources to capitalize on the relative strengths of each respective frame, resulting in a design that aims to increase overall representativeness while also oversampling subgroups of interest (**Exhibit 2**).

Exhibit 2. Relative Strengths and Limitations of Each Sample Source

	Strengths	Limitations
NORC's AmeriSpeak® Panel	▪ Nationally representative probability sample ▪ Response rates superior to contemporary phone and other online data collection strategies	▪ Designed as a nationally representative probability sample, so it will not necessarily be representative of the Massachusetts population
Supplemental Address-based sample (ABS)	▪ Designed to achieve a more representative statewide probability sample than other sources and to supplement the sample size based on probability samples	▪ Expensive to field ▪ Low response rates with or without incentives
Supplemental Nonprobability sample	▪ Less expensive than probability samples ▪ Able to target specific groups of adults who were underrepresented in the AmeriSpeak and ABS samples	▪ Not representative of Massachusetts population ▪ Requires more complex survey weighting procedures

Sample Implementation

NORC's AmeriSpeak® Panel

Funded and operated by NORC at the University of Chicago, AmeriSpeak® is a probability-based panel designed to be representative of the US household population. Randomly selected US households are sampled using area probability and address-based sampling based on the United States Postal Service (USPS) Delivery Sequence File. These sampled households are then contacted by US mail, telephone, and/or field interviewers (in-person) to gain their participation in the panel. The panel participation rate for AmeriSpeak at the time of the study was 20 percent. The panel provides sample coverage of approximately 97 percent of the US household population. Those excluded from the sample include individuals with a PO Box only address, addresses not listed in the USPS Delivery Sequence File, and some newly constructed dwellings that are not yet occupied. While most AmeriSpeak households participate in surveys by web, households without internet access can participate in AmeriSpeak surveys by telephone. Households without conventional internet access that have web access via smartphones are allowed to participate in AmeriSpeak surveys by web.

For this study, we selected all active adult panelists (19 and older) from Massachusetts. We randomly selected one adult in the household for panel households with more than one active adult panel member. The AmeriSpeak Panel contained 773 active panelists from the state of Massachusetts (in 12 of 14 counties) and invited all active panelists to participate in the survey.

Supplemental Address-Based Sample

Census-tract level stratification. To best support the research goals of the study, a stratified ABS sample was drawn using 5-year American Community Survey (ACS) estimates within the Census Planning Database,[1] which includes statistics from the ACS 5-year estimates (2014-18) and the 2010 Census to provide mean demographic data for all 1,478 Census tracts and 4,985 census-block groups in Massachusetts. In consultation with the Foundation, it was decided to use Census tracts as the unit for sample stratification. The sample error for Census tracts was anticipated to be lower than for block groups because of the tracts' relatively larger size.[a] The Census Planning Database is the standard available data source for designing stratified samples by geography. NORC limited stratification to ACS variables given concerns that population distribution shifted in Massachusetts since the 2010 Census.

Variable selection to achieve representation from subgroups of interest. We oversampled Census tracts with high concentrations of demographics that often are underrepresented in survey samples and would be of general of interest for the study. This includes racial and ethnic minorities, individuals with lower income, and individuals with fewer years of education. We selected stratification variables where there was evidence of higher concentrations of these groups at the Census-tract level from our analysis of the variables in Census Planning Database.

Sample stratification using high and low strata for those variables. We developed cut points to determine "high" and "low" composite sampling strata. Overall, the "high" stratum was defined as any Census tract[b] within the Commonwealth that met mean percentages above a specified point for a given

---

[a] The ACS is cluster design with a face-to-face component and does not interview households in every tract. As such, ACS estimates are modeled where they lack actual interviews. While generally reliable models, they are subject to a certain amount of error. Even within areas where there are interviews, there will also be sampling error.

[b] Census tracts with low rates of actual households due to high institutionalization rates are excluded from high stratum.

variable (see **Appendix A**). In collaboration with the Foundation, NORC classified 12 percent of Census tracts for each variable individually into the high strata to collectively define approximately a quarter of all tracts as “high strata” tracts. Statewide, this includes 26 percent of the 1,478 Census tracts.

The cut points included in **Appendix A** represent the percent of a tract within a given variable that “cut” that variable into high or low strata. Generally, these cut points were set to attain a distribution where the high stratum would comprise approximately one-fifth of all households, which allows for effective oversampling without the creating of large weight variances. A Census tract was categorized into the “high stratum” if it met one of the following criteria: 63 percent or more of persons were racial minority and/or Hispanic/Latinx; 20 percent or more of persons 25 and older did not have a high school diploma; 84 percent or more of persons 25 and older did not have a college degree; 23 percent or more of persons were below poverty; or 8 percent or more of persons in the civilian labor force were unemployed.

Maximize geographic representation. As part of the ABS design, we assessed the percent of interviews, by county, that we expected to occur in AmeriSpeak, based on the average characteristics of a Massachusetts sample. We took the appropriate reciprocal approach with ABS, so together the distribution of the combined AmeriSpeak and ABS would be more consistent with the ACS by county.[c] While 22.5 percent of Massachusetts households fall within the high strata, we sampled 51 percent of households from high strata. Notably, based on an analysis of the Census Low Response Score,[4] a metric that estimates survey response to Census surveys, we anticipated that 41 percent of interviews would come from the high strata due to greater expected non-response in high strata compared to low strata households. **Exhibit 3** provides the supplemental ABS sample allocation to the low and high stratum by county relative to the population in the low or high stratum; we also include the 2019 5-Year ACS by county for comparison.

Across each county, we oversampled the high strata by a factor of 2.5, (i.e., the probability of selecting a household in the high strata is 2.5 times higher than it would be in a simple random sample). We oversampled by a factor of 2.0 in counties that already contain a relatively high share of their population in the high strata (Nantucket, Essex, Hampshire, and Worcester counties) and a lower factor for counties with a very high share of population in the high strata (Bristol and Hampden set at 1.75; Suffolk set at 1.5). While this did increase the design effect (DEFF = 1.35) due to base weight variations, we expected that post-stratification weighting (i.e., raking) would generate less variance than utilizing AmeriSpeak without supplemental samples, given the expectation of a more representative sample due to the stratification.

Exhibit 3. Supplemental ABS Sample Design among High and Low Strata by Massachusetts County

County	ACS 5-Year (2019)*	Supplemental ABS Sample	Within County ABS Sample Allocation			
			Sample Allocated to High Stratum	Population in High Stratum	Sample Allocated to Low Stratum	Population in Low Stratum
Suffolk	11.8%	12.5%	80%	53%	20%	47%
Plymouth	7.2%	7.0%	36%	14%	64%	86%

[c] We could accomplish similar balancing using the Massachusetts EOHHS Regions (six regions in total), but we do not have this metric in the AmeriSpeak database and thus are balancing by county.

Middlesex	23.1%	23.4%	25%	10%	75%	90%
Bristol	8.3%	7.5%	61%	35%	39%	65%
Barnstable	3.6%	4.3%	18%	7%	72%	93%
Dukes	0.2%	0.4%	0%	0%	100%	100%
Nantucket	0.1%	0.2%	40%	20%	60%	80%
Norfolk	10.1%	12.4%	21%	8%	79%	92%
Essex	11.2%	12.0%	48%	24%	52%	76%
Worcester	11.8%	12.4%	41%	20%	59%	80%
Franklin	1.2%	1.5%	36%	14%	64%	86%
Hampshire	2.3%	1.4%	46%	23%	54%	77%
Hampden	6.9%	2.5%	69%	40%	31%	60%
Berkshire	2.1%	2.5%	43%	17%	57%	83%

***Source:** American Community Survey (ACS). 2019 ACS 5-Year Public Use Microdata Sample. Available at: https://www.census.gov/programs-surveys/acs/microdata/access.html

Supplemental Nonprobability Sample

The AmeriSpeak and ABS probability samples were supplemented with a nonprobability sample from the Dynata online opt-in panel.[d] The purpose of the Dynata sample was to increase the study sample size in a cost-effective manner. Dynata invited panelists who are 19 years of age or older and reside in Massachusetts to participate in the survey; these panelists received a link to the web-based survey. To help to reduce weight variation and potential bias in the final overall sample, Dynata targeted certain groups where more complete surveys were needed, including residents who identified as Black and Hispanic/Latinx, had relatively lower household income, and were over 65. Otherwise, Dynata targeted a demographically balanced respondent sample—i.e., the demographic distribution of the respondent sample was designed to approximate the distribution of the Massachusetts household population by age, race and ethnicity, and household income.

Survey Fielding

The field period was December 10, 2020 to March 12, 2021. We conducted a soft launch with the supplemental nonprobability sample on December 10, 2020 to ensure that the data were captured as anticipated in the online instrument. The AmeriSpeak and Wave 1 of supplemental ABS samples were launched following the December 10th soft launch, and the second wave of supplemental ABS was launched on January 20, 2021. We closed the fielding period on March 12, 2021.

The two-wave supplemental ABS approach enabled the research team to measure yield and eligibility in Wave 1 and to make associated adjustments in Wave 2, if necessary, to achieve the final sample. After Wave 1, which had a very low response rate (discussed below), we did not make any adjustments in the strata sample fractions for the second wave since there was concern that the low response rate reflected national mail delays in December 2020.[5] To assess potential mail delivery issues, we launched Wave 2 from two different locations. We mailed half from NORC's vendor in Indiana and half directly from a Boston post office to assess whether there were any differences in delivery timeliness and survey response. Changing the location of the mailing did not impact the response for Wave 2 (response rates are detailed in **Exhibit 5**). Given that Wave 2 performance did not improve over Wave 1, we did not field any additional ABS sample after we completed Wave 2. In collaboration with the Foundation, it was

[d] The Dynata vendor does not share state-level panel size.

decided to field additional supplemental nonprobability sample, targeting specific demographic groups to increase our response for key population subgroups (e.g., by race/ethnicity, household income, age).

NORC used a multi-modal approach to field the survey. We programmed the survey using NORC's Voxco software for both computer-assisted web interviewing (CAWI) and computer-assisted telephone interviewing (CATI) survey modes. Both the CAWI and CATI options were available for Amerispeak and supplemental ABS samples and allowed survey participants to select the mode that was most convenient. Amerispeak and ABS respondents received a link to the survey with a unique pin and a phone number to call if the respondent preferred to complete the survey by phone. The supplemental nonprobability respondents were only offered a CAWI option. We translated the CAWI and CATI survey into Spanish for residents who preferred to complete the survey in Spanish on the internet or by phone; respondents received a prompt asking if they preferred to take the survey in English or Spanish.

All CATI interviewers received training in advance of the fielding period. Interviewer training is continuously conducted and covers basic interviewing skills, including reading questions and answer categories verbatim, neutral probing, pacing, and how to use NORC's software. At the conclusion of training, interviewers complete a proficiency test and administer an interview under supervision in English and/or Spanish before they are certified to conduct data collection in that language.

Survey Content

The research team designed an instrument to capture the Foundation's research question:

1. What percentage of adults in Massachusetts expect to need behavioral health care (inclusive of mental health (MH) and substance use (SU) care) services for themselves or a family member in the coming months?
2. What percentage of adults in Massachusetts report they, or a close family member, have sought behavioral health care services for themselves or a family member?
3. Among those who have sought services, what was the experience like in finding a provider?
4. Among those who have sought services, were they able to connect with care? If not, why not?

The survey included questions on need for and receipt of MH and/or SU care and questions about facilitating factors and barriers to seeking or receiving care for the respondent and, with less detail, close family members. **Exhibit 4** provides an overview of the survey domains for this study. We included questions about overall health and mental health (adult respondents only), health insurance status (for adult respondents and children/stepchildren under 19), and the COVID-19 impact on the respondent's living arrangement and employment experiences.

We drew questions from established survey instruments and made adaptions as necessary for the CAWI/CATI format and time-period relative to COVID-19. As part of the survey design, we explored items from the following surveys and scales:

- American Community Survey (ACS)
- Boston University COVID-19 Survey
- Consumer Assessment of Healthcare Providers and Systems Experience of Care & Health Outcomes Survey (CAHPS ECHO)
- General Health Questionnaire (GHQ-12)
- Global Application of Individual Needs-Short Screener (GAIN-SS)
- Kessler Psychological Distress Scale (K6 or K10)
- Massachusetts Health Interview Survey (MHIS)

- Massachusetts Health Reform Survey (MHRS)
- Medical Outcomes Survey (MOS)
- National Health Interview Survey (NHIS)
- Patient Satisfaction Questionnaire (PSQ)
- Symptom Checklist-10
- UC Irvine/Amerispeak COVID-19 Survey

Like all survey-based research, the MHS relies on self-reported information. The quality of the data depends on the survey respondent's ability to understand the questions and the response categories, to remember the relevant information, and to report the information accurately. We expect the quality of the information reported to be better for more recent circumstances and for events with greater saliency. Problems with recall are more likely for events that are more distant in time, while problems with misreporting are more likely for sensitive questions (e.g., need for MH and/or SU care).

Cognitive testing. Prior to the beginning of the field period, we conducted cognitive interviews to ensure that the questions were interpreted as the research team intended. The cognitive testing identified additional response options related to MH and SU access and COVID-19 impact. In addition to the cognitive testing, the research team and the Foundation tested the instrument to identify improvements related to transitions between domains and the structure of complex matrix question.

Exhibit 4. Survey Domains

Domain	Questions about Respondent	Questions about Selected Close Relatives	Questions about Selected Child/Stepchild under Age 19	Content
(1) Need for MH/SU Care	✓	✓	✓	need for care ■ type of need (MH and/or SU)
(2) Process of Navigating/ Receiving MH/SU Care	✓		✓	time period of need (pre/post March 2020) ■ seeking care ■ reasons for not seeking care ■ receiving care ■ reasons for not receiving care ■ identifying a provider ■ timeliness of care ■ care setting
(3a) Anticipation of Need for Future MH/SU Care and Use	✓	✓	✓	anticipation of need for MH and/or SU care in the next six months
(3b) Overall Health Assessment	✓			usual source of care ■ self-assessment of overall health and MH ■ alcohol and cannabis use
(4) COVID-19 Context	✓			living arrangements and economic circumstances ■ COVID-19 exposure
(5) Health Insurance	✓		✓	health insurance status ■ gaps in coverage ■ type of health insurance ■ burden of out-of-pocket health care costs
(6) Demographics	✓			age ■ gender ■ race/ethnicity ■ county of residence ■ education attainment ■

and Socioeconomics				marital status ■ family size ■ family income ■ employment status

Survey Completion and Response Rate

Exhibit 5 includes sample performance information for the AmeriSpeak and supplemental ABS sample, including the number sampled/invited, eligible panelists, the incidence or eligibility rate (i.e., those who are eligible to complete the study among the invited sample), survey interviews completed, and response rate for each sample. The AmeriSpeak response rate (6.5%) calculation is based on the recruitment and retention rate for the full panel. While the response rates are low, they reflect declining response rates among all modes of household survey administration and are only one measure of survey quality.[6,7] The supplemental nonprobability sample completes are collected by our opt-in vendor, Dynata, which does not provide sample performance metrics. The total number of completed interviews was 1,719 across all three samples; 28 (1.6%) were completed in Spanish.

To be categorized as a completed survey, respondents answered at least 30 questions in the questionnaire. We conducted additional data cleaning and excluded surveys for respondents designated as speeders and straightliners. Speeders are those who complete the questionnaire too quickly (where duration, in minutes, is less than 33% of the median duration). Straightliners are those who answer multiple questions in a row with the same answer.

Exhibit 5. Sample Performance Summary, AmeriSpeak and Supplemental ABS

	Sampled/ Invited	No. Panelists Eligible for Interview	Incidence/ Eligibility Rate	No. Survey Interviews Completed	Panel Recruitment Rate*	Panel Retention Rate*	Response Rate
AmeriSpeak	773	773	100%	307	20.2%	80.9%	6.5%
Supplemental ABS Wave 1	9000	8764	97.4%	86	NA	NA	0.98%
Supplemental ABS Wave 2	4500	4239	94.2%	40	NA	NA	0.94%
Supplemental Nonprobability Sample				1286			

*Note: Reported recruitment and retention rates for AmeriSpeak are for the full national panel.

Weighting

The MHS sample design features a combination of probability and supplemental nonprobability samples. Nonprobability samples provide a lower cost supplement to probability samples; however, the quality of the data is less optimal given biases due to selection and unknown coverage. Therefore, it is important to develop pseudo weights that will help to reduce potential biases in the combined sample estimates. Weighting is a standard procedure utilized in sample surveys to account for probabilities of selection into a sample that might deviate from those for a simple random sampling approach; weighting is also used to reduce nonresponse bias and to balance completed interviews to known population benchmarks such as age, gender, educational attainment, and race/ethnicity.

Survey researchers and practitioners have proposed and implemented various methods for generating pseudo weights for nonprobability samples to support population estimates. Commonly used methods include calibration, propensity weighting, statistical matching, and doubly robust inference.[8,9] Since 2017, NORC statisticians have conducted extensive research to review and compare these weighting methods through simulations and case studies.

Our unique approach, TrueNorth, is a hybrid calibration weighting method that combines probability and nonprobability samples using small area estimation methods. Under the calibration method, survey weights are adjusted to match population benchmarks (typically from the ACS and other large-scale national surveys). TrueNorth builds upon the calibration method but introduces small area estimation modeling to achieve greater bias reduction. Small area estimation is a model-based method for improving estimation for subpopulations, called small areas. A small area may be defined by geographic, demographic, or socioeconomic variables. For example, one small area may be defined as male, Hispanic/Latinx or non-Hispanic Black, 65 and older, and with a bachelor's degree or above. The small area models under TrueNorth are used to derive model-based estimates for substantive survey variables by small areas. Along with Census demographic benchmarks, these small area estimates serve as additional calibration benchmarks under TrueNorth, hence hybrid calibration. Therefore, the TrueNorth weights will not only reproduce Census demographic benchmarks but also reproduce the small area estimates for the selected response variables. NORC research shows that TrueNorth can greatly reduce estimation bias, especially for survey variables that exhibit large biases associated with the nonprobability sample.[10,11]

Like other commonly used weighting methods for combined probability and nonprobability samples, TrueNorth does not remove all the bias. Data users may wish to take that into account when reporting power and precision. For example, in reporting 95% confidence intervals, a z-score larger than 1.96 may be used to account the extra uncertainty introduced by the nonprobability samples. Based on simulation studies, NORC uses a z-score of 2.11 for constructing 95% confidence intervals in TrueNorth studies.

The development of TrueNorth weights for this study involved the following steps:

1. Calculation of probability sample weights
2. Calculation of nonprobability sample weights
3. Development of small area models
4. Hybrid calibration for the combined probability and nonprobability samples

Calculation of Combined Probability Sample Weights

We first computed the base weights for the AmeriSpeak and ABS samples to account for the sample selection probabilities under the respective sample design. AmeriSpeak base weights are the final AmeriSpeak panel weights divided by the probabilities of the AmeriSpeak panelists being selected into the study sample. ABS base weights are the inverse of their probabilities of selection, adjusted for the number of eligible adults in the sampled residence.

Next, we inflated the base weights associated with the survey respondents to compensate for sample members that failed to complete the survey. For the AmeriSpeak samples these nonresponse adjustments were conducted separately within weighting classes defined by age, race/ethnicity, gender, and education:

- **Age:** 19-34, 35-49, 50-64, 65-74, 75 and older
- **Race and ethnicity:** Hispanic/Latinx or Black (non-Hispanic); all other
- **Education:** Some college or less, bachelor's degree or above
- **Gender:** male, female

A full cross of these variables leads to 40 weighting classes. However, to avoid excessive weight variation, we collapsed some classes associated with Hispanic/Latinx or Black (non-Hispanic), ending up with a total of 33 weighting classes.

For the ABS samples these nonresponse adjustments were conducted within weighting classes defined by high-low strata (as described in the sample design section above) and county:

- **Strata:** High, low
- **County:** Barnstable, Berkshire, Bristol, Dukes, Essex, Hampden, Hampshire, Middlesex, Nantucket, Norfolk, Plymouth, Suffolk, Worcester

A total of 22 weighting classes were used after collapsing classes for counties with small populations. We then combined the two sets of nonresponse adjusted base weights based on the number of complete surveys from each sample source.

Finally, we calibrated the nonresponse adjusted weights to key population benchmarks through a raking procedure so that weights sum to population benchmark totals. Operating only on the marginal distributions of the population, raking is an iterative proportional fitting procedure. To begin, the weights are adjusted to match the benchmark distribution of the first raking variable. Next, the weights are further adjusted to match the benchmark distribution of the second raking variable. The adjustment process is repeated for every raking variable, going back to the first variable for further adjustments as needed, until the weighted distribution of all of the raking variables matches their specified population benchmarks. The raking variables are defined below:

- **Race/Ethnicity:** Hispanic/Latinx, white (non-Hispanic), Black (non-Hispanic), other (non-Hispanic)
- **Age group:** 19-34, 35-49, 50-64, 65-74, 75 and older
- **Education:** less than high school, high school or equivalent, some college, bachelor's degree or above
- **Gender:** male, female
- **Massachusetts Executive Office of Health and Human Services Regions:** Western, Central, Northeast, MetroWest, Southeast, and Boston
- **Family Income:** MassHealth family income categories adjusted for family size, as defined by the health insurance unit (HIU)

We derived all the demographic benchmark data from the 2019 American Community Survey (ACS). For family income, we defined family as the health insurance unit (HIU) from the State Health Access Data Assistance Center (SHADAC). The HIU is an economic unit that consists of those members of a household who would likely to be eligible as a group for family health insurance coverage, or whose resources (i.e., income) would be considered in determining eligibility for public coverage.[12] The HIU differs from the Census definition of a family or a household that is used in many surveys. For example, household units – as the name suggests – consist of all individuals residing in a sampled household, regardless of interrelationships among members. The Census Bureau's definition of family includes all related members of a household. This would include parents and their children along with any other related individuals who are living with them (e.g., grandparents, adult siblings, aunts, uncles, nieces, nephews, cousins).

We used the ACS 2019 microdata from IPUMS to derive the family income benchmarks for adults in the state of Massachusetts. The ACS IPUMS data includes HIU variables from SHADAC. Since Massachusetts-specific income cutoffs from SHADAC were not available, benchmarks were created from ACS using the SHADAC HIU variables and national income cutoffs. The benchmarks were created to align with the questionnaire income cutoffs as seen below (**Exhibit 6**).

Exhibit 6. Designated FPL Cutoffs in the Questionnaire

HIU Persons	HIU Income (133% FPL)	HIU Income (300% FPL)	HIU Income (400% FPL)	HIU Income (500%)
= 1	$ 17,000	$ 38,300	$ 51,100	$ 63,800
= 2	$ 22,900	$ 51,700	$ 69,000	$ 86,200
= 3	$ 28,900	$ 65,200	$ 86,900	$ 108,600
= 4	$ 34,900	$ 78,600	$ 104,800	$ 131,000
= 5	$ 40,800	$ 92,100	$ 122,700	$ 153,400
= 6	$ 46,800	$ 105,500	$ 140,600	$ 175,800
= 7	$ 52,700	$ 118,900	$ 158,600	$ 198,200
= 8+	$ 58,700	$ 132,400	$ 176,500	$ 220,600

To create the ACS income benchmarks, we first calculated the HIU level income by summing the total reported income within each HIU ID number. This HIU level income was then merged to each ACS record by HIU ID. Therefore, HIU income was treated as an individual characteristic just like other raking variables. We then classified each ACS record into one of five income levels based on HIU level income and number of HIU persons using the thresholds from the questionnaire noted in Exhibit 8—less than or equal to 133% FPL, greater than 133% and less than or equal to 300% FPL, greater than 300% and less than or equal to 400% FPL, greater than 400% and less than or equal to 500% FPL, and greater than 500% FPL. For example, if person A had a calculated HIU level income of $20,000 and the HIU persons is one, then person A was assigned to income group 2 (greater than 133% FPL and less than or equal to 300% FPL). The population benchmark for the income levels was derived using the person-level weight from ACS. The population total for this five-level income variable was incorporated into the weighting as the family income raking variable (see **Exhibit 7** for the weighted distribution).

Exhibit 7. Person-level Weighted Distribution of HIU Income, Raking for Massachusetts Adult Population (19 and over)

≤133% FPL	20.0%
> 133% and ≤ 300% FPL	16.6%
> 300% and ≤ 400% FPL	8.5%
> 400% and ≤ 500% FPL	6.7%
> 500% FPL	48.2%

Calculation of Nonprobability Sample Weights

As there is no known “sample design” to the nonprobability sample, all members in this sample started with a base weight of one. The base weights were raked to the same set of population benchmarks as those used to calibrate the probability samples in the previous step.

Development of Small Area Models

The objective of small area modeling is to generate additional benchmarks for TrueNorth weighting to calibrate weights to in order to reduce potential bias. We conducted small area modeling in the following steps:

- First, we identified a set of key survey response variables using a machine learning approach called gradient boosted tree modeling. We modeled these variables under TrueNorth small area estimation. The machine learning approach identifies the key variables that are associated with the largest differences across the probability and nonprobability sample and are also highly correlated with other response variables. The survey response variables used were:
    - **Q57:** In the last 12 months, have any of your close relatives needed mental health and/or substance use (i.e., alcohol and/or drug use) care?
    - **Q39:** How is your overall health now compared to this time last year?
    - **Q46:** Since the COVID-19 related changes that started in March 2020, are you drinking... More often; About the same; Less often?
- Second, to support domain-level small area modeling, we defined a set of domains in the data, where each domain was a specific, relevant subgroup for data analysis and reporting. We used age, gender, race/ethnicity, and education to define these domains. The domains are, in effect, the unique cells in a four way cross-tabulation of these variables. The variables and categories used include:
    - **Age:** 19-64; 65 and over
    - **Gender:** male; female
    - **Race/Ethnicity:** white (non-Hispanic); all other
    - **Education:** high school education or less; some college or more
- Third, we fit domain-level small area models for each of the identified response variables identified earlier using weighted domain-level estimates as inputs and incorporating external data sources (e.g., ACS) as potential predictors in the models.
- Fourth, we used the fitted small area models to generate predicted values for each domain on each modeled response variable.

Hybrid Calibration

In this step, we developed hybrid calibration weights for the combined probability and nonprobability samples. The probability and nonprobability sample weights were first combined using a combination factor that is proportional to the relative sample size of the probability and nonprobability samples. Then, we calibrated the combined weights to the same ACS benchmarks and predicted values for each domain from the small area models. The final TrueNorth weights summed up to the total number of Massachusetts adults 19 years of age or older per ACS. **Exhibit 8** provides the sample size and associated margin of error by geographic, demographic, and socioeconomic characteristics; **Exhibit 9** includes the sample size by key outcomes.

Exhibit 8. MHS Sample Sizes by Region, Demographic, and Socioeconomic Characteristics

	Sample Size*	Margin of Error
Statewide	1719	4.0%
Geographic Region*		

Western	243	12.1%
Central	223	11.5%
Northeast	328	8.7%
MetroWest	382	7.9%
Boston	219	11.5%
Southeast	324	9.2%
Age of Respondent		
19-39	484	7.3%
40-64	785	5.6%
65 and over	450	8.4%
Gender		
Male (including transgender male)	690	6.1%
Female (including transgender female)	994	5.3%
Something Else	35	32.9%
Race/Ethnicity of Respondent		
Asian (Non-Hispanic)	69	19.1%
Black/African American (Non-Hispanic)	68	18.2%
Hispanic/Latinx	199	12.7%
White (Non-Hispanic)	1316	4.5%
Other (Non-Hispanic)	17	42.7%
2+ (Non-Hispanic)	50	23.0%
Family Income		
≤133% FPL	191	11.2%
> 133% and ≤ 300% FPL	390	8.6%
> 300%	1138	4.7%
Health Insurance Coverage of Respondent		
Yes, Covered Currently	1562	4.1%
No, Not Covered Currently	116	15.6%
Don't Know	41	38.7%
Insurance Coverage		
Fully Insured	1561	4.1%
Partial or Full-year Uninsured	157	14.5%

Note: Geographic regions are the Massachusetts Executive Office of Health and Human Services Regions: https://matracking.ehs.state.ma.us/eohhs_regions/eohhs_regions.html
* Sample sizes are unweighted counts

Exhibit 9. MHS Sample Sizes by Key Outcomes

	Sample Size*	Margin of Error
Q1. In the last 12 months, did you need mental health and/or substance use (i.e., alcohol and/or drug use) care for yourself?		

Mental Health and Substance Use Care	431	8.0%
Mental Health Care	401	8.4%
Substance Use Care	91	16.8%
Did Not Need Any Mental Health or Substance Use Care	1288	4.6%
Q57. In the last 12 months, have any of your close relatives needed mental health and/or substance use (i.e., alcohol and/or drug use) care?		
Mental Health and Substance Use Care	393	7.8%
Mental Health Care	321	8.5%
Substance Use Care	145	13.3%
Did Not Need Any Mental Health or Substance Use Care	1326	4.6%
Q63. In the last 12 months, did your child/stepchild need mental health and/or substance use (i.e., alcohol or drug use) care?		
Mental Health and Substance Use Care	43	22.4%
Mental Health Care	38	23.9%
Substance Use Care	7	52.5%
Did Not Need Any Mental Health or Substance Use Care	4	76.8%

* Sample sizes are unweighted counts

Item Nonresponse

We imputed the item missing data for some key demographic and socioeconomic variables to support weighting adjustments and data analysis. For most variables, where item nonresponse was low, we relied on random imputation. This included household size, race/ethnicity, gender, and marital status. For household size, race/ethnicity, gender, and marital status of respondents, we imputed all values of "don't know," "skipped on web," "refused," and "missing." In addition, values of "something else" for gender were imputed to either "male" or "female" for weighting purposes. Imputations were randomly drawn from the distribution of the characteristic for the observed portion of the sample. **Exhibit 10** lists the percent of observations imputed for each variable.

The hot deck imputation method was used to impute missing data for family income, where item nonresponse was 8.6%. In hot deck imputation, each missing value on a response variable is replaced with a value reported by a respondent (donor) who is similar to the respondent associated with the missing value (recipient). To identify a donor, we sort the data file by a set of variables so respondents that are similar with respect to these variables are next to each other. For each recipient, the donor is identified as the respondent that is closest to the recipient on the sorted file. The variables that are used to identify the donor are divided into two types: class variables and sort variables. The donor and the recipient have to match (i.e., have the same values) on the class variables, but they don't have to match on the sort variables. Within sort variables, the recipient's assigned value will be based on the donor with the closest value, if not an exact match). The hot deck imputation approach for family income was based on the following respondent characteristics:

- **Class variables:** gender, work status, race/ethnicity, census tract poverty level (three poverty levels were defined based on data from the ACS)
- **Sort variables:** age, marital status

Exhibit 10. Percent of Observations Imputed

Gender	2.1%
Race/Ethnicity	0.2%
Family Size	0.1%
Marital Status	0.1%
Family Income	8.6%

The distribution of variables before and after imputation is shown in **Exhibit 11** below.

Exhibit 11. Distribution of Variables Before and After Imputation

	Before Imputation	After Imputation
Gender		
Male (including transgender male)	40.9%	41.0%
Female (including transgender female)	59.1%	59.0%
Race/Ethnicity		
White (Non-Hispanic)	76.6%	76.6%
Black (Non-Hispanic)	4.0%	4.0%
Other (Non-Hispanic)	1.0%	1.0%
Hispanic/Latinx	11.5%	11.6%
Multi-Race (Non-Hispanic)	2.9%	2.9%
Asian (Non-Hispanic)	4.0%	4.0%
Family Size		
One Person	45.6%	45.7%
Two Persons	40.1%	40.1%
Three Persons	7.3%	7.3%
Four Persons	5.2%	5.2%
Five Persons	1.3%	1.3%
Six or More Persons	0.4%	0.4%
Marital Status		
Married, Living with Spouse	47.0%	47.0%
Married, Not Living with Spouse	2.2%	2.2%
Widowed	7.5%	7.5%
Divorced	10.6%	10.6%
Separated, Living with Spouse	0.9%	0.9%
Separated, Not Living with Spouse	1.1%	1.1%
Never Married	30.8%	30.8%
Family Income		
≤133% FPL	10.9%	11.1%
> 133% and ≤ 300% FPL	22.7%	22.7%
> 300% and ≤ 400% FPL	15.7%	15.7%
> 400% and ≤ 500% FPL	13.1%	13.0%
> 500% FPL	37.6%	37.5%

References

1. 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*. 2016.
2. QuickFacts United States. United States Census Bureau; 2020. https://www.census.gov/quickfacts/fact/table/US/PST045219.
3. NORC at the University of Chicago. *NORC's TrueNorth Calibration tool for probability and nonprobability samples.* Chicago, IL March 3 2020.
4. Erdman C, Bates N. The Low Response Score (LRS) A Metric to Locate, Predict, and Manage Hard-to-Survey Populations. *Public Opinion Quarterly*. 2017;81(1):144-156.
5. Badger E, Bui Q, Sanger-Katz M. The Postal Service Survived the Election. But It Was Crushed by Holiday Packages. *New York Times.* Jan. 19, 2021.
6. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public opinion quarterly*. 2008;72(2):167-189.
7. Davern M. Nonresponse rates are a problematic indicator of nonresponse bias in survey research. *Health services research.* 2013;48(3):905.
8. Elliott MR, Valliant R. Inference for nonprobability samples. *Statistical Science.* 2017:249-264.
9. Valliant R. Comparing alternatives for estimation from nonprobability samples. *Journal of Survey Statistics and Methodology*. 2020;8(2):231-263.
10. Ganesh N, Pineau V, Chakraborty A, Dennis JM. Combining Probability and Non-Probability Samples Using Small Area Estimation. *JSM Proceedings, Survey Research Methods Section.* 2017:1657-1667.
11. Yang M, Ganesh N, Mulrow E, Pineau V. Estimation methods for nonprobability samples with a companion probability sample. Paper presented at: Joint Statistical Meetings; Jul. 28, 2018; Vancouver, Canada.
12. Alarcon G, Fried B, Hest R. *Update of SHADAC's Health Insurance Unit.* Minneapolis, Minnesota: The State Health Access Data Assistance Center (SHADAC); Jul. 2020.

Appendices

Appendix A. Select Variables for Sampling Stratification and Associated Cut Points for Sample Strata

Variable Description	CPD Variables (from the CPD Tract Level Dataset)	Variable Definition (from the CPD Tract Level dataset)	Variable Transformation for the Stratification	Cut point (%)
Percentage of persons non-white or Hispanic	pct_NH_white_alone_ACS_14_18	The percentage of the ACS population that indicate no Hispanic origin and their only race as "White" or report entries such as Irish, German, Italian, Lebanese, Arab, Moroccan, or Caucasian	Percentage: 100 percent minus pct_NH_white_alone_ACS_14_18	63.1
Percentage of persons 25 and older who do not have a high school diploma Percentage of persons 25 and older who do not have a college degree	pct_Not_HS_Grad_ACS_14_18	The percentage of the ACS population aged 25 years and over that are not high school graduates and have not received a diploma or the equivalent	Low education defined as either: Percentage (25 and older) without a high school diploma pct_Not_HS_Grad_ACS_14_18 OR Percentage of population (25 and older) without a college degree= 100 percent minus pct_college_ACS_14_18 from the	20.3 (without a high school diploma) 83.5 (without a college degree)
	pct_college_ACS_14_18	The percentage of the ACS population aged 25 years and over that have a college degree or higher		
Percentage of persons who are below poverty	pct_Prs_Blw_Pov_Lev_ACS_14_18	The percentage of the ACS eligible population that are classified as below the poverty level given their total family or household income within the last year, family size, and family composition	n/a	22.9
Percentage of persons ages 25-64 in the civilian labor force who are unemployed	Civ_unemp_25_44_ACS_14_18	Number of civilians between the ages of 25 and 44 at the time of the interview who are unemployed in the ACS	Percentage of persons ages 25-64 in the civilian labor force who are unemployed	8.0
	Civ_unemp_45_64_ACS_14_18	Number of civilians between the ages of 45 and 64 at the time of the interview who are unemployed in the ACS		
	Civ_labor_25_44_ACS_14_18	Number of civilians between the ages of 25 and 44 at the time of		

		the interview who are in the labor force in the ACS		
	Civ_labor_45_64_ACS_14_18	Number of civilians between the ages of 45 and 64 at the time of the interview who are in the labor force in the ACS		

Source: United States Census Bureau. Census Bureau Planning Database. United States Census Bureau, 2020. Available at: https://www.census.gov/topics/research/guidance/planning-databases.html