MATCH 2023-2024 Methods Report

West Virginia Mountain State Assessment of Trends in Community Health (MATCH) Survey

2023-2024 Fielding

Prepared for: West Virginia Department of Human Services

Prepared by: West Virginia University Health Affairs Institute

For inquiries, please contact: Amanda Jones at aajones@hsc.wvu.edu

October 31, 2025

Executive Summary

Introduction

The Mountain State Assessment of Trends in Community Health (MATCH), a public health survey, was established in West Virginia (WV) through a partnership between the West Virginia Department of Human Services (DoHS) and West Virginia University (WVU) Health Affairs Institute (HAI). MATCH is a biennial, multi-modal (i.e., web, paper, phone), cross-sectional, population-based health survey that collects information on WV residents aged 18 and older living in non-group housing. The survey was designed to provide substate, population-level health data estimates to address data gaps within WV.

The 2023-2024 MATCH survey was administered to WV adult residents between October 13, 2023 and July 31, 2024 throughout 55 WV counties.

Sample

An address-based sampling (ABS) approach was employed. The primary ABS frame was a United States Postal Service (USPS) frame of WV residential addresses. This frame was merged with address level information from a WV Medicaid roster — providing additional details on potential race and socio-economic status of a respondent at the address. A total of 88,000 addresses were sampled and surveyed in three batches: soft launch, batch 1, and batch 2. The soft launch was a small (1,000 addresses) sample without stratification or oversampling used to test infrastructure prior to a larger release. Batch 1 (43,500 addresses) and batch 2 (43,500 addresses) were stratified at the county or subcounty level with oversampling for smaller counties, race (non-White), and poverty status. Batch 2 was sampled adaptively based on batch 1 responses.

Questionnaire

The primary aim of designing the questionnaire was to better understand the health of West Virginians to support matching of community health needs with resources designed to meet those needs. To accomplish this, each question was tailored to address a key aspect of health status and/or behavior, healthcare access, substance use, mental health, or other factors influencing health. Some changes were made to the 2023-2024 survey based on the special interests of the DoHS, such as the prevalence of vaping. Some questions and responses related to COVID-19 were removed as the information was no longer pertinent. Other changes included renaming question section headings to be more inclusive.

Data Collection

Up to four mailings were sent to selected participants that directed them to take the survey online or over the phone with an interviewer. The third mailing included a Pen and Paper Instrument (PAPI).

To increase the survey response rate, pre-incentives of $2 were included in the first mailing and participants who completed the survey were given $10. Outreach efforts were widespread across the state as well as targeted directly to areas with lower response rates, and more likely

to have non-White individuals and individuals on Medicaid. The overall response rate for the 2023-2024 MATCH survey was 20.5%.

Data Processing and Analysis

Collected data were source selected (CATI/CAWI vs PAPI), assessed for completeness, standardized, and weighted to make analyses representative down to the county-level for as many outcomes as possible.

Table of Contents

List of Tables

List of Figures

Document Acronyms

The following acronyms are used throughout this document:

Acronym	Definition
AA	Black or African American
AAPOR	American Association for Public Opinion Research
ABS	Address-Based Sampling
ACS	American Community Survey
ASPE	Assistant Secretary for Planning and Evaluation
CAWI	Computer-Assisted Web Interview
CATI	Computer-Assisted Telephone Interview
CDSF	Computerized Delivery Sequence File
Census	United States Census Bureau
Covid-19	Coronavirus Disease 2019
DH	West Virginia Department of Health
DoHS	West Virginia Department of Human Services
FPL	Federal Poverty Level
HAI	West Virginia University Health Affairs Institute
IRB	Institutional Review Board
IVR	Interactive Voice Response
MATCH	Mountain State Assessment of Trends in Community Health
MSG	Marketing Systems Group
NHSR	Not Human Subjects Research
NORC	National Opinion Research Center at The University of Chicago
PAPI	Pen and Paper Instrument
PC	Project Coordinator
P.O. Box	Post Office Box
PSU	Primary Sampling Unit
SES	Socioeconomic Status
USPS	United States Postal Service
UWE	Unequal Weighting Effect
WV	West Virginia
WVU	West Virginia University

1 Introduction

1.1 Project Overview

The Mountain State Trends in Community Health (MATCH) survey is a state-based population health monitoring system designed to capture various health indicators (e.g., health status, health behavior, and social determinants) at the state, region, and county level in West Virginia (WV). MATCH was created through a collaborative effort between the West Virginia University (WVU) Health Affairs Institute (HAI) and the West Virginia Department of Health and Human Resources (DHHR). In January 2024, DHHR officially reorganized into the three departments and, since that time, HAI has been primarily collaborating with West Virginia Department of Human Services (DoHS) for MATCH. The 2021-2022 survey served as the baseline for this second iteration, or wave, of the 2023-2024 MATCH survey. The data obtained from wave 2 can be used by DoHS staff, researchers, academics, legislators, policymakers, healthcare providers, insurance providers, and the public to better understand the health of West Virginians and match community health needs with resources specifically designed to meet those needs. Moreover, data from wave 2 of the survey can be compared to wave 1 and future survey iterations to monitor health trends over time. In addition, data could be used to measure the impact of health programs and initiatives across the state.

The 2023-2024 MATCH survey was fielded between October 13, 2023 and July 31, 2024. The 2023-2024 survey utilized an address-based sampling (ABS) design that was built using United States Postal Service (USPS) Computerized Delivery Sequence File (CDSF) comprising a licensed list of registered addresses. This design had a two-step global sampling procedure that first selected the household (as given by the address) and then selected one person from the household to take the survey. In the mailing, the adult in the household with the most recent birthday, age 18 years or over, was asked to complete the survey. The randomly selected respondents were invited to participate in one of three ways:

- Through a computer-assisted web interviewing (CAWI) module using Voxco software, where respondents could complete the survey on a computer, tablet, or smartphone.
- Pen and Paper Instrument (PAPI) survey, where respondents could complete the paper survey and then mail it back in a self-addressed, postage-paid envelope.
- Respondents who did not wish to complete the survey by web or by paper could call the phone number provided in the survey invitation materials to complete the survey by phone using the computer-assisted telephone interview (CATI) module in Voxco.

Spanish language option was made available for web survey if the respondents preferred Spanish. If a respondent wished to complete the survey but was physically or mentally unable to, a proxy respondent could complete the survey on the individual's behalf.

The survey implementation protocol used a sequential "push-to-web" design that utilized up to four mailings sent to sample members. Sample members were first invited to complete the survey by web with the initial invitation letter; this letter contained an explanation of the survey, a hyperlink to a landing page, and a unique survey access code. The second invitation, a postcard, also invited sample members to complete the survey. Sample members who did not respond to the first two invitations were then sent a paper survey packet in the third mailing and the fourth mailing consisted of a reminder postcard. Finally, a fifth mailing, a postcard, was sent

to a targeted sample of 5,000 invited participants either located in underperforming counties or considered as unrepresented demographics.

The ABS design produced 88,000 addresses which were sampled in three batches (soft launch, batch 1, and batch 2) without replacement. The mailing to the first 1,000 addresses, referred to as the soft launch, was sampled prior to stratification using a simple random design. Batch 1 contained 43,500 addresses. Sampling allocation was based on the 2021-2022 MATCH survey yield rate for attaining the minimum target survey completions from each county in West Virginia. This first batch of addresses was stratified to oversample smaller counties, non-White individuals, individuals enrolled in Medicaid, and individuals of low socioeconomic status (SES). Like batch 1, batch 2 contained 43,500 addresses. Minimum county targets and stratum were adjusted based on the batch 1 and soft launch survey completion yield rate. In addition, this second batch also oversampled counties with lower population sizes, non-White individuals, individuals enrolled in Medicaid, and individuals of low SES, but had an emphasis on the strata with the highest Black or African American yield rates.

Throughout the process, HAI coordinated with DoHS and other stakeholders, conducting extensive partner outreach to promote the survey during the fielding timeline. This report describes the procedures utilized to design and implement the MATCH survey, and to meet the objectives defined by HAI and DoHS.

1.2 Design Overview and Important Changes from 2021-2022 Fielding

The 2023-2024 MATCH survey covered numerous topics pertaining to health and health coverage of WV residents. Topics included the following:

- Your Health
- Healthcare Access
- Mental Health
- Lifestyle
- About You
- Household Finances
- Substance Use
- Other Topics (e.g. physical activity and satisfaction with life)

The 2023-2024 MATCH sampling frame was built from two distinct sources for addresses: The West Virginia (WV) Marketing Systems Group (MSG) ABS frame and a West Virginia Medicaid enrollment roster from March 1, 2022, to July 31, 2023. Additional details on the construction of the final sampling frame are found in *Section 2: Sampling*.

The key sampling differences between the 2023-2024 MATCH and 2021-2022 MATCH surveys include the use of a unified ABS frame, adjustments to the sampling strategy, and optimization of targeted mailings. Instead of two separate frames as used in the 2021-2022 survey, the 2023-2024 survey combined an initial ABS frame with addresses from a one-year Medicaid roster to create a unified frame, which aimed to reduce the design effect. Furthermore, adjustments were made to intra-county strata allocation to further mitigate county-level design effects. The oversampling strategy was also expanded from solely Black or African American households to include non-White households, thereby increasing the number of counties with oversampling.

Additionally, an adaptive design approach was introduced in batch 2, leveraging data from the soft launch and batch 1 to determine the optimal number of addresses needed to achieve target completion rates for counties and Black or African American respondents. These changes collectively aimed to improve the stability of estimates across a broader range of counties, independent of their population size.

1.3 Institutional Review Board (IRB) Determination

The WVU Office of Research Integrity and Compliance reviewed the MATCH survey instrument and a plan for data sharing with respect to human subjects' protection and determined that the proposed project was considered Public Health Surveillance that informs public health programs in the state. The MATCH project was classified as Not Human Subjects Research (NHSR) because it did not meet the criteria for original research involving human subjects as defined by the WVU Office of Research Integrity and Compliance. This designation allows certain research projects to be excluded from Institutional Review Board (IRB) review and approval.

As per the WVU Office of Research Integrity and Compliance, "Public Health Surveillance constitutes data and information to assess and characterize the burden and distribution of adverse health events, prioritize public health actions, monitor the impact of control measures, and identify emerging health conditions that may have a significant impact upon population health."

2 Sampling

2.1 Synopsis of Sample Design

An address-based sampling (ABS) design was employed for 2023-2024 MATCH. There were 88,000 addresses sampled from an optimized sampling frame using a probability-based design with known probabilities[1] of selection.

Oversampling occurred in the following areas:

- Counties with smaller populations of households,
- Addresses more likely to contain non-White persons (non-White), and
- Addresses linked with a Medicaid beneficiary at any time in the previous calendar year or otherwise flagged as a household with low SES status (Medicaid/low SES).

A single person within the sampled household address was selected to complete the questionnaire. In the mailing material, the adult in the household, age 18 years or over, with the most recent birthday, was asked to complete the survey.

The 88,000 addresses were sampled in three batches:

- (Soft Launch) The first 1,000 addresses were sampled prior to stratification using a simple random sample design.
- (Full Launch — Batch 1) The first batch contained 43,500 addresses. Allocation was based on the 2021 MATCH yield rates for attaining county targets. Using stratification, oversampling was performed for counties with fewer addresses and for non-White and Medicaid/low SES households.
- (Full Launch — Batch 2) The second batch contained 43,500 addresses. County targets (minimum 220 for the allocation design yield) and stratum allocation were adjusted based on the batch 1 and soft launch yield rates. Oversampling was performed for counties with fewer addresses and for non-White and Medicaid/low SES households with emphasis on the strata with the highest Black or African American yield rates.

2.2 Objectives of the Sample Design

The primary objective of the sample design was to maximize the number of stable estimates[2] for as many counties as possible while maintaining an acceptable overdistribution. A secondary objective was to maximize the number of stable estimates for Black or African American and non-White demographics when it could be done with acceptable impact on the primary objective. A tertiary objective was to allow subset analysis for low-income households (between 200%-400% of the federal poverty limit (FPL) or lower-income households (less than 200% of the FPL).

[1] Precisely, estimable probabilities based on sampling frame and sampling design.
[2] An estimate of a proportion is considered stable if the number of respondents for the question is at least 50 and the relative standard error is less than 0.3.

In the state of WV, the less populated counties[3] tend to have fewer residents from different racial backgrounds and the balancing of primary and secondary objectives forced creative solutions in the adaptive design allocation outlined in *Section 2.6.4 Batch 2 Sampling*.

2.3 Key Differences with 2021 MATCH Fielding

There are four primary differences in sampling between 2023-2024 MATCH and 2021-2022 MATCH. The first difference is that in the 2023-2024 MATCH, an initial ABS frame was combined with the March 1, 2022 – July 31, 2023 Medicaid roster addresses to form a single ABS frame, as opposed to two separated frames in the 2021-2022 MATCH. This change was made to lower design effects. The second key difference was that in order to lower county-level design effects, the intra-county strata was altered. The third difference is that non-White households, as opposed to only Black or African American households, were oversampled. This was done to increase counties with oversampling. The final key difference was the use of information obtained from the soft launch and batch 1 to determine the number of addresses needed to achieve target completions in counties and for Black or African American respondents. This adaptive design was used to increase the probability of a county obtaining the desired number of completions. Altogether, the changes were implemented to increase the number of stable estimates for the largest number of counties (irrespective of the county's intrinsic population).

2.4 Sampling Procedure

The global sampling procedure was a two-step process (1) for selection of household as given by the address and (2) selection of one person within the household.

2.4.1 Sampling of Household

The primary sampling unit (PSU) was the household as identified by its address. The selection of addresses from the sampling frame was done differently depending on the sample release (soft launch, batch 1, batch 2) and without replacement. The procedure for each release is described in the respective section below. Addresses from an earlier sample were excluded from subsequent sampling, i.e., soft launch addresses were excluded from batch 1 sampling and soft launch and batch 1 addresses were excluded from batch 2 sampling.

2.4.2 Selection of Respondent within the Household

The 2023 MATCH used the "most-recent-birthday" method to select the respondent within the household. The survey instrument instructed the person receiving the invitation as such:

"The survey should be completed by the adult age 18 or over in your household who had the most recent birthday."

2.5 Sampling Frame

This section describes the development of the 2023-2024 MATCH ABS sampling frame from initial sources through stratification of the frame.

2.5.1 Target Population

The target population for 2023-2024 MATCH included the total population of noninstitutionalized adults (aged 18 years or older) residing in residential households in WV. Excluded from this population were individuals who met at least one of the following criteria:

[3] McDowell notwithstanding.

- Less than 18 years of age,
- Resided in penal, mental, or other institutions,
- Resided on military bases covered by dedicated central office codes,
- Resided in other group quarters such as dormitories, barracks, convents, or boarding houses (with 10 or more unrelated residents),
- Did not read or speak English or Spanish well enough to be interviewed, or
- Had physical or mental impairments or a language barrier that prevented them from completing an interview (as defined by the interviewer or by another member of the household) if a knowledgeable proxy was not available

2.5.2 Sources for ABS Sampling Frame

The 2023-2024 MATCH sampling frame was built from two distinct sources for addresses. The primary source was the WV Marketing Systems Group (MSG) ABS frame[1], and the supplemental source was addresses from a March 1, 2022 – July 31, 2023 roster of WV Medicaid recipients. This section describes the two sources.

WV MSG ABS Frame

The WV MSG ABS Frame was built using USPS Computerized Delivery Sequence File (CDSF) as the core data source[1]. The CDSF contained various address types, including city style, rural route/highway contract, PO boxes, seasonal, long-term vacant, and drop points. The USPS CDSF did *not* include certain address types, including facilities classified as group quarters. Examples of group quarter locations include prisons, military installations, colleges/universities, and nursing homes.

The WV MSG ABS frame had geocoding information, including county, and indicators on whether an address was likely to contain a person who is Hispanic/Latino(a), Asian, or Black or African American. It was a mixture of proprietary and modeled data. Derived and modeled estimates for number of adults, number of children, and household income were also provided.

Additional details are found in *Appendix B, section 7.1.1*.

One-Year[4] Medicaid Roster

The one-year Medicaid roster provided by the Bureau of Medical Services (BMS) at DoHS is a collection of all adult persons (≥ 18 years of age) who were enrolled on Medicaid at any time from March 1, 2022, to July 31, 2023. The roster contains select demographic information including race, ethnicity, and death date.

Additional details are found in *Appendix B, section 7.1.1*.

2.5.3 Development of Source Frames

The section describes the manipulations on the source frames in preparation for the final frame construction.

NORC ABS Frame Construction

The National Opinion Research Center at The University of Chicago[2] (NORC) team constructed a stratified sampling frame using the WV MSG ABS frame that covered the residential household

[4] The "one-year" roster includes a 17-month window of enrollees.

population. The WV MSG frame started with a list of 1,023,356 addresses. From the initial list the following exclusions were applied:

- Mailing addresses whose FIPS code was not in WV
- Seasonal addresses, including seasonal education addresses
- Business, primary business with residence, and general delivery addresses
- Drop point addresses with more than four units
- PO Boxes which were not the only way to receive mail
- Duplicate records of street address, city, state, five-digit zip code

There were 807,949 addresses after applying the exclusion criteria.

An initial address level characteristic of likely low SES was established if the modeled household income was below 200% of the FPL or the number of associated persons (adults and children) according to the 2023 poverty limit guidelines of Assistant Secretary for Planning and Evaluation (ASPE).[3] *Table 7-1: Construction of Low SES Based on ASPE's 200% FPL Guidelines* provides information on the construction of the low SES characteristic.

NORC Medicaid Frame Construction

The NORC team started with the one-year Medicaid roster containing 505,293 records. Cases whose physical address was outside of WV, cases denoted homeless in their physical address, and cases who were deceased were excluded.

These exclusions resulted in 483,173 WV Medicaid cases with an address. Since multiple Medicaid enrollees could have the same address, all the Medicaid enrollee addresses were collapsed into unique addresses and the number of Medicaid enrollees at a given address was determined. Addresses with over 12 distinct Medicaid IDs were excluded from the frame under the assumption that these were addresses of group quarters.

Address level demographics were established for likely Hispanic/Latino(a), Asian/Pacific Islander, Black or African American, and American Indian or Alaskan Native[5]. An address was indicated as having the respective demographic *if any individual* associated with the address on the Medicaid roster was identified as being of that demographic group. For example, if one person indicated they were Hispanic/Latino(a), then, the address was designated with likely Hispanic/Latino(a) demographic. A household could have multiple demographics assigned.

2.5.4 2023 MATCH ABS Sampling Frame Construction

The NORC ABS frame and the NORC Medicaid frame were merged into the 2023-2024 MATCH ABS sampling frame.

The two frames were joined using the dplyr R package[4] on both mailing and physical addresses (excluding P.O. Boxes) using an exact match on each address data element: house number, pre-directional, street name, street suffix, post-directional, secondary address, city, state, ZIP5, and ZIP4. The P.O. Box addresses were then matched using the same address data elements as the initial step using "P.O. BOX" as the street name and the box number as the house number. Unmatched Medicaid addresses were matched on a reduced set of address data elements

[5] Other is not used because it contained unknown in the source data, see *Appendix A, Section 6.2, Common Statistical Terms for Survey Sampling* for details.

excluding city and ZIP4. Finally, for Medicaid cases with a match on both mailing addresses and physical address, the mailing address was prioritized to avoid duplication of Medicaid information in the frame.

There were 233,237 Medicaid addresses associated with an address in the NORC ABS Frame. Unmatched addresses in the NORC Medicaid frame were excluded from the 2023-2024 MATCH ABS sampling frame to ensure a robust frame with locatable addresses. Medicaid data are an administrative data source that may contain inaccurate/outdated addresses and records.

As a drop point is associated with multiple addresses, addresses with drop points were replicated with a unit indicator (in general Unit 1,…,Unit 4) to have a record for each address associated with the drop point suggested by the frame. With these additional addresses, the constructed final 2023 MATCH ABS sampling frame contained 809,823 addresses. The increase in records from the NORC ABS frame reflects the replication of drop point addresses with multiple units.

An address level demographic characteristic of likely non-White and an assessment of Medicaid/low SES were established for each address. The address was assigned to non-White characteristic if any of the WV MSG ABS frame characteristics of Hispanic/Latino(a), Asian, or Black or African American were indicated or if the address matched to a Medicaid Frame address and one of the Medicaid frame address level characteristics of Hispanic/Latino(a), Asian/Pacific Islander, Black or African American, American Indian, or Alaskan Native was indicated. An address was assessed as Medicaid/low SES if NORC ABS frame had an indication of likely low SES, below 200% FPL, or the address matched to a Medicaid frame address.

2.5.5 2023 MATCH ABS Sampling Frame Stratification

The county was the primary stratification level. This section contains the development of potential substrata, the selection of substrata, and global rules of allocation.

Potential County Substrata

To realize the final stratification, the 2023-2024 MATCH ABS sampling frame addresses were first stratified into one of the 55 WV counties using geocoded information. Depending on the distribution of the characteristics of interest in the county as determined by the sampling frame variables associated with the addresses, likely non-White and Medicaid/low SES, the county was then further stratified. The substrata considered were as follows:

- Non-Medicaid/low SES, White[6], abbreviated as NMW[7], Medicaid/low SES, White, abbreviated as MW
- Non-Medicaid/low SES, non-White abbreviated as NMNW
- Medicaid/low SES, non-White, abbreviated as MNW

County Sub-Stratification Eligibility and Prioritization

There were three eligibility rules for county sub-stratification. The first rule was that if the county had fewer than 8,000 addresses in the frame, then the county was *not* eligible for sub-

[6] White was the absence of non-White designation and was not a designation of likely White per se.
[7] For the abbreviations, the following conventions were used: N=non, M=Medicaid (or low SES), W=White.

stratification. Secondly, if the county had at least 8,000 addresses but fewer than 10,000 addresses in the frame, then the county was eligible for sub-stratification, but not eligible for oversampling by non-White or Medicaid/low SES status within the county. Lastly, if the county has at least 10,000 addresses in the frame, then the county was eligible for sub-stratification and oversampling for non-White or Medicaid/low SES status within the county.

Of the 55 counties, 24 were not eligible for sub-stratification, six were eligible for sub-stratification only, and 25 counties were eligible for both sub-stratification and oversampling.

Substrata prioritization: The prioritization of the substrata was the non-White strata followed by Medicaid/low SES indicator. Given the prioritization, out of the 15 possible county sub-stratifications built from the four strata off the Medicaid-race stratification only six were considered. To avoid small cell sizes, a substratum was required to have at least 1,500 distinct addresses in the 2023-2024 MATCH ABS sampling frame. A given *substrata eligible county* was assigned to the highest sub-stratification type to which it was eligible.

In order of prioritization, the six county sub-stratifications considered are:

- Type A (four substrata): NMW, NMNW, MW, and MNW
- Type B (three substrata): (NMW or MW), NMNW, and MNW
- Type C (three substrata): NMW, MW and (NMNW or MNW)
- Type D (two substrata): (NMW or MW) and (NMNW or MNW)
- Type E (two substrata): NMW and (NMNW or MW or MNW)
- Type F: no sub-stratification

Selection of County Substrata

In the 2023-2024 MATCH ABS sampling frame, there were at least 1,500 addresses in the NMW and MW substrata for all counties that were sub-stratification eligible[8]. This simplified the sub-stratification prioritization rules in two ways. No county satisfied the conditions for Type B and Type D sub-stratification, and the sub-stratification type assignment was based exclusively on its eligibility for sub-stratification (eligible, non-eligible) and its address counts in the two non-White substrata (NMNW and MNW).

Due to this simplification, an eligible county was assigned to the sub-stratification:

- Type A, if there were at least 1,500 addresses for both the NMNW and MNW substrata
- Type C, if the NMNW and MNW substrata had 1,500 addresses combined but either the NMNW or the MNW substrata had fewer than 1,500 addresses
- Type E, if the NMNW and MNW had fewer than 1,500 addresses combined, and all non-sub-stratification-eligible counties were implicitly assigned county sub-stratification
- Type F (i.e., no sub-stratification)

Of the 55 counties, four were Type A, eight were Type C, 19 were Type E, and 24 were Type F resulting in 102 total strata. *Table 2-1: WV Counties by Sub-Stratification Eligibility and Type* shows the assignment to sub-stratification and oversampling eligibility by county. Additional

[8] The counties which had a substrata of the form NMW and MW with fewer than 1,500 addresses are Calhoun, Doddridge, Gilmer, Pendleton, Pleasants, Pocahontas, Tucker, Tyler, Webster, and Wirt. These all have fewer than 4,000 addresses in the count – not eligible for sub-stratification.

details on the county and stratum address counts are found in *Appendix B, Section 7.1.3, Substratum Types*.

Table 2-1: WV Counties by Sub-Stratification Eligibility and Type

County	Sub-Stratification Eligibility	Sub-Stratification Type
Barbour	Not eligible	Type F
Berkeley	Eligible with oversample	Type A
Boone	Not eligible	Type F
Braxton	Not eligible	Type F
Brooke	Eligible with oversample	Type E
Cabell	Eligible with oversample	Type A
Calhoun	Not eligible	Type F
Clay	Not eligible	Type F
Doddridge	Not eligible	Type F
Fayette	Eligible with oversample	Type E
Gilmer	Not eligible	Type F
Grant	Not eligible	Type F
Greenbrier	Eligible with oversample	Type E
Hampshire	Eligible but without oversample	Type E
Hancock	Eligible with oversample	Type E
Hardy	Not eligible	Type F
Harrison	Eligible with oversample	Type C
Jackson	Eligible with oversample	Type E
Jefferson	Eligible with oversample	Type C
Kanawha	Eligible with oversample	Type A
Lewis	Not eligible	Type F
Lincoln	Eligible but without oversample	Type E
Logan	Eligible with oversample	Type E
Marion	Eligible with oversample	Type C
Marshall	Eligible with oversample	Type E
Mason	Eligible with oversample	Type E
McDowell	Eligible but without oversample	Type E
Mercer	Eligible with oversample	Type C
Mineral	Eligible with oversample	Type E

County	Sub-Stratification Eligibility	Sub-Stratification Type
Mingo	Eligible but without oversample*	Type E
Monongalia	Eligible with oversample	Type A
Monroe	Not eligible	Type F
Morgan	Eligible but without oversample	Type E
Nicholas	Eligible but without oversample	Type E
Ohio	Eligible with oversample	Type C
Pendleton	Not eligible	Type F
Pleasants	Not eligible	Type F
Pocahontas	Not eligible	Type F
Preston	Eligible with oversample	Type E
Putnam	Eligible with oversample	Type C
Raleigh	Eligible with oversample	Type C
Randolph	Eligible with oversample	Type E
Ritchie	Not eligible	Type F
Roane	Not eligible	Type F
Summers	Not eligible	Type F
Taylor	Not eligible	Type F
Tucker	Not eligible	Type F
Tyler	Not eligible	Type F
Upshur	Eligible with oversample	Type E
Wayne	Eligible with oversample	Type E
Webster	Not eligible	Type F
Wetzel	Not eligible	Type F
Wirt	Not eligible	Type F
Wood	Eligible with oversample	Type C
Wyoming	Not eligible	Type F

Table Notes: *Although Mingo had 10,005 addresses in the 2023 MATCH ABS Sampling Frame, its total at the time of assessment for oversample eligibility using a preliminary sampling frame was lower.

Global Rules of Allocation

For each batch[9] of the full launch sample (batch 1, batch 2), there was a two-step process for allocation of addresses: county allocation of addresses and stratum allocation of addresses.

[9] The Soft Launch uses a simpler process.

County (CY) allocation of addresses *(Acy)*: The approach for county level allocation of addresses depended on whether it was batch 1 or batch 2. Details are found in *Section 2.6, Phases of Distribution and Allocation*.

Stratum (j) allocation of addresses *(Aj)*: For counties that were eligible for oversampling, multiplicative factors were used in a way that controlled the inclusion probability relative to the inclusion probability of the reference category. Mathematically, county substratum allocation within CY was as follows:

$$A_j = \pi_j T_j = k_j * \pi_{NMW} * T_j = A_{CY} * \left(\frac{k_j T_j}{\sum_j k_j T_j} \right)$$

Where *Tj* was the total addresses in stratum *j*, *Aj* was the addresses located in the stratum *j*, A_{CY} was the addresses located in *CY*, (hence was the $\sum_j A_j$), $\pi_j$was the sampling probability stratum *j*, and k_j was the oversampling factor for the stratum relative to the reference stratum inclusion probability $\pi_{CY,NMW}$ or $\pi_j/\pi_{CY,NMW}$ and the sum is over all strata j that are substratum of CY. In batch 1, the oversampling factor depended on substratum type only: NMW, MW, NMNW, MNW, NMNW or MNW, NMNW or MW or MNW. In batch 2, the factor was a combination of substratum type and observed Black or African American yield rate observed in batch 1 at the time of batch 2 sampling. The application and oversampling coefficients for batch 1 and batch 2 are found in their respective allocation sections.

2.6 Phases of Distribution and Allocation

This section describes the global sampling procedure, the release of the sample(s), the sampling as applied for released batches (soft launch, batch 1), and as planned for unreleased batches (batch 2).

2.6.1 Sample Releases

A preliminary soft launch of 1,000 addresses was sampled and released on October 13, 2023, with the first collection of data on October 16, 2023. The full launch sample was released in two equal-sized batches (batch 1, batch 2). Batch 1, containing 43,500 addresses, was released on December 7, 2023, with the first collection of data on December 11, 2023. Batch 2, containing 43,500 addresses, was released between March 27, 2024, and March 29, 2024, with the first data collection on April 1, 2024.

2.6.2 Soft Launch

The soft launch was a small sample designed to test the process of materials before the full launch sample. The soft launch was released prior to the development and stratification of the 2023-2024 MATCH sampling frame. The sampling frame for the soft launch was the WV MSG ABS frame. This frame was constructed using the same exclusions as for the NORC ABS Frame with two distinctions: there was no removal of duplicates[10] as the duplicates were determined later and the additional exclusion of all drop points as their treatment was determined later.

Although the soft launch sample was prior to the stratification process, randomly selected counts will be found in the *Appendix, Table 7-6: County Soft Launch Sampling Totals and Batch*

[10] An address that was later determined to be a duplicate was in the Soft Launch. Its duplicate was excluded from later sampling.

1 and Batch 2 Final Allocations, this contains county soft launch sampling totals and Batch 1 and Batch 2 final allocations. The strata are assigned to the soft launch addresses *a posteriori* to the sampling, this table is also found in the *Appendix, Table 7-7: Stratum Soft Launch Sampling Totals and Batch 1 and Batch 2 Final Allocations*, this contains stratum soft launch sampling totals and batch 1 and batch 2 final allocations.

2.6.3 Batch 1

This section describes the allocation of addresses to strata for the first half (batch 1: 43,500 addresses) of the full launch sample (87,000 addresses). The process of the stratum allocation for batch 1 involved determining address allocation first for counties and then strata using the established rules under the assumption of having 87,000 addresses available. The actual allocation of addresses to a given stratum was one-half of this initial allocation with rounding and adjustment[11] to obtain exactly 43,500 addresses.

Batch 1 County Allocation Based on 87,000 Addresses

Allocation of the 87,000 addresses to the counties and strata were based on these governing principals and restrictions:

- Minimize design effect with a tendency towards proportionality
- Minimum of 1,000 addresses per county
- Minimum targeted number of completed surveys for counties
- Stratum allocation targeting demographics of interest

A balance of these constraints, using 2021-2022 MATCH yield rates, resulted in allocation of the 87,000 addresses to counties as follows:

- Minimum targets were set for the smallest counties (Types E and F): 220 for Type F counties, 230 for Type E counties without oversampling, or 250 for Type E counties with oversampling. Using 2021-2022 MATCH yield rates, this accounted for a preliminary 53,826 addresses.
- The remaining 33,174 addresses were proportionally allocated to the larger counties (Types A and C) using the frame counts for proportionality.

The details on the Batch 1 allocation to counties based on the available 87,000 addresses are provided in *Appendix B, Section 7.2, Sampling Counts and Allocation Details.* This allocation contained fractional counts and did not add precisely to 87,000. It was corrected in later steps at the time of final batch 1 allocation.

Batch 1 Stratum Allocation Based on 87,000 Addresses

For counties without substrata, the county total was the stratum total. For counties with substrata, but no oversampling, the county total was allocated proportionally to the substrata. For counties with oversampling (Types A, C, and a subset of E), the allocation of addresses to the substratum was performed using the formula in *Section 2.5.5, 2023 MATCH ABS Sampling Frame Stratification* and the oversampling factors found in *Table 2-2: Oversampling Factors by Substratum Type*.

[11] See Batch 1 Stratum Allocation.

Table 2-2: Oversampling Factors by Substratum Type

Substratum Type of Stratum *(j)*	Oversampling Factor (k_j)
NMW*	1
MW	1.25
NMNW	3.5
MNW	3.5
NMNW or MNW	3.5
NMNW or MW or MNW	1.5
Table Note: *Reference substratum	

Details on the county and stratum allocation based on 87,000 addresses are contained in *Appendix B, Section 7.2, Sampling Counts and Allocation Details*. Substratum descriptions are found in *Appendix A, Section 7.1.3 Substratum Types*.

Batch 1 Stratum Allocation

Batch 1 sample size was 50% of the total projected sample size allocated to each substratum up to rounding. This resulted in a preliminary allocation of 43,496 addresses, see *Appendix B, Section 7.2, Sampling Counts and Allocation Details*. To obtain precisely 43,5000 addresses, the four strata with the lowest inclusion probability had one additional address added. These strata were Kanawha-NMW, Berkeley-NMW, Raleigh-NMW, and Jefferson-NMW.

The final Batch 1 allocation totals are found in *Table 7-4: County and Stratum Batch 1 Allocation Based on 87,000 Addresses*.

2.6.4 Batch 2

This section describes the allocation of addresses to strata for the second half (batch 2; 43,500 addresses) of the full launch sample. Black or African American households were not distributed evenly across counties and tended to be concentrated in counties with a larger number of addresses. To balance the need for a sufficient number of completed surveys by Black or African Americans respondents with the total number per county, additional flexibility was added into county targets and stratum oversampling factors for batch 2.

The process of stratum allocation for batch 2 involved determining stratum oversampling factors, determining a targeted number of completed surveys per county, determining the number of addresses needed in the county to achieve the targeted number of completed surveys and then allocating the addresses to the strata using the established rules. These counts were rounded and adjusted to obtain exactly 43,500 addresses.

Batch Process of Simulated Sampling Results and Parameter Establishment

Although details are described sequentially in this document, the establishment of county targets and stratum oversampling factors were selected simultaneously. Due to projections of

Black or African American respondents falling to an undesirably low level at the time of batch 2 allocation, the process of selection of batch 2 parameters involved hundreds of simulations. The simulations were based on soft launch and batch 1 responses available on February 21, 2024.

After initial estimates, the objective of the manual[12] simulations were to maximize the estimated number of Black or African American respondents under these constraints:

- Total number of addresses had to be≤ 43,500
- Total Unequal Weighting Effect (UWE) < 1.5 and highest county UWE approximately 1.4
- The lowest projected county number of completes was 220

Different families of allocations were used for the simulations. The simulations contained some commonalities: (i) rules on allocation of addresses to the county, once the county targeted number of completes was established and (ii) the approach for allocation to substratum.

The description below is for the selected allocation approach, or more accurately, approach for selecting county targeted number of completes. It is outlined in a manner logical for describing steps, but it was selected from a family which had parameters/coefficients that were adjusted in various ways. In fact, the parameter assigned[13] last was the coefficient in the formula for Type A and C county targeted number of completes.

Batch 2 Stratum Oversampling Factors

The final stratum oversampling factors for batch 2 were built around the greatest needs: increased Black or African American response and achieving at least the minimum number of respondents in smaller counties. To meet reasonable projections for these areas of renewed effort in batch 2, two major distinctions from batch 1 in the use of oversampling factors were:

- oversampling was not used in batch 2 for the strata of type MW and NMNW or MW or MNW, and
- oversampling factors were allowed to vary depending on the on the yield rate of Black or African American respondents.

Each stratum was classified according to its Black or African American yield rate (simply defined as the number of respondents who identified only as Black or African American divided by the number of sampled addresses in the soft launch and batch 1). If the Black or African American yield rate was:

- Above 0.02, the stratum was classified as high yield,
- Between 0.01 and 0.02, the stratum was classified as medium yield, and
- Below 0.01, the stratum was classified as low yield.

The oversampling factors by substratum type and yield classification are found in *Table 2-3: Oversampling Factors by Substratum Type and Black or African American Yield Rate Classification* and the individual stratum classifications can be found in the *Appendix, Table 7-6: County Soft Launch Sampling Totals and Batch 1 and Batch 2 Final Allocations*.

[12] Time did not permit the programming of a full constrained optimization.
[13] Assigned to keep the number of allocated addresses below, but as close as possible to 43,500

Table 2-3: Oversampling Factors by Substratum Type and Black or African American Yield Rate Classification

Substratum Type (j)	High Yield Oversampling Factor (k_j)	Medium Yield Oversampling Factor (k_i)	Low Yield Oversampling Factor (k_j)
NMW*	1	1	1
MW**	1	1	1
NMNW	11.75	5	1.25
MNW	11.75	11.75	2.5
NMNW or MNW	10	10	2.5
NMNW or MW or MNW**	1	1	1

Table Notes: *Reference substratum; ** No additional oversampling for the substratum type was performed in Batch 2; High Yield=Black or African American yield rate greater than 0.02 at the time of Batch 2 sampling; Medium Yield=Black or African American yield rate greater than 0.01 at the time of Batch 2 sampling (but not High Yield); Low Yield=Black or African American yield rate at or below 0.01 at the time of Batch 2 sampling

Establishment of Updated County Level Targeted Number of Completions

Based on the soft launch and batch 1 data collected at the time of the batch 2 allocation, county targeted number of completions TGT_{CY} were adjusted. The final target used in assessment of allocation was defined depending on county stratification type (A, C, E, and F) and oversampling eligibility. Due to the need to include more Black or African American respondents, targeted completes for the larger counties were based on a county factor (F_CY) that was related to the stratum oversampling factors k_j and the total number of addresses in the strata T_j. T_{CY}

- Type F: $TGT_{CY} = 220$
- Type E not oversampling eligible: $TGT_{CY} = 230$
- Type E oversampling eligible[14]: $TGT_{CY} = 230 + \frac{10}{7907} * (T_{CY} - 10005)$
- Type C and A[15]: $TGT_{CY} = 240 + 0.7383 * F_{CY} * \left(\frac{T_{CY} - 17912}{17912}\right)$

The county factor was given by the following equation: $F_{CY} = 1 + 0.425 * \left(\frac{\sum_j k_j T_j}{\sum_j k_{j,low} T_j} - 1\right)$ where k_j was the batch 2 oversampling factor for stratum j and $k_{j,low}$ was the low yield oversampling factor for a stratum of the same substratum type as stratum j. Note, if all the substrata in a county were classified as low yield, the county factor reduced to 1.

Batch 2 County and Stratum Allocation

[14] No additional oversampling was performed for Type E counties in Batch 2. This equation is the line between the targeted completion of 230 set for Mingo (10,005 addresses) and 240 set for Wayne (17,912 addresses).

[15] The actual value for the correction was 0.73828125.

Once the county targeted number of completes was established, the preliminary allocation of addresses to the county $B2_{CY,prelim}$ was based on yields from soft launch $SLYLD_{CY}$ and batch 1 $B1YLD_{CY}$ and the yield rate for batch 1[16] $B1YLDRT_{CY}$ via

$$B2_{CY,prelim} = \frac{TGT_{CY} - SLYLD_{CY} - B1YLD_{CY}}{B1YLDRT_{CY}}.$$

The preliminary substratum allocations $B2_{j,prelim}$ were assigned using the formula above and the oversampling factors that correspond to the substratum's Black or African American yield classification and type found in *Table 2-3: Oversampling Factors by Substratum Type and Black or African American Yield Rate Classification*, rounded to a whole number.

This resulted in a preliminary allocation of 43,490. To obtain precisely 43,500 addresses, the ten strata with the highest Black or African American yield rate each had one additional address added. These strata were Kanawha-MNW, Berkeley-MNW, Cabell-NMNW or MNW, Wood-NMNW or MNW, Raleigh-NMNW or MNW, Mercer-NMNW or MNW, Marion-NMNW or MNW, Jefferson-MNW, Ohio-NMNW or MNW, and McDowell-NMNW or MW or MNW.

The final batch 2 allocation totals are found in *Appendix, Table 7-6: County Soft Launch Sampling Totals and Batch 1 and Batch 2 Final Allocations* and *Table 7-7: Stratum Soft Launch Sampling Totals and Batch 1 and Batch 2 Final Allocations*.

2.6.5 Programming the Samples

All sample batches were selected using SAS proc survey select using simple random sample without replacement using SAS 9.4.

For the soft launch sample, the frame was unstratified and sorted by BMS region (numeric, see *Table 7-3: BMS Regional Breakdown* for county to region assignment), county (alphabetic), and unique identifier from the WV MSG ABS frame prior to selection. The initializing seed was 600397 chosen by random.org.

For batch 1 and batch 2 samples, sampling was stratified and selected according to the designated allocations. The 2023-2024 MATCH ABS sampling frame was sorted by stratum, non-White status, Medicaid/low status, and the unique sampling unit identifier for the address. The initializing seed was 5811875 for batch 1 and 8965758 for batch 2 chosen by random.org.

[16] Soft launch was not used in the estimation of county yield rate because there was no stratification in the soft launch.

3 Questionnaire

3.1 Synopsis of Questionnaire Design

This section outlines the procedures used to update and program the CATI, CAWI and PAPI survey instruments for the 2023-2024 fielding of the MATCH survey.

3.2 Objectives of the Questionnaire Design

The primary aim of designing the questionnaire was to achieve the following objectives:

- To better understand the health of West Virginians to match community health needs with resources designed to meet those needs.
- To ensure questions are tailored to address key aspects of health status and behaviors, healthcare access, substance use, mental health, and other factors influencing health.

3.3 Key Differences with the 2021-2022 MATCH Fielding

The survey instrument was updated for the 2023-2024 fielding based on feedback from stakeholders.

Some of the questions changed based on the special interests of the DoHS, such as the prevalence of vaping. Questions and responses focused on COVID-19 were removed as the information was no longer pertinent. Other changes include renaming question section headings to be more inclusive. The key differences between the 2021-2022 and 2023-2024 survey instruments are summarized in *Appendix C, section 8.1 Questionnaire Content Changes from 2021 MATCH to 2023 MATCH* below.

3.4 Survey Instrument Content

The survey was organized into eight sections. Questions within these sections examined the main topics shown in *Table 3-1: Pen and Paper Interviewing Survey Content*.

Table 3-1: Pen and Paper Interviewing Survey Content

Questionnaire Section	Contents of Section
Your Health	▪ Overall health ▪ Difficulty in performing daily activities ▪ Diagnoses of common chronic conditions
Healthcare Access	▪ Insurance status and type ▪ Prescriptions filled and difficulty with acquiring prescription medication ▪ Need for medical care and reasons for not seeking needed medical care, if applicable ▪ Telehealth usage
Mental Health	▪ Overall mental health ▪ Impact of emotions on daily activities ▪ Mental health over the past 2 weeks ▪ Need for mental health care and reasons for not seeking mental health care, if applicable

Questionnaire Section	Contents of Section
	▪ Mental health prescriptions filled and difficulty with acquiring the
Lifestyle	▪ Difficulty in meeting financial obligations ▪ Food purchasing and consumption habits ▪ Reasons for being treated unfairly by a healthcare provider and actions taken in response
About You	▪ Age ▪ Gender ▪ Race/ethnicity ▪ Marital status ▪ Living situation ▪ Household composition and number of children in the household ▪ Educational attainment ▪ Employment status and occupation ▪ Use of public assistance programs ▪ Appalachian culture
Health Behaviors	▪ Alcohol consumption ▪ Smoking status ▪ Cannabis usage ▪ Usage of common substances (opioids, benzodiazepines, over-the-counter and prescription stimulants, cocaine, methamphetamine, heroin, and MDMA) ▪ Attempts to decrease use of common substances ▪ Withdrawal experiences from decreasing use of common substances ▪ Usage and reasons of use of opioids ▪ Overdose experiences and use of Narcan ▪ Need for substance use treatment and reasons for not seeking substance use treatment, if applicable ▪ Use of medications designed to reduce/stop the use of drugs or alcohol ▪ Impact of substance use on life events
Other Topics	▪ Outlook towards life ▪ Physical activity and access to physical activity resources ▪ Sexual orientation ▪ Family members in the household

Questionnaire Section	Contents of Section
	▪ Income level ▪ Thoughts/actions pertaining to suicide ▪ Consent to be recontacted by HAI in the future

3.5 Survey Instrument Development and Different Modes

3.5.1 CAWI

Participants who wanted to complete the survey online were able to do so using the CAWI module, Acuity4, from Voxco. The 2023-2024 survey was programmed into Voxco by NORC with assistance from the HAI Survey Research and Administration Program.

Although every question on the survey required an answer, a participant was allowed to skip any question they did not want to answer. A 'Web Skip' response was coded as a hidden default response option when the participant clicked the 'next arrow' on any question they did not want to answer.

The status options seen in Voxco were 'Complete', 'Dropped' (survey was inactive for over 90 minutes), and 'Interrupted' (participant closed browser). The participant had to submit the survey at the end to be marked as complete in the Voxco system.

3.5.2 CATI

Participants who wanted to complete the survey via telephone were able to do so using the Voxco CATI system. The Interactive Voice Response (IVR) pathway was developed to route participants who wanted to take the survey to an available interviewer after they entered their access code. If there was no interviewer available to answer their call, they were prompted to either wait for an available agent or leave a voicemail message. To maintain efficiency, the access code was embedded in the IVR process so that the interviewer was taken directly to the correct Case ID when they received an inbound call. An example of the IVR pathway is provided in *Appendix D, Section 9.1 Data Collection Figures*.

Interviewers logged in to take inbound calls in the Automatic Call Distribution (ACD) module. A 'WELCOME' screen was programmed for interviewers in the module with a script to introduce themselves and the survey and then ask the caller for consent to participate on the following 'CONSENT' screen. This allowed participants to take the survey later or refuse participation before beginning the survey.

The survey itself was launched using the CAWI module. The interviewers clicked on the URL on the 'LAUNCH SURVEY' screen to open the CAWI survey directly into a web browser. This allowed for a more uniform survey across the CAWI and CATI systems. All survey language was edited and programmed from the first-person narrative to the second-person narrative (e.g., "I don't know" was changed to "You don't know".) Instructions to interviewers on how to read questions were provided in CAPS LOCK. This indicated to the interviewer to not read these instructions aloud. There were two options available to interviewers to allow them to skip a question: "DON'T KNOW" and "REFUSED". These were added to all questions that did not

already have these options. Interviewers were instructed to not read them aloud to participants but select one of them to move to the next question. Interviewers closed out of the ACD module by selecting Complete on the 'LAUNCH SURVEY' screen if they completed the survey with the participant or returning to a previous screen to select another status to indicate why the survey was not finished (e.g., Will call back when ready).

3.5.3 PAPI

To update the survey instrument for the 2023-2024 fielding, HAI collaborated with NORC to develop the PAPI.

There are a few key differences in the PAPI versus the CAWI and CATI. Text boxes that specified 'other responses' were removed from the PAPI questions except for the lifetime job category questions. Two health behavior questions were included in the CAWI and CATI questionnaire but removed from the PAPI due to low responses on the first fielding. These questions were as follows:

- Q84 "(If yes to any in Q83) In the past 12 months, have you gotten sick or had flu-like symptoms when trying to cut down or stop using any of the following?"
- Q87 "(If yes) Thinking about the last time you used prescription opioids, or 'pills', in any way a doctor did not direct you to use it, what were the reasons you used it the last time?"

3.5.4 Spanish Language

The 2023-2024 survey instrument was translated to Spanish by a vendor and then reviewed by both NORC and HAI. After final approval, NORC programmed the Spanish version into Voxco. The survey invitation mailings included Spanish translation instructions to complete the survey either by going to the website or scanning the code, then entering their personal access code. At the top of the survey on the web, a language drop-down menu allowed participants to switch the survey from English to Spanish. There was no option to take the survey over the phone or the PAPI in Spanish.

3.5.5 Changes During Fielding

Changes that occurred during the survey fielding include:

- An error found in the skip logic from Q29 to Q36 during the soft launch that was corrected on 12/6/23 before batch 1 launched.
- A 'Will call back when ready' option was added to the CATI module on 11/27/23.
- A 'Schedule callback' option was removed from the CATI on 12/7/23.
- The language for the incentive was changed from '10-dollar gift card' to '$10 incentive'.

4 Data Collection

4.1 Synopsis of Data Collection and Procedures

This section includes the implementation protocol used for the survey, outlining the mailing materials and selection process, and overviewing survey completion options. This section also describes the differences between the batches of the launch, how surveys were determined to be complete, and how partial completions were treated.

4.2 Objectives of the Data Collection and Procedures

The primary aim of designing the questionnaire is to achieve the following objectives:

- Collect robust mental and behavioral health sub-state population-level information;
- Collect data generalizable at the county level with the ability to identify subsets of vulnerable populations.

4.3 Key Differences with the 2023-2024 MATCH Fielding

There were key differences between data collection during the 2021-2022 MATCH fielding and the 2023-2024 MATCH fielding. The differences are described in detail below in Key Differences with the 2021-2022 MATCH Fielding, Section 5.3.

4.4 Procedures

4.4.1 CAWI Process

To access the survey, the participant was able to scan the QR code on their mailing or to navigate to www.TakeMATCHsurvey.org and enter their unique access code.

Participants were able to stop survey completion and later re-enter the survey at any time until it was completed. Participants had the ability to go back to previous questions and change their answers. Upon completion, participants were unable to log into the survey again.

4.4.2 CATI Process

A telephone number was provided on the mailings, which connected the participant to the IVR. The IVR process required the access code to be entered in the automated recordings to enter the queue to speak to an agent during call center hours. Interviewers began their inbound calls with a 'WELCOME' screen where they would introduce themselves and ask if the participant wanted to take the survey now. They could then select "Continue" to proceed with the survey or "Refuse", "Language Barrier", "Take me off your list", and "Will call back when ready" to not continue with the survey.

Call Center

- For the soft launch of the survey, NORC took all inbound calls and provided a scheduling plan to staff agents based on call volume.
- For batch 1, the call center was open from 8 a.m. to 8 p.m., Monday through Saturday and 12 p.m. to 8 p.m. on Sunday beginning on 12/11/23. NORC covered inbound calls before and after work HAI work hours (9 a.m. to 5 p.m.) and on weekends. NORC also provided coverage during all hours of surge periods, which were high call volumes surrounding mailings being received. HAI trained nine research specialists as interviewers who staffed the call center in two shifts at 9 a.m. to 1 p.m. and 1 p.m. to 5 p.m. Monday through Friday. The HAI Survey Research Program monitored call center

for any issues and call volume. There were two interviewers per shift in December 2023, which shifted to one interviewer per shift in January 2024. Due to diminished call volume, the call center was paused in February 2024 until the launch of batch 2.

- For batch 2, the call center opened on March 29, 2024, but hours were changed from Monday to Friday, 9 a.m. to 5 p.m. NORC provided coverage for inbound calls from March 29 through April 26, 2024. HAI call center staffing remained the same, with two interviewers per shift that reduced with call volume.

Voicemails

Voicemails were monitored and returned by the MATCH project team and HAI Survey Research and Administration Program. Participants left voicemails to inquire about taking the survey or with questions about their incentives, the survey, or the project itself. Voicemails were processed and entered on the voicemail tracking sheet by the HAI Survey Research and Administration Program. Interviewers called back those who wanted to complete the survey over the phone, and project or program coordinators responded to the incentive questions, placed general callbacks, and monitored the document for completion of all callbacks. A callback schedule was made to separate call center shifts from callbacks for staff trained to work in the call center. Callbacks were placed within 24 to 48 hours from when the call was reviewed during high call volume. When call center volume declined, callbacks were done once a week. Two callback attempts were made before marking a case as resolved. During call center shifts, one interviewer did callbacks and one interviewer took inbound calls during low call volume.

4.4.3 PAPI Process

PAPIs were sent in the third mailing of both batches by MSG. PAPIs were returned to the mailroom at WVU and indexed in the HAI office. Each returned PAPI contained a unique barcode that identified the participant from their Case ID. The barcode was scanned onto a tracking sheet, and the date of the scan was entered. A visual review was performed to determine if a PAPI was complete or incomplete, and these results were entered in the tracking sheet along with the age and the race of the participant (batch 1 only).

The PAPIs were boxed and numbered with an indexed spreadsheet which recorded the number of boxes filled, how many PAPIs were in each box, the PAPI Case IDs indicated by the unique barcode, and the date each PAPI was placed in the box. PAPIs were stored in a secure office before they were transferred to MSG for scanning. MSG was responsible for PAPI data processing. Operators from MSG used ABBYY software to scan the PAPIs for data extraction, and the data scans were then sent in Excel format to HAI analysts. After completing the process for batch 1 PAPIs, a member of the MATCH project team retrieved the PAPIs from the MSG office. The completed PAPIs are stored for six years in accordance with a memorandum of understanding.

4.5 Methods Used to Increase Response Rate

4.5.1 Incentive Structure

In the MATCH 2023-2024 fielding, the survey invitation included $2 which was visible through the clear address box on the mailing envelope. This pre-incentive was included to encourage recipients to open the mailing and participate in the survey.

In all batches of the survey, participants are offered a $10 incentive upon completing the survey. When participating via CAWI or CATI, individuals could receive the incentive by USPS or by email address. The virtual post incentives were distributed by Medmatics, LLC dba Virtual Incentives. When participating via PAPI, individuals could receive their incentive as cash, facilitated by MSG. An email option for receiving the incentive was not offered on the PAPI to avoid errors in deciphering the participants handwritten email addresses.

Participants could also choose to opt out of receiving an incentive.

4.5.2 Mailing Schedule

The mailing protocol was a push-to-web design to increase the number of CAWI completions, decrease the amount of item-missing data, and to allow additional timing for PAPI content finalization and printing.

A soft launch of 1,000 addresses was sent prior to the first mailing, which was an invitation letter, in batch 1. The first invitation letter was sent to all batch 1 sample members on December 7, 2024. The first invitation letter was sent to batch 2 sample members on March 29, 2024. The full mailing schedule is shown in *Table 4-1: MATCH Mailing Schedule, All Batches*.

Table 4-1: MATCH Mailing Schedule, All Batches

Mailing	Soft Launch Date	Batch 1 Date	Batch 2 Date
Mailing 1: Invitation Letter	10/13/2023	12/7/2023	3/29/2024
Mailing 2: Reminder Postcard	12/21/2023	12/21/2023	4/12/2024
Mailing 3: PAPI Packet	1/11/2023	1/11/2024	4/30/2024
Mailing 4: Second Reminder Postcard	1/24/2024	1/24/2024	5/14/2024
Mailing 5: Targeted Mailing Postcard (5,000 Sample)			7/3/2024

4.5.3 Mailing Material Content

In the first two mailings, respondents were encouraged to complete the survey by web, and both the invitation letter and the reminder postcard contained a link to the landing page, www.TakeMATCHsurvey.org, a survey access code, and the number to call to complete the survey by phone. The third mailing included the PAPI packet with a printed questionnaire. The fourth mailing consisted of a reminder postcard. Efforts were made to not send reminder mailing materials to individuals who had already completed the survey.

See *Appendix D, Section 9.2, Mailing Materials* for a copy of these mailing materials.

The full list of materials is listed below with a more detailed description of each:

- Mailing 1: Invitation Letter: This letter was mailed in a 6- by 9-inch WVU-branded envelope with the invitation letter inside it. The letter explained the purpose of the project, inviting the respondent to participate, and provided a URL

(www.TakeMATCHsurvey.org) for the respondent to take the survey, a survey access code for the respondent to enter once they got to that URL, and a QR code that they could scan to be instantly taken to the website without having to manually type in the address. In addition, contact information was provided should the sample member have any questions about the project or wish to complete the project by phone.

- Mailing 2: Reminder Postcard: A fold-over postcard was sent after the invitation letter. It included the same URL and QR code, the survey access code and contact information, and reminded respondents to complete the survey if they had not done so already.
- Mailing 3: PAPI Packet: This 8- by 12-inch mailing packet contained an external HAI-branded envelope. Inside was an invitation letter, a paper survey with a cover, and a postage-paid and self-addressed business return envelope that the respondent could use to send the survey back to HAI free of charge. The invitation letter asked respondents to complete the survey either by web or by mail. The URL, QR code, and survey access code were all provided in this letter as well, and the contact information was prominently displayed.
- Mailing 4: Second Reminder Postcard: A fold-over postcard was sent that included the same URL and QR code, and the survey access code and contact information, and reminded respondents to complete the survey if they had not done so already.
- Mailing 5: Targeted Mailing Postcard: A front and back postcard was sent which included the same URL and QR code, and the survey access code and contact information, and reminded respondents to complete the survey if they had not done so already. This was sent to a targeted sample of 5,000 in batch 2.

4.5.4 Community Outreach

Reaching all 55 counties in WV, especially regions with low internet connectivity, requires a flexible strategy by understanding the geography, demographics, and other unique characteristics that define the WV population.

The purpose of the outreach efforts is twofold: (1) To ensure connections with trusted partners (e.g., civic organizations, newspapers, churches, local schools) who can assist with amplifying the MATCH message in their communities and create awareness of the survey, and (2) to establish deep organizational partnerships with local communities that facilitate trust, connection, and future collaboration on survey fielding.

Organizations were strategically targeted with the intent of capturing a diverse mix of potential community connections. To effectively connect with these partners, various outreach methods were utilized (e.g., call, mail, email, social media). This approach engaged a broad audience of partners that helped amplify the MATCH message in their respective communities, and it helped keep overall costs at a minimum.

For a list of community outreach efforts by month, please see *Appendix D, Section 9.3 Community Outreach Efforts by Month*.

4.6 Defining a Completed Survey

The final processed survey has two completion categories: i.e., Completed and Partially Completed surveys. If a respondent answered at least 52 (≥ 95%) out of the 54 complete-tag questions and at least four out of six of the weighting questions, then a survey was considered

'Complete'. If a respondent answered at least 33 (≥ 95%) out of the 34 partial-tag questions or at least 38 (≥ 70%) out of the 54 complete-tag questions along with at least four out of the six weighting questions, then a survey was considered 'Partial Complete'. Please refer to *Table 4-2: Questions and Accepted Responses Involved in Survey Completion Assessment* for the list of survey questions that were tagged for completion, partial completion, and weighting.

Table 4-2: Questions and Accepted Responses Involved in Survey Completion Assessment

Survey Question Number	Variable Name	Accepted Response or Value	Complete (C) Partial Complete (P) Weight (W)
Q2	HTS_GENHLTH	Not missing (999)	C, P
Q3	HTS_DIFDACT	Not missing (999)	C, P
Q7	HTS_TRSLP2W	Not missing (999)	C, P
Q8	HAD_INSCVRG	Not missing (999)	C, P
Q10	HAS_RXHP12M	Not missing (999)	C, P
Q13	HAS_CRND12M	Not missing (999)	C, P
Q16	HAS_TLVT12M	Not missing (999)	C, P
Q21	HAS_ERCR12M	0-99 and not missing (999)	C, P
Q22	MHS_MTLHLTH	Not missing (999)	C, P
Q23	MHS_MTLPASK	Not missing (999)	C, P
Q24	MHR_EINF12M_CHRS MHR_EINF12M_SCLF MHR_EINF12M_RLSP MHR_EINF12M_PFRM	None are missing (999)	C, P
Q25	MHR_KSSLR2W_NERV MHR_KSSLR2W_HOPE MHR_KSSLR2W_FIDG MHR_KSSLR2W_DEPR MHR_KSSLR2W_WRTL MHR_KSSLR2W_ISLT	None are missing (999)	C, P
Q26	MHS_EMOSPRT	Not missing (999)	C, P
Q27	MHS_MHDX12M	Not missing (999)	C, P
Q28	MHS_MCND12M	Not missing (999)	C, P
Q36	MHS_MHRX12M	Not missing (999)	C, P
Q40	MHS_UFTX12M	Not missing (999)	C, P
Q43	LSR_RENT12M	Not missing (999)	C, P
Q45	LSR_DEBT12M	Not missing (999)	C, P
Q46	LSS_FOOD12M	Not missing (999)	C, P
Q47	LSS_FDSK30D	Not missing (999)	C, P
Q48	LSS_FRSHBUY	Not missing (999)	C, P
Q50	LSR_FRML30D_FPNT LSR_FRML30D_MOW LSR_FRML30D_RORG	At least one is not missing (999)	C, P

Survey Question Number	Variable Name	Accepted Response or Value	Complete (C) Partial Complete (P) Weight (W)
	LSR_FRML30D_SPKT LSR_FRML30D_OTHR LSR_FRML30D_NONE		
Q51	DMS_AGE	0-119 and not missing (999)	C, P, W
Q52	DMS_BRTHSEX	Not missing (999)	C, P, W
Q53	DMR_GENDER_MALE DMR_GENDER_FMLE DMR_GENDER_TRNS DMR_GENDER_OTHR	At least one is not missing (999)	C, P
Q54	DMS_HISPANC	Not missing (999)	C, P
Q57	DMR_RACE_WHTE DMR_RACE_BLCK DMR_RACE_AIAN DMR_RACE_ASIA DMR_RACE_NHPI DMR_RACE_OTHR	At least one is not missing (999)	C, P, W
Q58	DMS_MRTLSTA	Not missing (999)	C, P, W
Q59 Q60	DMS_LIVEWTH_ALNE DMD_RESADLT	Q59 is "No one other than yourself" or Q60 is not missing (999)	C, P, W
Q64	DMS_HOMETYP	Not missing (999)	C, P
Q65	DMS_HOMEPMT	Not missing (999)	C, P
Q66	DMS_MAXEDUC	Not missing (999)	C, P, W
Q67	DMR_EMPSTAT_EMPL DMR_EMPSTAT_SELF DMR_EMPSTAT_STUD DMR_EMPSTAT_HMKR DMR_EMPSTAT_RETD DMR_EMPSTAT_UEMP	At least one is not missing (999)	C, P
Q70	DMS_VETSTAT	Not missing (999)	C
Q71	HHS_JEXP12M_TLYD HHS_JEXP12M_PLYD HHS_JEXP12M_LTPW HHS_JEXP12M_LWRK HHS_JEXP12M_UTO HHS_JEXP12M_LPAY HHS_JEXP12M_NPAY HHS_JEXP12M_LBNF HHS_JEXP12M_QJOB HHS_JEXP12M_FJOB	None are missing (999)	C

Survey Question Number	Variable Name	Accepted Response or Value	Complete (C) Partial Complete (P) Weight (W)
	HHS_JEXP12M_MPMT HHS_JEXP12M_UEMP		
Q72	HHS_FINA12M_USAV HHS_FINA12M_CTFD HHS_FINA12M_IDBT HHS_FINA12M_URET HHS_FINA12M_BRWM HHS_FINA12M_PAWN	None are missing (999)	C
Q73	HHS_PBNF12M_TANF HHS_PBNF12M_SNAP HHS_PBNF12M_WIC HHS_PBNF12M_MDCD HHS_PBNF12M_LEAP HHS_PBNF12M_LFLN HHS_PBNF12M_VCHR HHS_PBNF12M_JHWV	None are missing (999)	C
Q74	BHS_ALCD30D	0-30 and not missing (999)	C
Q77	BHS_ECIGFRQ	Not missing (999)	C
Q78	BHS_SMKFREQ	Not missing (999)	C
Q79	BHS_MARJ30D	0-30 and not missing (999)	C
Q82	BHS_SUBU12M_MARJ BHS_SUBU12M_OPID BHS_SUBU12M_BENZ BHS_SUBU12M_CNTR BHS_SUBU12M_STML BHS_SUBU12M_COCN BHS_SUBU12M_METH BHS_SUBU12M_HRIN BHS_SUBU12M_MDMA	None are missing (999)	C
Q84	BHS_OPID12M	Not missing (999)	C
Q87	BHS_OVRDEVR	Not missing (999)	C
Q92	BHS_ODFM12M	Not missing (999)	C
Q93	BHS_SCND12M	Not missing (999)	C
Q100	BHR_SUXP12M_FAMH BHR_SUXP12M_RELH BHR_SUXP12M_CPS BHR_SUXP12M_APS BHR_SUXP12M_LSHM BHR_SUXP12M_LSJB BHR_SUXP12M_LEO	None are missing (999)	C

Survey Question Number	Variable Name	Accepted Response or Value	Complete (C) Partial Complete (P) Weight (W)
	BHR_SUXP12M_SUDX BHR_SUXP12M_LSDL		
Q101	OTS_LIFESAT_IDL OTS_LIFESAT_XLNT OTS_LIFESAT_STFD OTS_LIFESAT_WANT OTS_LIFESAT_NOCH	None are missing (999)	C
Q102	OTS_EXER30D	Not missing (999)	C
Q103	OTS_SEXATRC	Not missing (999)	C
Q104	OTS_INCFMSP	>= 1 and not missing (999)	C
Q106	OTS_INCCATG	Not missing (999)	C
Q108	OTS_SUICIDE	Not missing (999)	C

4.7 Response Rates

4.7.1 Council of American Survey Research Organizations and AAPOR Response Rates

Response rates[17] are indicators of survey and sample performance and response. American Association for Public Opinion Research (AAPOR) standard definitions of Response Rate No.2 (RR2, Equation 4-1)[5] were calculated for the combined sampling frame. All response rates are affected by the procedure of assigning final status dispositions. The results of each mail attempt were assigned a disposition status according to AAPOR standard guidelines.

4.7.2 Response Rates

Out of the 111 survey questions, only 54 were assessed for completion because they are not gate questions or post-survey incentive questions. Please see *Section 4.6, Defining a Completed Survey* and the associated table for detailed list of questions that were used for assessing completeness. Completed surveys (S) were defined as respondents that answered at least 95% of the 54 questions that were assessed for completion. Partially completed surveys (P) were defined as respondents who answered at least 70% of the 54 questions assessed for completion. Refusals (R) were defined as anyone who began a survey but did not complete enough to be considered partial or fully complete. Noncontacts (NC) were defined as cases that did not respond to any survey invitations. Ineligibility and unknown eligibility had precedence over R and NC in terms of designations. There were no instances that fit the AAPOR standard for O in Equation 4-1. Definitions and counts of cases by eligibility are reported in Table 4-3: Number of Undeliverable Mailings Across all 2023 MATCH Mailings. Cases where occupancy of the housing unit was unknown were classified as UH. Based on these, the equation for RR2 is below.

[17] Calculated response weights were unweighted.

Equation 4-1

$$RR2 = \frac{\mathrm{S} + \mathrm{P}}{(\mathrm{S} + \mathrm{P}) + (R + NC + O) + (UH + UO)}$$

In the frame, cases where one or more undeliverable messages from USPS were received and were not categorized as vacant[18] were considered as not eligible. Undeliverable mailings classified as deceased were considered other unknown eligibility (UO). The RR2 calculation treats these addresses of unknown eligibility as if they were eligible. *Table 4-3: Number of Undeliverable Mailings Across all 2023 MATCH Mailings* presents the number of undeliverable cases across the four mailings by reason for being undeliverable. Cases with an undeliverable message received from USPS indicating the address is vacant were considered not eligible.

The RR2 was 20.5% (*Appendix D, Table 9-2: Response Rates)*. Response rates by region, county, and survey strata can be found in *Appendix D, Table 9-2: Response Rates*.

Table 4-3: Number of Undeliverable Mailings Across all 2023 MATCH Mailings

Standardized Deliverability Description	Deliverability Description	Soft Launch	Batch 1	Batch 2
Cannot be Delivered [IA], Unknown Eligibility	Attempted, not known	3	359	479
	Illegible	-	2	2
	Insufficient address	-	215	283
	No mail receptacle	5	523	580
	Not deliverable as addressed/unable to forward/ forwarding order expired	65	2,582	2,729
No Such Number [NSN], Not Eligible	No such number	11	603	732
No Such Street [NSS]	No such street	1	68	69
Refused by Addressee [REF], Eligible	Refused	-	23	56
	Unclaimed	-	52	49
Returned to Sender, Unknown Eligibility	In dispute	-	2	-

[18] Categories from the USPS include Temporarily Away, Not Delivered – No Forwarding, Attempted – Not Known, No Such Number, Insufficient Address, No Mail Receptacle, Refused, Other

Standardized Deliverability Description	Deliverability Description	Soft Launch	Batch 1	Batch 2
Vacant [VAC], Not Eligible	Vacant	47	2,606	2,965
Other, Unknown Eligibility	Deceased	-	5	7
Total		132	7,040	7,951

Table 4-4: AAPOR Response Rate Calculation Components Based on Final Eligibility Assessments

Eligibility Assessment	Formula Alignment	Value
Complete	S	14,866
Partial Complete	P	1,837
All Eligible	(S+P) + (R+NC+O)	76,718
Unknown Eligible	(UH+UO)	4,628
Not Eligible	Not included in response rate	6,654
Response Rate No.2	RR2 = (S+P) / [(S+P) + (R+NC+O) + (UH+UO)]	0.205 or 20.5%

Table 4-5: Number of Fully and Partially Completed Surveys by Frame and Mode

-

Mode	Fully Completed	Partially Completed
Web completes	11,126	939
Paper Completes	3,740	898

5 Data Processing and Analysis

5.1 Synopsis of Data Processing and Analysis

The data processing and analysis section of this report describes the tools and approach for cleaning, standardizing, and analyzing 2023 MATCH survey data.

The 2023 MATCH survey dataset was processed in SAS version 9.4 software using a double programming approach for quality control. The cleaned 2023 MATCH survey dataset was developed using key variables from sample data (soft launch, batch 1, and batch 2), CATI/CAWI response data (SPSS format), PAPI response data (csv format), and undeliverable data (csv format). The 2023 MATCH sample data has the addresses sampled for the soft launch, batch 1, and batch 2 of the survey, with key information such as address, batch number, and survey PIN. The Voxco survey dataset management system stored the entire 2023 MATCH sample data which included web (CATI/CAWI) response data. The PAPI survey responses were processed by MSG and linked to the web survey responses to produce the final processed survey dataset. Lastly, the undeliverable mailing list provided by MSG had undeliverable dates (attempted delivery) and codes which were used to determine survey eligibility.

5.2 Objectives of Data Processing and Analysis

1. Establish format rules for scanning paper surveys and downloading web survey.
2. Determine missing data rules and exception cases for missing data for all question types (single-select, select-all-that-apply, skip-logic).
3. Develop data cleaning rules for:
    a. Mutually exclusive responses in select-all-that-apply questions.
    b. Mode specific cleaning for insurance questions, numeric entry values, multiple responses to single-select questions (for PAPI).
4. Establish criteria for determining survey completion.
5. Create data standardization rules for survey dataset.
6. Weight the dataset to be used in providing representative results.

5.3 Key Differences with the 2021-2022 MATCH Fielding

2021-2022 MATCH fielding was conducted by a third-party professional, Research Triangle Institute International (RTI), whereas 2023-2024 MATCH fielding was conducted by HAI. As a result, there is extensive documentation for the 2023-2024 MATCH dataset lifecycle, spanning data lineage and metadata.

Listed below are some key differences between 2021-2022 and 2023-2024 MATCH data processing:

- Open-ended responses were downloaded but not processed to be included in the cleaned 2023-2024 MATCH dataset.
- 2021-2022 MATCH data had only derived and recoded variables with similar definitions, but 2023-2024 MATCH dataset had more variable types with specific scenarios and criteria.
- 2021-2022 MATCH dataset did not have a data standardization procedure.

5.4 Data Processing

The data cleaning process of 2023-2024 MATCH data was comprised of rules for cleaning, survey disposition assignment, and standardization. Data cleaning rules focused on missing data, survey skip logic, different types of survey questions (single-select, select-all-that-apply), free numeric responses, unexpected multiple responses to single-select questions in paper survey.

5.4.1 Cleaning Missing Data

Single-select questions

Any question in the survey could have been skipped or not answered. The incoming CATI/CAWI response dataset had "Missing", "Don't know", and "Refused" responses coded as 999, 888, and 889, respectively. Since the options "Don't know" and "Refused" were only presented for survey administrators of CATI survey mode, all instances of 888 and 889 in CATI and CAWI modes were recoded to 999, which served as the standard code for missing data. The incoming PAPI responses dataset had only a "Missing" response which was programmed as 999.

Select-all-that-apply questions

When a select-all-that-apply question was skipped or had unselected response options, the incoming value was 0. To differentiate between questions that were skipped and those with unselected options, the incoming value of 0 was either recoded to 999 or retained the value of 0. When none of the response options for a question was selected, all the responses to the select-all-that apply question were recoded to 999 to indicate that the question was skipped. When some response options for the question were selected, the incoming value of 0 was retained for unselected options to indicate that the question was not skipped.

Rule exception(s)

The incoming values for "Missing", "Don't know", and "Refused" for the annual income survey question, No. 107. To be more precise, please enter your best estimate of you and your family members' total 2022 annual income before taxes and other deductions?" were 99999999, 88888888, and 88888889, respectively, because 999, 888, and 889 codes were considered valid within-range responses. To have a consistent missing value code across all survey modes, the income question's incoming values for "Missing", "Don't know", and "Refused" were recoded to -999.

The survey question, "Do you identify as Appalachian?" offers "I don't know" and "Prefer not to answer" response options as valid responses across all survey modes, so the incoming values were not recoded to 999, as they were for other single-select questions from CAWI mode.

Cleaning gate questions

Gate questions in a survey define skip logic, whether a respondent should or should not have answered the associated follow-up questions. When a gate question is 'satisfied', it means that the follow-up questions should have been answered. If a gate question is 'not satisfied', it means that the follow-up questions should not have been answered.

If a gate question was 'satisfied', then the follow-up questions' missing data was recoded as 999 to indicate that they should have answered but were skipped. If a gate question was 'not satisfied', then the follow-up questions' missing data was recoded as 777 to indicate that they should not have been answered.

If the skip logic of gate questions included select-all-that-apply questions, then the cleaning rules for missing data combined rules defined for gate questions and select-all-that-apply questions.

Health insurance gate question

The health insurance gate question, "Do you have any kind of health insurance coverage, including private health insurance or government plans such as Medicare or Medicaid?" had an additional cleaning step. Web (CATI/CAWI) respondents were asked the follow-up insurance question if they skipped the gate insurance question because of the possibility that respondents may still select an insurance type. Similarly, PAPI surveys required additional cleaning because respondents do not always follow gate instructions on paper. Regardless of the response to the insurance gate question, if a respondent selected an insurance type in the follow-up question, "What kinds of health insurance or healthcare coverage do you have?" then a new variable was created with a revised incoming value to indicate "Yes" to the gate question. The incoming values to the follow-up question were assigned 777 code if and only if none of the insurance types was selected. New variables created to recode the incoming values of gated questions in accordance with the follow-up responses are called derived variables.

Cleaning mutually exclusive responses

Some select-all-that-apply questions may have mutually exclusive responses. For example, the question "In the past 30 days, did you or anyone in your household get free groceries or free meals from the following?" has two unique responses that cannot be selected along with other responses. The unique responses to this question were "I received free groceries or free meals but not from any of the above" and "I did not receive free groceries or free meals." If a respondent selected either one of these unique responses, they couldn't select any of the other non-unique responses that specify sources of free groceries or meals: "Food pantries or food banks" "Meals on Wheels" or "Religious organizations" or "Shelters or soup kitchens." Additional cleaning was completed on questions with mutually exclusive responses to ensure that two unique responses are not selected simultaneously, or a unique response was not selected along with the other non-unique responses. If any of the non-unique response options was selected, then the unique value was assigned 0 to indicate it was not selected along with non-unique responses. If two unique response options were selected, then the 'primary' unique value was retained with a code of 1, while the 'secondary' unique value was coded 0 to indicate it was not selected. All questions with mutually exclusive responses were examined and had unique value assignments determined by a subject matter expert and primary analyst prior to data cleaning.

5.4.2 Cleaning Paper Survey

The PAPI survey response data went through additional cleaning steps because the survey was printed on paper and did not have restrictions programmed as in a web survey.

Numeric responses

Some of the many incorrectly processed handwritten values include unexpected symbols (example: 100*) or out of range numeric responses (example: birthdate instead of age). A short list of unexpected values for critical variables such as age were corrected after careful visual inspection of the original paper survey submission.

Multiple responses to single-select questions

In scenarios where multiple responses were selected to a single-select question, the PAPI scanning program was assigned to capture all responses indicated, so specific cleaning rules can be applied based on the question type. Generally, when multiple responses were indicated for single-select questions, the incoming value was recoded to the standard missing code 999.

Rule exception(s)

If multiple options were selected to the survey question, "What is the highest degree or year of school you completed?" then the highest level of education was retained.

If multiple responses were captured for single-select Likert scale questions, then all the multiple incoming values were examined by subject matter experts to determine the most appropriate option. The subject matter experts used all the single-select Likert scale questions from the 2023 MATCH survey to create different combinations of multiple incoming values and the best corresponding recoded value. Please refer to the table below for more details.

Table 5-1: Recommended Recoded Value for Multiple Incoming Responses to Single-Select Questions in PAPI Survey

Question Number in PAPI Survey	Different Combinations of Multiple Incoming Values	Recommended Recoded Value*
Q4	Mostly physical health (1); Mostly mental health (2); Both physical and mental health equally (3) Mostly physical health (1); Both physical and mental health equally (3) Mostly mental health (2); Both physical and mental health equally (3)	Both physical and mental health equally (3)
Q24	Not at all (3); Does not apply (4)	Does not apply (97)
Q43	You do not pay rent or a mortgage (1); Stayed the same (3)	You do not pay rent or a mortgage (4)
Q48	Sometimes (4); I do not shop for food (6)	I do not shop for food (6)

Question Number in PAPI Survey	Different Combinations of Multiple Incoming Values	Recommended Recoded Value*
Q55	Yes, it's my culture (1); Yes, it's my ethnicity (2) Yes, it's my culture (1); Yes, it's my ethnicity (2); Yes, it's both my culture & ethnicity (3) Yes, it's my culture (1); Yes, it's both my culture & ethnicity (3)	Yes, it's both my culture & ethnicity (3)
Q65	Purchased home with no payments due (3); Inherited home with no payments due (4)	Inherited home with no payments due (4)
Q66	Any combination of multi-response	Select highest level of education
Q100	No (2); Does not apply (3)	Does not apply (97)

* Recoded value after standardization rules have been applied.

Excluded questions

Some questions in the CATI/CAWI survey mode were not present in the PAPI survey mode. The missing values to all PAPI respondents for these excluded questions were coded as 777 to indicate the absence of question. Note that the code 777 was also used to gate-missing values.

Survey disposition

CATI/CAWI and PAPI data were cleaned separately and assigned to a completion level. There were three preliminary completion categories: Complete, Partial Complete, and Incomplete. If a respondent answered at least 52 (≥ 95%) out of the 54 complete-tag questions and at least four out of six of the weighting questions, then a survey was considered 'Complete'. If a respondent answered at least 33 (≥ 95%) out of the 34 partial-tag questions or at least 38 (≥ 70%) out of the 55 complete-tag questions along with at least four out of the six weighting questions, then a survey was considered 'Partial Complete'. Please refer to the table in Section 4.7 Table 4-2: Questions and Accepted Responses Involved in Survey Completion Assessment for the list of survey questions that were tagged for completion, partial completion, and weighting.

After cleaning CATI/CAWI and PAPI data separately, they were combined into a single dataset using survey disposition for further processing. Survey disposition was developed using survey eligibility and survey completion. There were three survey eligibility categories: Eligibility (Presumed), Not Eligible, or Unknown Eligibility. Undeliverable sampled addresses had one of the three eligibility categories assigned based on AAPOR's disposition definitions. Survey eligibility and completion levels were combined to create seven preliminary survey disposition codes: Complete, Partial Complete, Incomplete, Eligible/Not Started, Unknown Eligibility, Not Eligible, and Started/Completed by Ineligible Respondent.

The disposition code "Started/Completed by Ineligible Respondent" was assigned to respondents aged 0 to 18 using the age variable and were excluded from the final dataset.

Addresses with "Eligible/Not Started", "Unknown Eligibility", or "Not Eligible" eligibility codes were assigned "Not Applicable" for survey mode and were excluded from the final dataset. Remaining respondents were assigned "Complete" or "Partial Complete" or "Incomplete" based on the survey completion.

The processed 2023-2024 MATCH dataset has three categories for mode: CATI, CAWI, and PAPI and two disposition codes: Complete and Partial Complete.

Data standardization

After all the CATI/CAWI and PAPI response data were cleaned and combined into one dataset for final survey mode and disposition codes, the 2023-2024 MATCH cleaned dataset was standardized. The goal of standardization was to improve data quality, increase data reuse potential, and ensure compatibility with various software programs and programming languages including SAS, SPSS, Stata, and R. Standardizing involved creating a structured dataset with variable names, labels, formats, and response codes.

5.4.3 Variable Names

Survey question variable names follow XXT_RRRRRRR_SSSS naming convention. The three-letter prefix conveys the survey module (XX) and variable type (T). The root is limited to seven letters and conveys the survey question's key content. The four-letter suffix only appears for select-all-that-apply, scale, and yes-no questions conveying the response options.

Below are a couple of examples demonstrating the variable naming convention:

- Single-Select Question Variable
  - Variable Name: MHS_MTMTNOW
  - Module: Mental Health (MH)
  - Type: Source (S)
  - Key Content: Taking medicine or treatment (TMT) for mental health(M) currently (NOW)
  - Suffix: Not applicable (single-select response)
- Multi-Select Question Variable
  - Variable Names: DMR_GENDER_MALE, DMR_GENDER_FMLE, DMR_GENDER_TRNS, DMR_GENDER_OTHR
  - Module: Demographics (DM)
  - Type: Recode (R)
  - Key Content: Gender (GENDER)
  - Suffixes: Male, Female

Variable type

Variable types were determined by the data cleaning and processing steps applied to incoming variables. There were six possible variable types in the 2023-2024 MATCH dataset: source (S), derived (D), recoded (R), constructed (C), imputed (I), analytical (A), and question-alignment (Q). All variable types had basic data cleaning steps and standardized codes. When variables

had multiple data processing transformations, the variable label indicates all and any transformation while the variable name indicated the main step.

Table 5-2: Variable Types and Definitions

Variable Type	Definition
Source	Reflects the survey instrument question and response values with standardized response codes.
Recoded	Reflects the survey instrument question and responses values, which are reordered to have standardized response codes.
Derived	Has standardized final values that are coded from responses to follow-up questions.
Constructed	Has standardized final values that are built from constructing new response by combining one or more source variables.
Imputed	Has standardized final values that are built from specific demographic variables using imputation techniques.
Question-alignment	Has recoded values to align with previous fielding's question type. This recoding was applied to questions that changed from select-all-that-apply to yes-no format. The recoding was to align to the previous and next fielding's format which was select-all-that-apply.

Variable labels

The standardized variable labels include the up to three-digit question number (range: 1.00-111.00), complete question and responses for select-all-that-apply questions, variable type, and any other important critical information not conveyed in the variable name. It is important to note that a variable label does not replace the question in the survey instrument. Due to character count limitations, certain descriptive information was omitted from the survey question when creating the variable label for the codebook, while ensuring that the original meaning of the question was preserved.

Below are a couple of examples demonstrating the variable label convention:

- 1.00 Are you completing this survey for yourself or for someone else in your household?
    - For single-select questions, the question number is followed by a '00' after the period.

- 15.02 What were the reasons why you did not get the medical care you needed in the past 12 months? I did not have insurance coverage.
    - For multi-select questions, the two-digit number after the period indicated the option number for multi-response questions. The associated response was added at the end of the question as well.
- 57.66-C02 Reported race in three categories (Constructed)
    - The '66' indicates constructed variable, which was also described at the end of the variable label in parentheses.
- 83.77.01 In the past 12 months, have you gotten sick or had flu-like symptoms when trying to cut down or stop using any of the following? Marijuana (Web only)
    - The '77' indicates that variable is missing from a survey mode, which was also described at the end of the variable label in parentheses.
- 9.01-D What kinds of health insurance or healthcare coverage do you have? A plan purchased through your or someone else's employer or union (Derived from gate)
    - Variable type was indicated by a single letter after the question number; separated by a dash. It was also mentioned at the end of the label in parentheses.
- 57.66-C02 Reported race in three categories (Constructed)
    - If multiple variables are created for a particular variable type, the letter is followed by a two-digit number indicating the number and order.

Variable response codes

The response codes of incoming CATI/CAWI and PAPI datasets were non-standardized and varied by question type. The table below shows standardized response codes to which all incoming raw values will be reassigned.

Table 5-3: Standardized Variable Response Codes

Standardized Response Code	Response Value/Description
1	Yes
2	No (for non-analytical variables)
97	Does not apply
98	I do not know
99	Prefer not to answer
777	Gated-Missing
999	Missing

Variable format

Standardized format names for Yes/No responses, Likert scale responses, and other common or variable specific responses have been derived from the variable names and other well-researched abbreviations. The standardized format names for each variable can be found in the format column of the codebook.

Quality control

The entire data cleaning process was validated using a double programming approach. A primary programmer developed the cleaning specifications (programming instructions) and wrote the corresponding program. A second analyst then reviewed and validated both the specifications and the program independently. Lastly, the final dataset and specifications were thoroughly reviewed both internally and by third-party vendors to ensure accuracy and consistency.

5.5 Imputation

Imputations and constructions were performed ensuring complete inputs for alignment with ACS population benchmarks, ensuring complete inputs for alignment with population benchmarks and supporting the generation of representative weighted estimates.

Logical rules and derived variables such as Lives Alone and the number of adults in a household[19], were applied as needed to maintain internal consistency and support subsequent weighting procedures.

5.5.1 Sequential Hot Deck Imputation

Missing values were imputed using a sequential hot deck approach implemented in SAS PROC SURVEYIMPUTE with unweighted selection (UWSHD). Donor values were drawn from cases with similar characteristics, and the dataset was ordered by key variable to maximize similarity between donors and recipients.

Imputation was conducted in a structured sequence to ensure completeness of all weighting variables, including age, sex, race, Hispanic origin, marital status, education, and household composition. Previously imputed variables were used as auxiliary inputs for subsequent imputations to improve consistency and predictive accuracy.

Logical adjustments were applied where necessary. For example, Lives Alone was assigned "No" if Living with Spouse was imputed as "Yes." The number of adults (n_h) multiplier for weighting (see 5.6, used an estimator for imputation. The assumption was that living with a spouse (S=Y), parent (P=Y), or grandparent (GP=Y) indicated at least one additional adult per yes – in addition to the respondent – "living arrangement." This was combined with the DMD_RESADLT to form a lower bound – capped at four. If the DMD_RESULTS was missing, the estimator on living arrangement was used:[20]

$$n_h = \begin{cases} \min\{\max\{(1 + S = Y + P = Y + GP = Y), DMD\_RESADLT\}, 4\} & not\ missing\ DMD\_RESADLT \\ mean\ n_h\ over\ respondents\ with\ the\ same\ living\ arrangement & otherwise \end{cases}$$

The resulting dataset contains fully imputed demographic and household variables, ready for weighting calibration to ACS-based and demographic totals. This approach ensures a complete and internally consistent dataset suitable for downstream survey analysis.

[19] It is only used in weighting to assign a multiplier for the fact that only one respondent is selected in the address.

[20] A household-level estimator was used to get number of adults in a household; the value was imputed via estimation. Living arrangements assessment ignored missing in those variables.

5.5.2 Variables Imputed

Key sociodemographic and household variables with missing responses were imputed to produce a complete dataset for weighting calibration. Variables were selected for their necessity in the weighting process. Imputed variables are identified in the final dataset with the DMI_ prefix.

Imputations were implemented using an unweighted sequential hot-deck approach (UWSHD) in SAS PROC SURVEYIMPUTE. Donor values were drawn from cases with similar characteristics, and records were ordered by key variable to maximize similarity between donors and recipients. Previously imputed variables were used as auxiliary inputs for subsequent imputations to improve internal consistency. When less detailed variables were required (e.g., age groups, simplified education categories), these were derived post-imputation from the more granular variables.

The main sociodemographic and household variables imputed include:

- Age: DMS_AGE → DMI_AGE
- Sex: DMS_BRTHSEX → DMI_BRTHSEX
- Hispanic origin: DMS_HISPANC → DMI_HISPANC
- Race: DMC_RACEG5 → DMI_RACEG5
- Marital status: DMS_MRTLSTA → DMI_MRTLSTA
- Education: DMS_MAXEDUC → DMI_MAXEDUC
- Live with spouse: DMS_LIVEWTH_SPOU → DMI_LIVEWTH_SPOU
- Live alone: DMS_LIVEWTH_ALNE → TMP_LIVEWTH_ALNE

Table 5-4: Number and Percentage of Missing Data for Imputed Variables

Variable	Nonrespondents	Respondents	Missing
DMS_AGE	42	16,661	0.3
DMS_BRTHSEX	46	16,657	0.3
DMS_HISPANC	107	16,596	0.6
DMC_RACEG5	63	16,640	0.4
DMS_MRTLSTA	92	16,611	0.6
DMS_MAXEDUC	49	16,654	0.3
DMS_LIVEWTH_SPOU	1,124	15,579	6.7
DMS_LIVEWTH_ALNE	217	16,486	1.3

5.6 Weighting

5.6.1 Weighting Process Overview

This section outlines the weighting scheme used to develop analytic survey weights for the MATCH wave 2 data. The weighting procedure consists of a series of sequential adjustments from the construction of household-level design weights to the creation of person-level analytic weights ensuring that survey estimates are representative of the target adult household

population in West Virginia. The first four steps correspond to Horvitz-Thompson inverse-probability weights, while the final calibration step adjusts for residual coverage bias.

5.6.1.1 Weighting Plan Synopsis

The sample weighting process has five steps:

1. Create a base weight
2. Eligibility adjustment
3. Nonresponse adjustment
4. Person-level design weight
5. Smoothed person-level weight
6. Coverage adjustments
7. UWE reduction (analytic weights)

These steps are the same steps as the ABS portion of the weighting in the 2021-2022 MATCH.

5.6.1.2 Key Weighting Differences with 2021-2022 (Wave 1) Fielding

The main difference in weighting from 2021-2022 MATCH to 2023-2024 MATCH was due to the improvements in sampling, namely, the use of a single ABS frame.

The main change due to the difference in sampling was a simplification of process (i.e., no longer a convex combination of the samples coming from the two frames - as there is only one frame).

Other significant changes include:

- Power transformation for the base weights instead of truncation for dealing with excessive weight variation prior to the coverage adjustments
- Moderate change to developing calibration totals from ACS data
- Slightly different calibration categories (e.g., lives alone)
- Final truncation and rescaling of weights

5.6.1.3 Household-Level Design Weights (WTHHDES)

The household-level design weight is the first of the Horvitz-Thompson weights. After weighting the weighted addresses are representative of all residential addresses in WV. The ABS frame used a stratified simple random sample to select housing units within each of the 102 sampling strata (j).

Purpose: The household-level design weight corrects for unequal probabilities of inclusion across stratum for the addresses.

Construction: For batch 1 and 2, a simple random sample without replacement was used to select addresses within each stratum j. Based on this, the inclusion probability of household h in strata π_{hj} was estimated 1 by $\pi_{hj} = \frac{T_j}{S_j}$, where T_j was the total number of addresses in stratum j and S_j was the total number of addresses selected between the soft launch, batch 1 and batch 2 combined. Hence, the weight is estimated via the equation:

$$\boldsymbol{WT_{HHDES}} = \frac{1}{\pi_{hj}} = \frac{S_j}{T_j}.$$

There are two things to note. First, as constructed, the sum of weights over selected addresses in a stratum are equal to the number of addresses in the stratum. Second, the estimates of the weights brushed over a few technicalities: that soft launch did not have allocation to strata, that there was a slight difference in the addresses between the soft launch and batch 1 and batch 2, and that the allocation in batch 2 depended on the observed results of this sample. Given the process, the authors feel these concerns are negligible and this simple approach is preferrable.

5.6.1.4 Household-Level Eligibility Adjustment Weight WT_{HHELGB}

The household-level eligibility weight is the second of the Horvitz-Thompson weights. Following this adjustment, the weights are representative of households when summed across eligible addresses in the sample.

Purpose: The household-level eligibility adjustment accounts for addresses in the frame that do not contain a member of the target population (e.g., the address is vacant or non-existent).

Construction: Details on how USPS undeliverable codes were used in assessing address eligibility are found in the data processing plan. If a selected addresses received one or more undeliverable codes deemed ineligible (e.g., vacant, no such number), they were "dropped" by assigning no weight. Hence, for an address h, the eligibility weight is given by

$$WT_{HHELGB} = \begin{cases} 0 & \text{if } h \text{ is deemed ineligible} \\ WT_{HHDES} & \text{otherwise} \end{cases}.$$

Note that this adjustment results in a lower weighted count in a stratum because the dropped addresses should not be counted.

5.6.1.5 Household-Level Nonresponse Adjustment Weights $\boldsymbol{WT_{HHNR}}$

This is the third of the Horvitz-Thompson weights. Following this adjustment, the weights are representative of households when summed across addresses with a respondent in the sample.

Purpose: The household-level nonresponse adjustment accounts for eligible households which do not have any respondents.

Construction: Addresses without a respondent are "dropped" by assigning no weight. However, as the dropped addresses are presumed to contain an eligible person, the remaining weights are adjusted via a multiplicative factor to account for this loss. For simplicity, the adjustment is done at the stratum level by

$$WT_{HHNR} = \begin{cases} \frac{\sum_{\text{eligible in } j} WT_{HHELGB}}{\sum_{\text{respondent in } j} WT_{HHELGB}} * WT_{HHELGB} & \text{if } h \text{ has a respondent} \\ 0 & \text{otherwise} \end{cases},$$

where, h is the address, j is the stratum containing h, eligible in j represents the set of eligible addresses in j and respondent in j represents the set of addresses in j with a respondent.

Here are a couple of notes. First, this adjustment results in the same sum of weights. Second, there is an implicit assumption here that address level nonresponse depends only on stratum. A more complicated modeling approach could be implemented, but the coverage adjustment corrects for this later in the process and intermediary weights are not used in analysis.

5.6.1.6 Person-Level Design Adjustment Weights $\boldsymbol{WT_P}$

This is the last of the Horvitz-Thompson weights. Following this adjustment, the weights are theoretically representative of adults living in households when summed across addresses with a respondent in the sample.

Purpose: The person-level design adjustment corrects for the reduction in probability of selection for coresident adults because only one adult is chosen per address.

Construction: Under reasonable assumptions, the probability selection of an adult within an address h with n_h coresident adults (conditional on selection of the address) is $\frac{1}{n_h}$. The correction, in a Horvitz-Thompson logic, is used and given by multiplying by the inverse of this factor or

$$WT_P = n_h * WT_{HHNR}.$$

To control for UWE prior to calibration, the estimated n_h is maxed at four in the programming.

Note that although this weight is theoretically representative for the target population — model issues of nonresponse (which, in addition to not being robustly modelled, likely is missing not at random) and households not represented in the frame — it is empirically clear that this selection step introduces bias due to the overrepresentation of female respondents in the final sample.

5.6.1.7 Smoothed Person-Level Weights $\boldsymbol{WT_{P,80,RSC}}$

The person level weight is smoothed and scaled to county totals. Following this weight, population counts agree to target totals for counties.

Purpose: The smoothed person-level weights are to improve the ability to calibrate.

Construction: The first of the construction is to smooth the weights using a power transformation. The power of 0.80 was chosen to maintain recollection of the design while reducing variation to allow for correction. This smoothing weight was:

$$WT_{P,80} = WT_P^{0.8}$$

This smoothed weight was then rescaled at the county level to the American Community Survey's adult household population. The weight was given by:

$$WT_{P,80,RSC} = WT_{P,80} * \frac{T_{CY,ACS}}{\sum_{P\ in\ CY} WT_{P,80}},$$

where P is the individual, CY is the county and $T_{CY,ACS}$ is the adult household population of the county given by the ACS 5-year 2022 table B09021.

5.6.1.8 Person-Level Calibration Weights/Coverage Adjustment $\boldsymbol{WT_{CLB}}$

The smoothed person-level weights are calibrated using external population totals. Following this adjustment, the weights align with known population totals that are designated and are representative of adults living in households (i.e., the target population), correcting for coverage due to frame, nonresponse, and weaknesses in assumptions used above.

Purpose: The person-level calibration corrects for the absence of covariates in the non-response adjustment and other similar issues (e.g., nonapplication of the most recent birthday selection).

Construction: PROC WTADJUST ADJUST = POST is used to calibrate (or rake) to select population totals. The WT_{CLB} is the output of this procedure.

The development of population counts that can be used in calibration are found in section 5.6.4 Final Adjustment and Calibration.

5.6.1.9 Person-Level Analytic Weights/UWE Reduction $\boldsymbol{WT_A}$

A final truncation and rescale is done to obtain the analytic weights. Following this adjustment weights still align closely with known population totals, while having lower unequal weighting effect. The counts agree with county totals.

Purpose: To reduce variation in weights to minimize the impact of outliers, especially relevant for subset analyses.

Construction: To maintain minimal impact on raking totals, the truncation was performed at the 1st and 99th percentile. The truncated weight was given by:

$$WT_{CLB,TR} = \begin{cases} WT_{CLB,P01} & \text{if less than 1st percentile} \\ WT_{CLB} & \text{if between the 1st and 99th percentile,} \\ WT_{CLB,P99} & \text{if greater than the 99th percentile} \end{cases}$$

where $WT_{CLB,P01}$ and $WT_{CLB,P99}$ represent the 1st and 99th percentile of WT_{CLB} respectively.

This was then rescaled at the county level to obtain the analytic weight:

$$WT_A = WT_{CLB,TR} * \frac{T_{CY,ACS}}{\sum_{P\ in\ CY} WT_{CLB,TR}}.$$

The final Unequal Weighting Effect (UWE) was 2.32, and the Kish Effective Sample Size (KESS) was 89.54 for the worst-performing county indicating notable variation in survey weights. The table *Table 10-3: County Level Unequal Weighting Effect and Kish Effective Sample Size* can be found in the Appendix.

5.6.2 Demographic Population Totals

ACS provides demographic and related tables that are used as a gold-standard for population totals. These tables are provided at different geographic levels (e.g., ST, CY, CT, CBG) and can have various target populations in their person level counts (e.g., overall population, group housing totals, household totals) or even have housing units or other more general "populations" as the target.

Conventions Used in Formulas

This subsection outlines the notation conventions applied throughout the formulas, including variable naming for geographic levels (e.g., CBG, CY), demographic dimensions (e.g., sex, age group, education), and data sources (e.g., raw, rescaled, adjusted). It provides a consistent

framework to interpret how ACS data is processed and integrated into the estimation and calibration steps.

Geographic Units: ST - State, CY - County, CT - Census Tract, CBG - Census Block Group

Demographic Dimensions

- Sex (S): Male and Female
- Eight-category age group divided into eight categories (AG8): 18–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, 85+
- Age group divided into six categories (AG6): 18–24, 25–34, 35–44, 45–64, 65–74, 75+
- Age group divided into five categories (AG5): 18–24, 25–34, 35–44, 45–64, 65+
- Age group divided into three categories (AG3): 18–34, 35–64, 65+
- Education level classification with seven categories (EL7): Less than 9th grade, 9th–12th (no diploma), High school graduate, Some college (no degree), Associate degree, Bachelor's degree, Graduate/professional degree.
- Marital status classification with five categories (MS5): Now married, Widowed, Divorced, Separated, Never married.
- Race classification with four categories (R4): White alone, Black or African American alone, Asian alone, Other.
- Ethnicity-Race classification with three categories (ER3): White (not Hispanic), Hispanic or Latino(a), Other.

Notation for Total Counts

The general notation for total population counts is $\mathbf{T}_{*}$, where the * contains the indicators identifying the count (geographic, demographic, type, and source table).

The type of counts include:

- raw (RAW), which is when the reference is total adult population including group housing
- rescale (RSC) is for when the total count is scaled to the adult household population
- adjusted (ADJ) is reserved for the final rescaled total count.

The raw ACS totals and adjusted totals (by construction) agree across tables when applicable, so their table reference is dropped to simplify equations without risk of ambiguity.

The total counts used in formulas include the following list.

- $\mathbf{T}_{\mathbf{CBG,AG3,B09021}}$: Count from ACS Table B09021 at CBG level by age group (AG3).
- $\mathbf{T}_{\mathbf{CBG,S,A8,RAW}}$: Raw counts at CBG level by sex (S) and age group (AG8).
- $\mathbf{T}_{\mathbf{CBG,AG3,RAW}}$: Raw counts at CBG level by age group (AG3).
- $\mathbf{T}_{\mathbf{CY,S,AG8,RAW}}$: Raw counts at CY level by sex (S) and age group (AG8).
- $\mathbf{T}_{\mathbf{CY,S,AG8,MS5,RAW}}$: Raw marital status counts at CY level by sex(S), age group (AG8), and marital status (MS5).
- $\mathbf{T}_{\mathbf{CY,S,AG5,EL7,RAW}}$: Raw education counts at CY level by sex (S), age group (AG5), and education level (EL7).
- $\mathbf{T}_{\mathbf{CY,S,AG8,R4,RAW}}$: Raw race counts at CY level by sex (S), age group (AG8), and race (R4).

- $\mathbf{T_{CY,S,AG8,ER3,RAW}}$: Raw ethnicity-race counts at CY level by sex (S), age group (AG8), and ethnicity race (ER3).
- $\mathbf{T_{CY,S,AG5,RAW}}$: Raw education counts at CY level by sex (S) and age group (AG5).
- $\mathbf{T_{CBG,S,AG8,RSC}}$: Rescaled counts at CBG level by sex (S) and age group (AG8).
- $\mathbf{T_{CY,S,AG8,RSC}}$: Rescaled counts at CY level by sex (S) and age group (AG8).
- $\mathbf{T_{CY,S,AG5,ADJ}}$: Adjusted counts at CY level by sex (S) and age group (AG5).
- $\mathbf{T_{CY,S,AG8,ADJ}}$: Adjusted counts at CY level by sex (S) and age group (AG8).
- $\mathbf{T_{CY,S,AG8,MS5,ADJ}}$: Adjusted counts at CY level by sex (S), age group (AG8), and marital status (MS5).
- $\mathbf{T_{CY,S,AG5,EL7,ADJ}}$: Adjusted counts at CY level by sex (S), age group (AG5), and education level (EL7).
- $\mathbf{T_{CY,S,AG8,R4,ADJ}}$: Adjusted counts at CY level by sex (S), age group (AG8), and race (R4).
- $\mathbf{T_{CY,S,AG8,ER3,ADJ}}$: Adjusted counts at CY level by sex (S), age group (AG8), and ethnicity race (ER3).

Multipliers

Multipliers are rescaling factors used to adjust raw totals to match target population benchmarks. They are applied at various geographic and demographic levels to ensure consistency for eligible totals.

$\mathbf{M_{CY,S,AG8,ADJ}}$: Adjustment multiplier at the county (CY) level for sex (S) and AG8.

Terms Used in Demographic Calibration

This section defines key terms specific to the demographic calibration process, including population groups, table types, and methods used to align survey estimates with ACS demographic benchmarks.

- Adjustable tables: Required population corrections that are more involved than simple additions or subtractions (e.g., rescaling, calibration).
- Calibration: Adjustment of base weights, so the survey's weighted totals align with demographic benchmarks.
- Direct tables: ACS tables where the ACS table's target population coincides with MATCH's target population or can be computed by simple additions or subtractions (e.g., if overall population and group housing totals are known).
- Adult household population: 18 years or older living in households (i.e., excluding group housing).

5.6.2.1 Demographic Count Development Synopsis

The primary source for demographic counts is the ACS 5-Year Estimates (2018–2022). All tables mentioned below. Direct tables (i.e., B26101, B09021) were developed and then used to correct adjustable tables (B01001* family, B01001, B15001, B12002). The B01001 is rescaled and then calibrated. The others use their raw proportions as a multiplier to the adjusted B01001 counts.

The choice of tables is based on availability of items that can be aligned.

5.6.2.2 Significant Differences from 2021- 2022 Fielding

- Lives alone was developed.
- Additional granularity was developed for marital status, race, education level.
- The computation of county level estimates for adjustable tables was made to align on all shared demographics.

5.6.2.3 Differences from the Plan

The ACS Public Use Micro Sample (PUMS) was mentioned as potential supplement but was deemed nonimplementable due to complications in alignment. Lives with spouse/partner was not added because the totals were only available in reference to the householder, which would exclude married/partnered parents or adult children of the householder, for example.

5.6.2.4 Development of Direct Adult Household Population Totals

This section contains development of the adult household population estimates for the direct tables. They are used in later development. There were two tables considered direct: B26101 and B09021.

State-Level Age × Sex Totals (ACS Table B26101)

The ACS Table B26101 is only available at the ST level and has Universe: Total population. It contains an age group (AG8) by sex (S) breakdown with a focus on group housing. The non-group housing (i.e., household population), was derived by subtracting the overall group quarters population totals from the overall total population totals.

County-Level Age × Lives Alone Status (ACS Table B09021)

The ACS Table B09021 is available at all levels (i.e., ST, CY, CT, CBG) and has Universe: Population 18 years and over in households. The universe is the same as our target population, so totals were pulled directly for age group (AG3) and lives alone. Lives with someone counts were derived by subtracting lives alone totals from the respective overall totals. Caution was used in utilizing other counts as they were not mutually exclusive or had references which we could not align.

Although calibration for MATCH used no estimates below county level, these were calculated all the way to CBG level as the subcounty totals were used for area level counts (see *Section 5.6.3 Geographic Population Totals* for ACS tables used) and adjustment of totals.

5.6.2.5 Development of Adjusted Adult Household Population Totals

This section contains the process of developing adult household population totals from adjustable tables: B01001, B01001A, B01001B, B01001D, B01001H, B01001I, B15001, B12001 (see Section 5.1- ACS tables used)

These tables do not provide easy access to the household population, and we adjusted their totals accordingly. B01001 is developed first and used in the adjustments for the other tables.

County-Level Age Group × Sex Totals (ACS Table B01001)

The ACS Table B01001 is available at all levels (i.e., ST, CY, CT, CBG) and has Universe: Total population. The raw totals are rescaled at the CBG level after summing to age group (AG8) by sex (S). For a given S and AG8, the rescaled count is via the following formula:

$$\mathrm{T_{CBG,S,AG8,RSC}} = \mathrm{T_{CBG,S,AG8,RAW}} * \frac{\mathrm{T_{CBG,AG3,B09021}}}{\mathrm{T_{CBG,AG3,RAW}}}..$$

The rescaled totals are then summed to the county level and adjusted via calibration. The adjusted sex by age group totals at the county level are given by:

$$\mathrm{T_{CY,S,AG8,ADJ}} = \mathrm{T_{CY,S,AG8,RSC}} * \mathrm{M_{CY,S,AG8,ADJ}}$$

where the adjustment multiplier $\mathrm{M_{CY,S,AG8,ADJ}}$ is given by SUDAAN PROC WTADJUST, with the ADJUST=POST option with upper bound of $1.002 * \frac{\mathrm{T_{CY,S,AG8,RAW}}}{\mathrm{T_{CY,S,AG8,RSC}}}$, and incoming weight $\mathrm{T_{CY,S,AG8,RSC}}$, model $\mathrm{S} * \mathrm{AG6}\ \ \mathrm{CY} * \mathrm{AG3}$, and POSTWGT totals given by B26101 for $\mathrm{S} * \mathrm{AG6}$ and B09021 for $\mathrm{CY} * \mathrm{AG3}$.

After adjustment, the total counts align with B26101 – age group (AG6) by sex (S) – at the state level retain alignment with B09021 – age group (AG3) – at the county level.

County-Level Education Level × Age Group × Sex Totals (ACS Table B15001)

The ACS Table B15001 is available at sub county levels (i.e., ST, CY, CT) and has Universe: Population 18 years and over. The breakdown is highest education level (EL7) by age group (AG8) by sex (S). The adjusted totals at the county level were given by:

$$\mathrm{T_{CY,S,AG5,EL,ADJ}} = \mathrm{T_{CY,S,AG5,ADJ}} * \frac{\mathrm{T_{CY,S,AG5,EL7,RAW}}}{\mathrm{T_{CY,S,AG5,RAW}}}$$

This formula maintains the original distribution of education levels within each sex and age group based on raw ACS data, while ensuring that the total population for each group matches the adjusted counts from B01001. Following this adjustment, age group (AG5) by sex (S) by county agree with adjusted B01001 totals and the education level proportions within county, sex, age group are consistent with the raw proportions for the same combination.

County-Level Marital Status × Age Group × Sex Totals (ACS Table B12002)

The ACS Table B12002 is available at subcounty levels (i.e., ST, CY, CT) and has Universe: Population 15 years and over. The breakdown is marital status (MS5), by age group (AG8) by sex (S). The adjusted totals at the county level were given by:

$$\mathrm{T_{CY,S,AG8,MS5,ADJ}} = \mathrm{T_{CY,S,AG8,ADJ}} * \frac{\mathrm{T_{CY,S,AG8,MS5,RAW}}}{\mathrm{T_{CY,S,AG8,RAW}}}$$

This formula applies the observed marital status proportions from the raw ACS data to the adjusted population totals by sex and age group, thereby ensuring consistency with the original distribution while maintaining alignment with the calibrated B01001 marginal totals. Following this adjustment, age group (AG8) by sex (S) by county agree with adjusted B01001 totals and

the marital status proportions within county, sex, age group are consistent with the raw proportions for the same combination.

County-Level Race × Age Group × Sex Totals (ACS Tables B01001, B01001A, B01001B, B01001D)

The ACS Tables have the following details:

- B01001 Universe: Total population
- B01001A Universe: People who are White alone
- B01001B Universe: People who are Black or African American alone
- B01001D Universe: People who are Asian alone

The breakdown is race (R4) by age group (AG8) by sex (S). For White, Black or African American alone, and Asian alone the race totals were pulled from the respective table (B01001A, B01001B, B01001D) and summed to the appropriate level. The other category was calculated as raw total for the sex and age group category from B01001 minus the sums from B01001A, B01001B, and B01001D. The adjusted totals at the county level were given by:

$$T_{CY,S,AG8,R4,ADJ} = T_{CY,S,AG8,ADJ} * \frac{T_{CY,S,AG8,R4,RAW}}{T_{CY,S,AG8,RAW}}$$

This formula distributes the adjusted sex-by-age group totals across the four race categories using proportions from the raw ACS data, ensuring that the final estimates reflect the original race composition while conforming to the calibrated B01001 totals. Following this adjustment, age group (AG8) by sex (S) by county agree with adjusted B01001 totals and the race proportions within county, sex, age group are consistent with the raw proportions for the same combination.

County-Level Ethnicity × Age Group × Sex Totals (ACS Tables B01001, B01001H, B01001I)

The ACS Tables have the following details:

- B01001 Universe: Total population
- B01001H Universe: White alone, not Hispanic, or Latino population
- B01001I Universe: People who are Hispanic or Latino

The breakdown is ER3 by AG8 by sex (S). For White only – not Hispanic or Latino(a), Hispanic or Latino(a), the ethnicity-race totals were pulled from the respective tables (B01001H, B01001I) and summed to the appropriate level. The other category was calculated as raw total for the sex and age group category from B01001 minus the sums from B01001H, and B01001I. The adjusted totals at the county level were given by:

$$T_{CY,S,AG8,ER3,ADJ} = T_{CY,S,AG8,ADJ} * \frac{T_{CY,S,AG8,ER3,RAW}}{T_{CY,S,AG8,RAW}}$$

This formula allocates the adjusted sex-by-age group totals to the three ethnicity-race categories based on raw ACS proportions, ensuring that the final estimates retain the original distribution while remaining consistent with the adjusted B01001 totals. Following this adjustment, AG8 by sex (S) by county agree with adjusted B01001 totals and the race

proportions within county, sex, age group are consistent with the raw proportions for the same combination.

5.6.3 Geographic Population Totals

To ensure survey estimates reflect the full West Virginia population, MATCH uses demographic and geographic totals from the U.S. Census American Community Survey (ACS). The ACS provides detailed data that helps calibrate survey weights and align results with population benchmarks. The process for determining the cutoffs was made more systematic in 2023- 2024 MATCH fielding.

5.6.3.1 Constructing Geographic Totals

The ACS geographic tables are constructed to summarize area-level characteristics (e.g., poverty rate, renter status, internet access) at Census Block Group (BG), Census Tract (CT), and County (CY) levels. These totals are developed through a consistent multi-step process across all tables, using standard ACS sources and derived calculations.

5.6.3.2 ACS Geographic Tables Synopsis

The geographic calibration utilized ACS 2022 five-year tables to assign local area characteristics to survey respondents based on their address. Each respondent was linked to their Census Block Group or Census Tract using geographic identifiers. The tables provided area-level rates for poverty status, housing occupancy, internet and computer access, insurance coverage, and housing tenure. These rates were categorized into ranges to classify respondents (e.g., low, middle, or high poverty areas).

Total Population Counts (Denominator for Proportions):

Proportions are calculated using characteristic-specific ACS tables, where both the numerator and denominator come from the same table to ensure consistency.

- The denominator represents the total population of universe relevant for the characteristic (e.g., population for whom status is determined).
- Extracted from the relevant ACS table at each geographic level (BG, CT, CY).
- Filter conditions: LEVEL = "BG" (or CT/CY) were matched using GEO_ID.
- This ensures that proportions are always based on the correct underlying universe defined by ACS for that characteristic.

Characteristic Counts (Numerator):

Characteristic-specific counts are extracted from the same ACS table as the denominator, ensuring validity and comparability.

- The numerator represents the count of population with specific characteristics (e.g., below poverty level, renter status).
- Extracted at the Block Group level and matched geographic identifiers using GEO_ID.

Proportion Calculation

Proportions make area-level variables comparable across geographies of different sizes. Substituting with county-level rates when totals are zero avoids missingness, supporting

complete and consistent downstream analysis. A proportion is computed for each block group to represent the characteristic rate:

Proportion = Characteristic Count (Numerator) / Total Population Count (Denominator)

- If the total count is greater than zero, the proportion is calculated directly.
- If the total count is zero, the county-level proportion is substituted (imputed) to avoid missing data.

PLVL (Proportion Level) Construction

- PLVL flags whether the proportion is directly observed or imputed.
- If the total population count (ACS_TOTAL_N) is greater than zero, PLVL is set to "BG" (indicating a valid Block Group level estimate).
- If the total count is zero, PLVL is set to 'CY', indicating that a county-level imputed value is used instead.

Threshold Classification

To categorize geographic areas:

- Weighted quartiles (Q1 for 25th percentile, Q3 for 75th percentile) are computed using PROC UNIVARIATE in SAS, weighted by the total population count coming from B09021 (VALUE).
- Each area is then assigned a threshold group:
    - 1 = Low: Proportion < Q1
    - 2 = Middle: Proportion between Q1 and Q3
    - 3 = High: Proportion > Q3
- A corresponding descriptive label is assigned (e.g., "POVERTY RATE < Q1", "Q1 <= POVERTY RATE <= Q3", "POVERTY RATE > Q3").

A detailed breakdown table is provided in the appendix, *Table 10-6: Geographic 2022 5-year ACS Tables*.

Table 5-5: Threshold Classification Logic Table

Threshold Group	Rule (Proportion Compared to Quartiles)	Label Example
1 (Low)	Proportion < Q1	"RATE < Q1"
2 (Middle)	Q1 ≤ Proportion ≤ Q3	"Q1 <= RATE <= Q3"
3 (High)	Proportion > Q3	"RATE > Q3"

5.6.4 Final Adjustment and Calibration

This section documents the final calibration step: moderating extreme weights and aligning survey estimates with population benchmarks at state, region, and county levels.

5.6.4.1 Model Specification

The calibration model incorporated key demographic and regional interactions to improve alignment between survey estimates and population benchmarks.

5.6.4.2 Model used for weighting:

*SEXNUM*AGEGP6NUM REGIONBMS*AGEGP6NUM REGIONBBH*AGEGP6NUM REGIONRBF*AGEGP6NUM COUNTYFIPS*SEXNUM COUNTYFIPS*AGEGP3NUM COUNTYFIPS*DMI_LIVEWTH_ALNE SEXNUM*EDLVL7NUM REGIONBMS*EDLVL7NUM REGIONBBH*EDLVL7NUM REGIONRBF* EDLVL7NUM SEXNUM* AGEGP3NUM*MRTL3NUM REGIONBMS*MRTL3NUM REGIONBBH*MRTL3NUM REGIONRBF*MRTL3NUM HISP3NUM RACE4NUM REGIONBMS*ACS_POVERTY_THRS REGIONBBH*ACS_POVERTY_THRS REGIONRBF*ACS_POVERTY_THRS REGIONBMS*ACS_NINTRNT_THRS REGIONBBH*ACS_NINTRNT_THRS REGIONRBF*ACS_NINTRNT_THRS ACS_MCD1964_THRS ACS_NCV1964_THRS ACS_HSNGOCC_THRS ACS_HSE1UNT_THRS*

Table 5-6: Calibration in the Final Model

Model Interactions	Model Description	Source ACS Tables
Age by birth sex, state	18-24, 25-34, 35-44, 45-54, 55-64, 65+ by male, female	2022 5-year ACS table B26101
Age, regional	18-24, 25-34, 35-44, 45-54, 55-64, 65+ by all four BMS Regions, by all six BBH Regions, and by all seven RBF Regions	2022 5-year ACS table B01001
Birth sex, county	male, female by all 55 counties	2022 5-year ACS table B01001
Age, county	18-34, 35-64, 65+ by all 55 counties	2022 5-year ACS table B09021
Lives Alone, county	Lives alone, does not live alone by all 55 counties	2022 5-year ACS table B09021
Education level by birth sex, state	Less than 9th Grade, 9th Grade to 12th Grade-no diploma, Some college- no Degree, associate degree, Bachelor's degree, Graduate or Professional Degree by Male, Female	2022 5-year ACS table B15001
Educational level, regional	Less than high school, high school/GED to associates, Bachelor's degree or more by all four BMS Regions, by all six BBHS Regions, and by all seven RBF Regions	2022 5-year ACS table B15001
Marital status by age by birth sex, state	Married, widowed/divorced/separated, never married by 18-34, 35-64, 65+ by male, female	2022 5-year ACS table B12002

Model Interactions	Model Description	Source ACS Tables
Marital status, regional	married, widowed/divorced/separated, never married by all four BMS Regions, by all six BBH Regions, and by all seven RBF Regions	2022 5-year ACS table B12002
Hispanic, state	White only Non-Hispanic, Hispanic, Other	2022 5-year ACS tables B01001H, B01001I, and B01001
Race, state	White only, Black only, Asian only, Multi racial or Other	2022 5-year ACS tables B01001A, B01001B, B01001D, and B01001
GC Proportion of households Poverty rate in CBG, regional	Poverty rate < 0.0677, 0.0677 ≤ poverty rate ≤ 0.2346 , poverty rate > 0.2346 by all four BMS Regions, by all six BBH Regions, and by all seven RBF Regions	2022 5-year ACS table B17021
GC Proportion of households in CBG without internet access, regional	proportion < 0.0578, 0.0578≤ proportion ≤ 0.1976, proportion > 0.1976 by all four BMS Regions, by all six BBH Regions, and by all seven RBF Regions	2022 5-year ACS table B28002
GC Proportion of 19-64 population in CBG on Medicaid, state	proportion <= 0. 0983, 0. 0983< proportion < 0.2926, proportion >= 0.2926	2022 5-year ACS table B27010
GC Proportion of 19-64 population in CBG without insurance, state	proportion <= 0.0383, 0.0383 < proportion < 0. 1363, proportion >= 0. 1363	2022 5-year ACS table B27010
GC Proportion of Occupied housing Units in CBG, state	proportion <= 0.785, 0.785 < proportion < 0.935, proportion >= 0.935	2022 5-year ACS table B25002
GC Proportion of 1 unit in Housing in CBG, state	proportion <= 0.8447, 0.8447<proportion <1, proportion = 1	2022 5-year ACS table B25032

5.7 Proper Estimation Techniques

In this section, examples of the appropriate was to calculate estimates from a standard analytic data set from the 2023-2024 MATCH survey. Examples of code that can be used in SAS[6], R[7], STATA[8], and SPSS[9] statistical software are presented assuming '999' are appropriately coded as missing. Examples for estimating means or frequencies are presented for respondents told they have Chronic Obstructive Pulmonary Disease (COPD) by a healthcare provider (HTS_NCONDDX_COPD) and self-assessed general health status (HTS_GENHLTH).

5.7.1 Description of HTS_NCONDDX_COPD

"Have you ever been told by a doctor, nurse, or other healthcare provider that you have any of the following conditions:" Respondents were presented with a list of 13 conditions, including "Chronic Obstructive Pulmonary Disease, or COPD." Respondents could answer "Yes" or "No" for each condition. In the examples below "Yes" was coded to the value "1" and "No" was coded to the value "0".

5.7.2 Description of HTS_GENHLTH

In the survey, respondents were asked the question: "In general, how would you describe your health?" The following responses were offered, and only one could be selected:

"Excellent"
"Very good"
"Good"
"Fair"
"Poor"

Each of the above responses was treated as a category.

5.7.3 Example Analysis in SAS Software

```
LIBNAME IN '[location of 2023 MATCH dataset]'

PROC SURVEYFREQ
        DATA = MATCH23;
        WEIGHT WT_A;
        STRATA STRATUM;
        TABLE HTS_NCONDDX_COPD;
RUN;


PROC SURVEYMEANS
        DATA = MATCH23;
        WEIGHT WT_A;
        STRATA STRATUM;
        CLASS HTS_GENHLTH;
        VAR HTS_GENHLTH;
RUN;
```

5.7.4 Example Analysis in R Software

Required packages to install: survey[10]

```
library(survey)
```

The following assumes data are read in and named "match23.dta."

```
options(survey.lonely.psu = "adjust"
```

```
match23.design <- svydesign(
        id = ~1,
        strata = ~STRATUM,
        weights = ~WT_A,
        data = match23.dta)

svymean(~HTS_MCONDDX_COPD,
        match23.design,
        na.rm = TRUE)

svymean(~HTS_GENHLTH,
        match23.design,
        na.rm = TRUE)
```

5.7.5 Example Analysis in STATA Software

```
svyset [pweight = WT_A], strata(STRATUM) singleunit(centered)

svy: mean HTS_MCONDDX_COPD

svy: proportion HTS_GENHLTH
```

5.7.6 Example Analysis in SPSS Software

SPSS syntax for the Syntax Editor:

```
CSPLAN ANALYSIS
/PLAN FILE=[file path to plan file]
/PLANVARS ANALYSISWEIGHT=WT_A
/SRSESTIMATOR TYPE=WOR
/PRINT PLAN
/DESIGN STRATA=STRATUM
/ESTIMATOR TYPE=WR.

CSTABULATE
/PLAN FILE = [file path to plan file]
/TABLES VARIABLES=HTS_MCONDDX_COPD
/CELLS POPSIZE TABLEPCT
/STATISTICS SE
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.

CSTABULATE
/PLAN FILE = [file path to plan file]
/TABLES VARIABLES=HTS_GENHLTH
/CELLS POPSIZE TABLEPCT
/STATISTICS SE
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

6 Appendix A – Glossary of Common Terms

6.1 General Terms

Prevalence - The proportion of a population affected by specific health-related indicators, expressed as a percentage, rate, or frequency.

Population Health - The distribution of health statuses and outcomes among defined groups of individuals.

Respondents - Individuals from the sample who completed the survey.

Sampling Frame - A list of individuals used to select a sample from the larger population.

Sample - The group of potential respondents chosen based on the sampling frame.

Stability - The reliability of prevalence estimates, where stable estimates remain consistent across repeated surveys, while unstable estimates may not accurately reflect true prevalence.

Stratification - A method of analyzing prevalence estimates by dividing data into subgroups to observe differences.

Confidence Intervals (CI) - A range of values representing the level of uncertainty in prevalence estimates, within which the true value is expected to fall, with 95% confidence.

Relative Standard Error (RSE) - A measure of the reliability of a prevalence estimate, calculated by dividing the standard error by the estimated value.

Significance - A measure of whether the difference between two prevalence estimates is likely not due to chance.

6.2 Common Statistical Terms for Survey Sampling

The language and notations used in this report are consistent with standard texts in probability and sampling theory. This includes use of terminology such as sampling frame, target population, primary sampling unit (PSU), and address-based sampling. The statistical terms used commonly throughout this report are defined below:

Yield Rate – In this context, yield rate refers to the proportion of complete survey responses obtained from a sampling of addresses, calculated as the number of complete survey responses divided by the total number of addresses sampled:

(complete survey responses) / (total number of addresses sampled)

The yield rate is used to assess the effectiveness of the sampling methodology in producing valid data for projections. It does not consider the eligibility of the sampled addresses, meaning that it includes all addresses in the denominator, regardless of whether they represent members of the target population. This metric helps determine the overall response output of the sampling effort.

Unequal Weighting Effect (UWE) - Unequal Weighting Effect (UWE) refers to a statistical measure used to evaluate the impact of unequal weighting in survey data or other sampling methods. UWE is calculated using the formula:

$$UWE = \frac{mean\ of\ weights^2}{(mean\ of\ weights)^2}$$

This ratio serves as a global measure of the design effect, which quantifies how much the effective sample size is reduced due to the unequal weighting of respondents. The design effect itself is generally influenced by the specific estimates used in the analysis, but UWE provides a generalized view of this effect. In practical terms, UWE indicates how much the presence of unequal weights affects the effective sample size of a subset of respondents.

6.3 Common Terms for Sampling Data Collection

Batches

MATCH 2023 sampled a total of 88,000 addresses across three batches:

Soft Launch – The initial mailing to 1,000 addresses is known as the soft launch. These addresses were selected using a simple random sample before stratification to identify areas where oversampling was necessary.

Batch 1 – The first sampling batch included 43,500 addresses and was designed based on the yields from the 2021 MATCH to meet minimum county targets. Stratification was applied to conduct oversampling in West Virginia counties with smaller populations, as well as among non-White, Medicaid, and low SES populations.

Batch 2 – The second batch consisted of the remaining 43,500 addresses. The minimum county targets for the allocation design were set at 220, and the stratum allocations were adjusted based on the yields from batch 1 and the soft launch. Like batch 1, oversampling was conducted in counties with smaller populations, as well as among non-White, Medicaid, and low SES populations. However, this second batch placed a greater emphasis on strata with the highest Black/African American yield rates.

Full Launch Sample – The combined total of batch 1 and batch 2 is referred to as the full launch sample.

Survey Modalities

The PAPI questionnaire differed from the CAWI and CATI versions by removing most 'other response' text boxes, except for lifetime job category. PAPI participants could receive a $10 cash incentive by mail but were not offered the option to receive it via email to avoid errors in interpreting handwritten addresses.

CATI (Computer-Assisted Telephone Interviewing) – A data collection method where an interviewer conducts the survey over the phone using a computer system to guide the questions, record responses, and manage the survey process.

CAWI (Computer-Assisted Web Interviewing) – A self-administered survey method where participants complete the questionnaire online via a web-based interface, with the computer assisting by presenting questions and recording answers.

PAPI (Pen and Paper Instrument) – A traditional survey method where participants complete a physical paper questionnaire by hand and return it by mail or another collection method for data entry.

7 Appendix B – Sampling

7.1 Frame Development

This section contains additional details for the frame development.

7.1.1 Additional Details on Frame Sources

This section provides supplemental information on sources for the 2023 MATCH Sampling Frame.

WV MSG ABS Frame

The MSG ABS frame had appended geocodes (latitude/longitude) using industry accredited GIS data from TomTom and Census TIGER. Geocoding provides a one-to-one link between postal geography and census geography. Census geography (FIPS state/county, Census Tract, Block Group, Census Block) was appended to each address using the latitude and longitude derived from geocoding.

The residential WV MSG ABS frame provided for 2023 MATCH includes all WV residential addresses – i.e., full or primarily business addresses were excluded – from an August 2023 update.

The addresses were appended with MSG derived household flags on likely race/ethnicity (Hispanic/Latino(a), Asian, or Black), household size (adult and children), and household income. The household flags were derived by matching the entire WV MSG ABS Frame to multiple consumer files identifying all individuals at each address. Approximately 84% of non-P.O. Box addresses were matched.

The three race/ethnicity flags, Hispanic/Latino(a), Asian, or Black, were derived using a mixture of MSG proprietary data and commercially available consumer modeled data. MSG has proprietary ethnic surname tables for Hispanic and several Asian subgroups (e.g., Chinese, Japanese, Korean, Thai, and Asian Indian). In addition, MSG used multiple commercially available consumer files that had individual level race/ethnic indicators derived from models. These models included rules for first names, surnames, name prefixes and suffixes, and geographic criteria to identify an individual's race/ethnicity. If *any* of the individuals at a given address had a surname matched to an MSG surname file or a modeled Hispanic/Latino(a), Black, or Asian code, the respective household flag was set to yes.

The number of adults aged 18 years and over in a household was derived from the number of individuals matched to a particular address. The number of children aged less than 18 years in a household came from a combination of modeled data and reported data collected from syndicated surveys. Household income was a modeled[21] value using multiple sources to derive the estimate.

One-Year Medicaid Roster

The one-year Medicaid roster contained adult enrollees aged 18 years and over from March 1, 2022 to July 31, 2023. Only the most recent record for the Medicaid ID[22] (representing an

[21] This was a proprietary model performed by MSG; details are unavailable.
[22] The actual Medicaid ID is replaced with an alternative identifier for the provided roster.

individual) was kept for the address. Each record comes with both a mailing address and a physical address.

The raw roster came with a race characteristic that had the following options: Caucasian, African American, Asian/Pacific Islander, American Indian/Alaska Native, Other. The other contained Unknown and for this reason, it was not included in the household non-White flag.

Additional characteristics in the roster that were not used in the frame construction are age and sex/gender.

7.1.2 Determination of Likely Low SES

Low SES was based on 200% of FPL. The following table provides detailed assessment based on the estimated values associated with the address from the WV MSG ABS Frame.

Table 7-1: Construction of Low SES Based on ASPE's 200% FPL Guidelines

Household/Family Size*	200% FPL Gross or Net Annual Income	Household Income Threshold Used for Low SES
1	$29,160	<= $25,000
2	$39,440	<= $35,000
3	$49,720	<= $45,000
4	$60,000	<= $55,000
5	$70,280	<= $65,000
6	$80,560	<= $75,000
7	$90,840	<= $85,000
8+	$101,120	<= $95,000
Table Notes: *Calculated from the number of adults and children. If there was no available estimate of the household size, the size of one was applied.		

7.1.3 Substratum Types

Table 7.2 contains an interpretive description of the substratum types.

Table 7-2: Substratum Types Descriptions

Substratum Type	Description
NMW	Non-Medicaid/low SES, White
MW	Medicaid/low SES, White
NMNW	Non-Medicaid/low SES, non-White
MNW	Medicaid/low SES, non-White

Substratum Type	Description
NMNW or MNW	Non-White
NMNW or MW or MNW	Non-White or Medicaid/low SES

7.2 Sampling Counts and Allocation Details

This section contains additional details for the sampling and allocation.

7.2.1 BMS Regions

County to BMS regional breakdown is contained in the following table.

Table 7-3: BMS Regional Breakdown

BMS Region Number	Counties Contained in the Region
Region 1	Brooke, Calhoun, Clay, Doddridge, Gilmer, Hancock, Jackson, Marion, Marshall, Monongalia, Ohio, Pleasants, Ritchie, Roane, Tyler, Wetzel, Wirt, Wood
Region 2	Boone, Cabell, Kanawha, Lincoln, Logan, Mason, Putnam, Wayne
Region 3	Barbour, Berkeley, Braxton, Grant, Hampshire, Hardy, Harrison, Jefferson, Lewis, Mineral, Morgan, Pendleton, Preston, Randolph, Taylor, Tucker, Upshur
Region 4	Fayette, Greenbrier, McDowell, Mercer, Mingo, Monroe, Nicholas, Pocahontas, Raleigh, Summers, Webster, Wyoming

7.2.2 Batch 1 Preliminary Allocation Based on an 87,000 Address Release

This table contains the application of the algorithm for initial allocation of addresses to strata for batch 1. The process involves allocating addresses to counties is addresses, then to substrata. The final batch 1 allocations were a small correction of the rounded halved totals.

Table 7-4: County and Stratum Batch 1 Allocation Based on 87,000 Addresses

County						Substratum				
Name	Type	ALL	Target	YR21	87k	Name	ALL	K_j	87k	B1
Barbour	Type F	6,970	220	0.1945	1,130.9	No substrata	6,970	-	1,130.9	565
Berkeley	Type A*	53,739	-	-	3,799.0	NMW	29,954	1.00	1,465.1	733
						MW	15,791	1.25	965.4	483
						NMNW	3,445	3.50	589.7	295
						MNW	4,549	3.50	778.7	389
Boone	Type F	6,542	220	0.1354	1,624.2	No substrata	6,542	-	1,624.2	812
Braxton	Type F	5,836	220	0.2076	1,059.7	No substrata	5,836	-	1,059.7	530
Brooke	Type E*	10,676	250	0.1875	1,333.3	NMW	6,160	1.00	635.0	318
						NMNW or MW or MNW	4,516	1.50	698.3	349
Cabell	Type C*	45,721	-	-	3,232.0	NMW	25,069	1.00	1,340.2	670
						MW	16,398	1.25	1,095.8	548
						NMNW or MNW	4,254	3.50	796.0	398
Calhoun	Type F	2,867	220	0.1886	1,166.2	No substrata	2,867	-	1,166.2	583
Clay	Type F	3,476	220	0.1652	1,331.8	No substrata	3,476	-	1,331.8	666
Doddridge	Type F	3,211	220	0.1992	1,104.2	No substrata	3,211	-	1,104.2	552
Fayette	Type E*	16,710	250	0.1901	1,315.1	NMW	8,300	1.00	521.9	261
						NMNW or MW or MNW	8,410	1.50	793.2	397

County						Substratum				
Name	Type	ALL	Target	YR21	87k	Name	ALL	K_j	87k	B1
Gilmer	Type F	2,884	220	0.2051	1,072.9	No substrata	2,884	-	1,072.9	536
Grant	Type F	5,218	220	0.2067	1,064.5	No substrata	5,218	-	1,064.5	532
Greenbrier	Type E*	15,099	250	0.1901	1,315.2	NMW	8,571	1.00	613.9	307
						NMNW or MW or MNW	6,528	1.50	701.3	351
Hampshire	Type E	9,596	230	0.1970	1,167.4	NMW	5,289	1.00	643.4	322
						NMNW or MW or MNW	4,307	1.00	524.0	262
Hancock	Type E*	13,984	250	0.1768	1,413.9	NMW	7,822	1.00	648.1	324
						NMNW or MW or MNW	6,162	1.50	765.8	383
Hardy	Type F	6,256	220	0.2082	1,056.8	No substrata	6,256	-	1,056.8	528
Harrison	Type C*	31,628	-	-	2,235.0	NMW	17,973	1.00	1,000.5	500
						MW	11,384	1.25	792.1	396
						NMNW or MNW	2,271	3.50	442.4	221
Jackson	Type E*	12,652	250	0.1924	1,299.2	NMW	6,537	1.00	540.6	270
						NMNW or MW or MNW	6,115	1.50	758.6	379
Jefferson	Type A*	23,204	-	-	1,640.0	NMW	14,655	1.00	733.3	367
						MW	5,245	1.25	328.1	164
						NMNW	1,763	3.50	308.8	154
						MNW	1,541	3.50	269.9	135
Kanawha	Type A*	91,075	-	-	6,438.0	NMW	50,338	1.00	2,630.5	1,315
						MW	30,985	1.25	2,023.9	1,012
						NMNW	3,204	3.50	586.0	293

County						Substratum				
Name	Type	ALL	Target	YR21	87k	Name	ALL	Kj	87k	B1
						MNW	6,548	3.50	1,197.6	599
Lewis	Type F	7,881	220	0.2198	1,000.9	No substrata	7,881	-	1,000.9	500
Lincoln	Type E	8,592	230	0.1842	1,248.6	NMW	3,763	1.00	546.8	273
						NMNW or MW or MNW	4,829	1.00	701.7	351
Logan	Type E*	12,687	250	0.1257	1,989.3	NMW	6,563	1.00	829.0	414
						NMNW or MW or MNW	6,124	1.50	1,160.3	580
Marion	Type C*	26,328	-	-	1,861.0	NMW	15,024	1.00	827.0	414
						MW	9,236	1.25	635.5	318
						NMNW or MNW	2,068	3.50	398.4	199
Marshall	Type E*	14,556	250	0.2024	1,235.1	NMW	8,535	1.00	600.1	300
						NMNW or MW or MNW	6,021	1.50	635.0	317
Mason	Type E*	10,852	250	0.1706	1,465.1	NMW	5,541	1.00	601.0	301
						NMNW or MW or MNW	5,311	1.50	864.1	432
McDowell	Type E	9,133	230	0.1107	2,077.8	NMW	5,150	1.00	1,171.7	586
						NMNW or MW or MNW	3,983	1.00	906.2	453
Mercer	Type C*	28,333	-	-	2,003.0	NMW	13,248	1.00	693.0	347
						MW	12,336	1.25	806.7	403
						NMNW or MNW	2,749	3.50	503.3	252
Mineral	Type E*	11,538	250	0.2292	1,090.9	NMW	6,688	1.00	522.5	261

County						Substratum				
Name	Type	ALL	Target	YR21	87k	Name	ALL	K_j	87k	B1
						NMNW or MW or MNW	4,850	1.50	568.4	284
Mingo	Type E	10,005	230	0.1080	2,130.5	NMW	4,931	1.00	1,050.0	525
						NMNW or MW or MNW	5,074	1.00	1,080.5	540
Monongalia	Type A*	51,476	-	-	3,639.0	NMW	33,671	1.00	1,839.9	920
						MW	13,064	1.25	892.3	446
						NMNW	2,557	3.50	489.0	245
						MNW	2,184	3.50	417.7	209
Monroe	Type F	5,106	220	0.2356	1,000.0	No substrata	5,106	-	1,000.0	500
Morgan	Type E	8,305	230	0.1873	1,228.1	NMW	4,883	1.00	722.1	361
						NMNW or MW or MNW	3,422	1.00	506.0	253
Nicholas	Type E	9,947	230	0.2074	1,108.8	NMW	4,665	1.00	520.0	260
						NMNW or MW or MNW	5,282	1.00	588.8	294
Ohio	Type C*	20,777	-	-	1,468.0	NMW	12,536	1.00	692.2	346
						MW	6,575	1.25	453.8	227
						NMNW or MNW	1,666	3.50	322.0	161
Pendleton	Type F	2,821	220	0.2085	1,055.2	No substrata	2,821	-	1,055.2	528
Pleasants	Type F	2,951	220	0.1927	1,141.9	No substrata	2,951	-	1,141.9	571
Pocahontas	Type F	3,737	220	0.2142	1,027.3	No substrata	3,737	-	1,027.3	514
Preston	Type E*	13,183	250	0.2539	1,000.0	NMW	7,353	1.00	456.8	228

County						Substratum				
Name	Type	ALL	Target	YR21	87k	Name	ALL	K_j	87k	B1
						NMNW or MW or MNW	5,830	1.50	543.2	272
Putnam	Type C*	25,608	-	-	1,810.0	NMW	16,615	1.00	951.2	476
						MW	7,322	1.25	524.0	262
						NMNW or MNW	1,671	3.50	334.8	167
Raleigh	Type C*	31,028	-	-	2,193.0	NMW	16,132	1.00	839.6	420
						MW	11,614	1.25	755.6	378
						NMNW or MNW	3,282	3.50	597.8	299
Randolph	Type E*	11,528	250	0.1959	1,276.0	NMW	5,866	1.00	521.3	261
						NMNW or MW or MNW	5,662	1.50	754.7	377
Ritchie	Type F	4,187	220	0.1918	1,146.9	No substrata	4,187	-	1,146.9	573
Roane	Type F	6,532	220	0.2039	1,078.8	No substrata	6,532	-	1,078.8	539
Summers	Type F	5,049	220	0.1851	1,188.4	No substrata	5,049	-	1,188.4	594
Taylor	Type F	7,304	220	0.2186	1,006.2	No substrata	7,304	-	1,006.2	503
Tucker	Type F	2,754	220	0.2604	1,000.0	No substrata	2,754	-	1,000.0	500
Tyler	Type F	3,912	220	0.1794	1,226.5	No substrata	3,912	-	1,226.5	613
Upshur	Type E*	10,462	250	0.2317	1,079.0	NMW	5,107	1.00	419.4	210
						NMNW or MW or MNW	5,355	1.50	659.6	330
Wayne	Type E*	17,912	250	0.1441	1,734.9	NMW	9,028	1.00	700.7	350
						NMNW or MW or MNW	8,884	1.50	1,034.2	517
Webster	Type F	3,388	220	0.2220	1,000.0	No substrata	3,388	-	1,000.0	500

County						Substratum				
Name	Type	ALL	Target	YR21	87k	Name	ALL	K_j	87k	B1
Wetzel	Type F	7,048	220	0.1866	1,178.7	No substrata	7,048	-	1,178.7	589
Wirt	Type F	2,235	220	0.1891	1,163.2	No substrata	2,235	-	1,163.2	582
Wood	Type C*	40,327	-	-	2,850.0	NMW	21,726	1.00	1,219.4	610
						MW	16,023	1.25	1,124.2	562
						NMNW or MNW	2,578	3.50	506.4	253
Wyoming	Type F	4,997	220	0.1473	1,493.2	No substrata	4,997	-	1,493.2	747
TOTALS		809,823	-	-	86,994.5		809,823	-	86,994.5	43,496

NOTES: Type=Sub-stratification type (=oversampling eligible); ALL=Total number of addresses in the 2023 MATCH ABS Sampling Frame; Target=Target number of completed surveys for the county, if applicable; YR21=2021 Yield Rate (for counties with targets); 87k=Allocation based on 87,000 addresses; k_j=Oversampling factor (1.00=reference or no oversampling); B1=Preliminary batch 1 allocation (rounded one-half of substratum 87k allocation); - =Not applicable;*

7.2.3 Batch 2 County Factors and Final Targeted Completions with Stratum Black or African American Yield Rate Classifications and Oversampling Factors

Table 7-5: Batch 2 Final County High and Low Targeted Completions with County and Stratum Factors

County					Substratum			
Name	Type	All	Factor	Analytic Target	Name	ALL	BAA CLSS [1]	k_j
Barbour	Type F	6,970		220.00	No substrata	6,970	Low	
Berkeley	Type A*	53,739	1.3805	729.27	NMW	29,954	Low	1
					MW	15,791	Low	1
					NMNW	3,445	Med	5
					MNW	4,549	High	11.75
Boone	Type F	6,542		220.00	No substrata	6,542	Low	
Braxton	Type F	5,836		220.00	No substrata	5,836	Low	
Brooke	Type E*	10,676		230.85	NMW	6,160	Low	1
					(NMNW MW MNW)	4,516	Low	1
Cabell	Type C*	45,721	1.2603	586.68	NMW	25,069	Low	1
					MW	16,398	Low	1
					(NMNW MNW)	4,254	Med	10
Calhoun	Type F	2,867		220.00	No substrata	2,867	Low	
Clay	Type F	3,476		220.00	No substrata	3,476	Low	
Doddridge	Type F	3,211		220.00	No substrata	3,211	Low	
Fayette	Type E*	16,710		238.48	NMW	8,300	Low	1
					(NMNW MW MNW)	8,410	Low	1
Gilmer	Type F	2,884		220.00	No substrata	2,884	Low	
Grant	Type F	5,218		220.00	No substrata	5,218	Low	
Greenbrier	Type E*	15,099		236.44	NMW	8,571	Low	1

County					Substratum			
Name	Type	All	Factor	Analytic Target	Name	ALL	BAA CLSS [1]	k_j
					(NMNW MW MNW)	6,528	Low	1
Hampshire	Type E	9,596		230.00	NMW	5,289	Low	1
					(NMNW MW MNW)	4,307	Low	1
Hancock	Type E*	13,984		235.03	NMW	7,822	Low	1
					(NMNW MW MNW)	6,162	Low	1
Hardy	Type F	6,256		220.00	No substrata	6,256	Low	
Harrison	Type C*	31,628	1.2066	403.71	NMW	17,973	Low	1
					MW	11,384	Low	1
					(NMNW MNW)	2,271	Med	10
Jackson	Type E*	12,652		233.35	NMW	6,537	Low	1
					(NMNW MW MNW)	6,115	Low	1
Jefferson	Type A*	23,204	1.2334	304.57	NMW	14,655	Low	1
					MW	5,245	Med	1
					NMNW	1,763	Low	1.25
					MNW	1,541	High	11.75
Kanawha	Type A*	91,075	1.3033	1,183.27	NMW	50,338	Low	1
					MW	30,985	Low	1
					NMNW	3,204	Med	5
					MNW	6,548	High	11.75
Lewis	Type F	7,881		220.00	No substrata	7,881	Low	
Lincoln	Type E	8,592		230.00	NMW	3,763	Low	1
					(NMNW MW MNW)	4,829	Low	1

County					Substratum			
Name	Type	All	Factor	Analytic Target	Name	ALL	BAA CLSS [1]	k_j
Logan	Type E*	12,687		233.39	NMW	6,563	Low	1
					(NMNW MW MNW)	6,124	Low	1
Marion	Type C*	26,328	1.2240	341.90	NMW	15,024	Low	1
					MW	9,236	Low	1
					(NMNW MNW)	2,068	High	10
Marshall	Type E*	14,556		235.76	NMW	8,535	Low	1
					(NMNW MW MNW)	6,021	Low	1
Mason	Type E*	10,852		231.07	NMW	5,541	Low	1
					(NMNW MW MNW)	5,311	Low	1
McDowell	Type E	9,133		230.00	NMW	5,150	Low	1
					(NMNW MW MNW)	3,983	High	1
Mercer	Type C*	28,333	1.2700	370.92	NMW	13,248	Med	1
					MW	12,336	Med	1
					(NMNW MNW)	2,749	High	10
Mineral	Type E*	11,538		231.94	NMW	6,688	Low	1
					(NMNW MW MNW)	4,850	Low	1
Mingo	Type E*	10,005		230.00	NMW	4,931	Low	1
					(NMNW MW MNW)	5,074	Low	1
Monongalia	Type A*	51,476	1.0000	572.02	NMW	33,671	Low	1
					MW	13,064	Low	1
					NMNW	2,557	Low	1.25
					MNW	2,184	Low	2.5

County					Substratum			
Name	Type	All	Factor	Analytic Target	Name	ALL	BAA CLSS [1]	k_j
Monroe	Type F	5,106		220.00	No substrata	5,106	Low	
Morgan	Type E	8,305		230.00	NMW	4,883	Low	1
					(NMNW MW MNW)	3,422	Low	1
Nicholas	Type E	9,947		230.00	NMW	4,665	Low	1
					(NMNW MW MNW)	5,282	Low	1
Ohio	Type C*	20,777	1.2281	274.81	NMW	12,536	Low	1
					MW	6,575	Low	1
					(NMNW MNW)	1,666	Med	10
Pendleton	Type F	2,821		220.00	No substrata	2,821	Low	
Pleasants	Type F	2,951		220.00	No substrata	2,951	Low	
Pocahontas	Type F	3,737		220.00	No substrata	3,737	Low	
Preston	Type E*	13,183		234.02	NMW	7,353	Low	1
					(NMNW MW MNW)	5,830	Low	1
Putnam	Type C*	25,608	1.0000	316.13	NMW	16,615	Low	1
					MW	7,322	Low	1
					(NMNW MNW)	1,671	Low	2.5
Raleigh	Type C*	31,028	1.2910	407.50	NMW	16,132	Low	1
					MW	11,614	Low	1
					(NMNW MNW)	3,282	High	10
Randolph	Type E*	11,528		231.93	NMW	5,866	Low	1
					(NMNW MW MNW)	5,662	Low	1
Ritchie	Type F	4,187		220.00	No substrata	4,187	Low	
Roane	Type F	6,532		220.00	No substrata	6,532	Low	

County					Substratum			
Name	Type	All	Factor	Analytic Target	Name	ALL	BAA CLSS [1]	k_j
Summers	Type F	5,049		220.00	No substrata	5,049	Low	
Taylor	Type F	7,304		220.00	No substrata	7,304	Low	
Tucker	Type F	2,754		220.00	No substrata	2,754	Low	
Tyler	Type F	3,912		220.00	No substrata	3,912	Low	
Upshur	Type E*	10,462		230.58	NMW	5,107	Low	1
					(NMNW MW MNW)	5,355	Low	1
Wayne	Type E*	17,912		240.00	NMW	9,028	Low	1
					(NMNW MW MNW)	8,884	Low	1
Webster	Type F	3,388		220.00	No substrata	3,388	Low	
Wetzel	Type F	7,048		220.00	No substrata	7,048	Low	
Wirt	Type F	2,235		220.00	No substrata	2,235	Low	
Wood	Type C*	40,327	1.1859	502.96	NMW	21,726	Low	1
					MW	16,023	Low	1
					(NMNW MNW)	2,578	Med	10
Wyoming	Type F	4,997		220.00	No substrata	4,997	Low	

NOTES: Type=Sub-stratification type (*=oversampling eligible); ALL=Total number of addresses in the 2023 MATCH ABS Sampling Frame; Analytic Target=Target number of completed surveys for the county used to estimate allocation; Factor=The county factor used in establishing the target (see Section 7); BAA CLSS = Black or African American Yield Rate Classification; k_j=Oversampling factor (1.00=reference or no oversampling); - =Not applicable;

[1] Black or African American Yield Rate Classification were as follows: High=Black or African American yield rate greater than 0.02 at the time of batch 2 sampling; Med =Black or African American yield rate greater than 0.01 at the time of batch 2 sampling (but not High Yield); Low=Black or African American yield rate at or below 0.01 at the time of batch 2 sampling.

7.2.4 Final Allocations for Batch 1 and Batch 2 with Soft Launch Sampling Totals

The following tables contains the final batch 1 and batch 2 allocation to strata and counties. Although the process of assigning preliminary allocation went from county to stratum, the final allocation was assigned to the strata and hence the county totals are simply the sum of the allocation to its substrata. Soft launch counts are also provided.

Table 7-6: County Soft Launch Sampling Totals and Batch 1 and Batch 2 Final Allocations

County	ALL	B1 [1]	SL	B2
Barbour	6,970	565	6	874
Berkeley	53,739	1,901	67	2,157
Boone	6,542	812	9	474
Braxton	5,836	530	11	471
Brooke	10,676	667	15	714
Cabell	45,721	1,616	53	2,224
Calhoun	2,867	583	1	523
Clay	3,476	666	8	654
Doddridge	3,211	552	7	639
Fayette	16,710	658	22	797
Gilmer	2,884	536	4	526
Grant	5,218	532	8	622
Greenbrier	15,099	658	24	470
Hampshire	9,596	584	17	595
Hancock	13,984	707	14	745
Hardy	6,256	528	11	451
Harrison	31,628	1,117	36	864
Jackson	12,652	649	11	804
Jefferson	23,204	821	27	742
Kanawha	91,075	3,220	111	3,290
Lewis	7,881	500	8	584
Lincoln	8,592	624	11	922
Logan	12,687	994	17	1,042
Marion	26,328	931	33	741
Marshall	14,556	617	14	653
Mason	10,852	733	12	599
McDowell	9,133	1,039	13	1,040
Mercer	28,333	1,002	29	1,219
Mineral	11,538	545	9	517
Mingo	10,005	1,065	12	1,246
Monongalia	51,476	1,820	69	691
Monroe	5,106	500	2	929

County	ALL	B1 [1]	SL	B2
Morgan	8,305	614	10	676
Nicholas	9,947	554	6	521
Ohio	20,777	734	26	500
Pendleton	2,821	528	3	436
Pleasants	2,951	571	3	494
Pocahontas	3,737	514	7	731
Preston	13,183	500	13	536
Putnam	25,608	905	35	780
Raleigh	31,028	1,098	43	1,003
Randolph	11,528	638	17	605
Ritchie	4,187	573	2	523
Roane	6,532	539	8	479
Summers	5,049	594	6	622
Taylor	7,304	503	8	551
Tucker	2,754	500	4	569
Tyler	3,912	613	6	602
Upshur	10,462	540	18	533
Wayne	17,912	867	14	514
Webster	3,388	500	8	595
Wetzel	7,048	589	7	523
Wirt	2,235	582	3	522
Wood	40,327	1,425	54	922
Wyoming	4,997	747	8	1,444
TOTALS	809,823	43,500	1,000	43,500

NOTES: ALL=Total number of addresses in the 2023 MATCH ABS Sampling Frame; B1=Addresses allocated in batch 1; SL=Addresses randomly sampled in soft launch; B2=Addresses sampled for batch 2; PROP=Proportion

Proportion is defined as the sum of B1, SL, and B2 divided by ALL.

Table 7-7: Stratum Soft Launch Sampling Totals and Batch 1 and Batch 2 Final Allocations

County		Substratum				
Name	**Type**	**Description**	**ALL**	**B1**	**SL**	**B2**
Barbour	Type F	all	6,970	565	6	874
Berkeley	Type A*	non-Medicaid/non-low SES, White	29,954	734	37	555
		Medicaid/low SES, White	15,791	483	18	292
		non-Medicaid/non-low SES, non-White	3,445	295	6	319
		Medicaid/low SES, non-White	4,549	389	6	991
Boone	Type F	all	6,542	812	9	474
Braxton	Type F	all	5,836	530	11	471
Brooke	Type E*	non-Medicaid/non-low SES, White	6,160	318	9	412
		non-White OR Medicaid/low SES	4,516	349	6	302
Cabell	Type C*	non-Medicaid/non-low SES, White	25,069	670	24	663
		Medicaid/low SES, White	16,398	548	25	434
		non-White	4,254	398	4	1,127
Calhoun	Type F	all	2,867	583	1	523
Clay	Type F	all	3,476	666	8	654
Doddridge	Type F	all	3,211	552	7	639
Fayette	Type E*	non-Medicaid/non-low SES, White	8,300	261	10	396
		non-White OR Medicaid/low SES	8,410	397	12	401
Gilmer	Type F	all	2,884	536	4	526
Grant	Type F	all	5,218	532	8	622
Greenbrier	Type E*	non-Medicaid/non-low SES, White	8,571	307	12	267
		non-White OR Medicaid/low SES	6,528	351	12	203
Hampshire	Type E	non-Medicaid/non-low SES, White	5,289	322	13	328
		non-White OR Medicaid/low SES	4,307	262	4	267
Hancock	Type E*	non-Medicaid/non-low SES, White	7,822	324	5	417
		non-White OR Medicaid/low SES	6,162	383	9	328
Hardy	Type F	all	6,256	528	11	451

County		Substratum				
Name	Type	Description	ALL	B1	SL	B2
Harrison	Type C*	non-Medicaid/non-low SES, White	17,973	500	20	298
		Medicaid/low SES, White	11,384	396	11	189
		non-White	2,271	221	5	377
Jackson	Type E*	non-Medicaid/non-low SES, White	6,537	270	6	415
		non-White OR Medicaid/low SES	6,115	379	5	389
Jefferson	Type A*	non-Medicaid/non-low SES, White	14,655	368	17	270
		Medicaid/low SES, White	5,245	164	5	97
		non-Medicaid/non-low SES, non-White	1,763	154	3	41
		Medicaid/low SES, non-White	1,541	135	2	334
Kanawha	Type A*	non-Medicaid/non-low SES, White	50,338	1,316	62	950
		Medicaid/low SES, White	30,985	1,012	42	585
		non-Medicaid/non-low SES, non-White	3,204	293	1	302
		Medicaid/low SES, non-White	6,548	599	6	1,453
Lewis	Type F	all	7,881	500	8	584
Lincoln	Type E	non-Medicaid/non-low SES, White	3,763	273	6	404
		non-White OR Medicaid/low SES	4,829	351	5	518
Logan	Type E*	non-Medicaid/non-low SES, White	6,563	414	7	539
		non-White OR Medicaid/low SES	6,124	580	10	503
Marion	Type C*	non-Medicaid/non-low SES, White	15,024	414	16	247
		Medicaid/low SES, White	9,236	318	14	152
		non-White	2,068	199	3	342
Marshall	Type E*	non-Medicaid/non-low SES, White	8,535	300	8	383
		non-White OR Medicaid/low SES	6,021	317	6	270

County		Substratum				
Name	Type	Description	ALL	B1	SL	B2
Mason	Type E*	non-Medicaid/non-low SES, White	5,541	301	3	306
		non-White OR Medicaid/low SES	5,311	432	9	293
McDowell	Type E	non-Medicaid/non-low SES, White	5,150	586	8	586
		non-White OR Medicaid/low SES	3,983	453	5	454
Mercer	Type C*	non-Medicaid/non-low SES, White	13,248	347	7	304
		Medicaid/low SES, White	12,336	403	16	283
		non-White	2,749	252	6	632
Mineral	Type E*	non-Medicaid/non-low SES, White	6,688	261	6	300
		non-White OR Medicaid/low SES	4,850	284	3	217
Mingo	Type E*	non-Medicaid/non-low SES, White	4,931	525	7	614
		non-White OR Medicaid/low SES	5,074	540	5	632
Monongalia	Type A*	non-Medicaid/non-low SES, White	33,671	920	42	420
		Medicaid/low SES, White	13,064	446	19	163
		non-Medicaid/non-low SES, non-White	2,557	245	5	40
		Medicaid/low SES, non-White	2,184	209	3	68
Monroe	Type F	all	5,106	500	2	929
Morgan	Type E	non-Medicaid/non-low SES, White	4,883	361	8	397
		non-White OR Medicaid/low SES	3,422	253	2	279
Nicholas	Type E	non-Medicaid/non-low SES, White	4,665	260	2	244
		non-White OR Medicaid/low SES	5,282	294	4	277
Ohio	Type C*	non-Medicaid/non-low SES, White	12,536	346	16	175
		Medicaid/low SES, White	6,575	227	7	92
		non-White	1,666	161	3	233
Pendleton	Type F	all	2,821	528	3	436

County		Substratum				
Name	Type	Description	ALL	B1	SL	B2
Pleasants	Type F	all	2,951	571	3	494
Pocahontas	Type F	all	3,737	514	7	731
Preston	Type E*	non-Medicaid/non-low SES, White	7,353	228	7	299
		non-White OR Medicaid/low SES	5,830	272	6	237
Putnam	Type C*	non-Medicaid/non-low SES, White	16,615	476	18	461
		Medicaid/low SES, White	7,322	262	13	203
		non-White	1,671	167	4	116
Raleigh	Type C*	non-Medicaid/non-low SES, White	16,132	421	25	267
		Medicaid/low SES, White	11,614	378	13	192
		non-White	3,282	299	5	544
Randolph	Type E*	non-Medicaid/non-low SES, White	5,866	261	10	308
		non-White OR Medicaid/low SES	5,662	377	7	297
Ritchie	Type F	all	4,187	573	2	523
Roane	Type F	all	6,532	539	8	479
Summers	Type F	all	5,049	594	6	622
Taylor	Type F	all	7,304	503	8	551
Tucker	Type F	all	2,754	500	4	569
Tyler	Type F	all	3,912	613	6	602
Upshur	Type E*	non-Medicaid/non-low SES, White	5,107	210	9	260
		non-White OR Medicaid/low SES	5,355	330	9	273
Wayne	Type E*	non-Medicaid/non-low SES, White	9,028	350	7	259
		non-White OR Medicaid/low SES	8,884	517	7	255
Webster	Type F	all	3,388	500	8	595
Wetzel	Type F	all	7,048	589	7	523
Wirt	Type F	all	2,235	582	3	522
Wood	Type C*	non-Medicaid/non-low SES, White	21,726	610	30	315
		Medicaid/low SES, White	16,023	562	22	232
		non-White	2,578	253	2	375
Wyoming	Type F	all	4,997	747	8	1,444

County		Substratum				
Name	**Type**	**Description**	**ALL**	**B1**	**SL**	**B2**
TOTALS			809,823	43,500	1,000	43,500
NOTES: Type=Sub-stratification type (*=oversampling eligible); ALL=Total number of addresses in the 2023 MATCH ABS Sampling Frame; B1=Addresses allocated in batch 1; SL=Addresses randomly sampled in soft launch; B2=Addresses sampled for batch 2; PROP=Proportion						
Proportion is defined as the sum of B1, SL, and B2 divided by ALL.						

8 Appendix C – Questionnaire

8.1 Questionnaire Content Changes from 2021 MATCH to 2023 MATCH Survey

General:

- Added text boxes for participants to specify their response to all questions that response options include 'another response'. This was done to determine if any response options need to be added in the next fielding of the survey.

Section 1: Your Health

- Question No.5: 'Have you ever been told by a doctor, nurse, or other healthcare provider that you have any of the following conditions?'
  - Response option of 'Hashimoto's disease' was removed.
  - 'COVID-19' was added.
- Question No.6. 'Do you currently have endocarditis, or have you recovered from it?'
  - Question was added.

Section 2: Healthcare Access

- Question No.9: '(If yes) What kinds of health insurance or healthcare coverage do you have?'
  - Expanded the definition of 'Medigap' as a kind of health insurance or healthcare coverage: 'Medicare Supplement Insurance (Medigap)'
- Question No.10: 'Has a doctor or healthcare provider prescribed any medications for you in the past 12 months?'
  - Added to replace the question: 'I did not have any prescription medication over the past 12 months' to determine if prescription medications were not filled or delayed.'
- Question No.12: 'What were the reasons why you delayed or did not get any medications that a doctor or healthcare provider prescribed for you in the past 12 months?'
  - The responses 'I did not have childcare', 'I lost the prescription', and 'I couldn't remember which pharmacy they sent it to' were removed. 'Because of COVID-19' was edited to 'I was worried about getting COVID-19.'
- Question No.13: 'Was there a time in the past 12 months when you needed medical care? Do not include mental health, dental care, or preventive care/annual screenings.)'
  - Mental health and preventive care/annual screenings were added in what to not include when asking about medical care in the past 12 months.
- Question No.15: '(If no) What were the reasons why you did not get the medical care you needed in the past 12 months?'
  - Removed the response option of 'Because of COVID-19' when asking about reasons for not getting needed medical care.
- Question No.16: 'In the past 12 months, have you had a telehealth visit with a doctor or healthcare provider? Telehealth would include phone, video chat, mobile app, or online patient portals such as MyWVUChart or MyCareCorner.'
  - Telehealth visits were edited to refer to any visits rather than a singular visit (i.e., Were any of these telehealth visits... vs. Was this a telehealth visit...)

- Question No.19: 'In the past 12 months, during any telehealth did a healthcare provider recommend you see a specialist, have lab testing, and/or get a prescription?'
    - The timeframe 'In the past 12 months' was added to question the telehealth provider recommendations.
    - 'Did your provider follow up with you about the recommendation' was removed from this question to be a separate item.

Section 3: Mental Health

- Question No.30: '(If no) Sometimes people do not see a doctor or healthcare provider even when they might need to. Which of the following are reasons why you did not see a doctor or healthcare provider for problems with your mental health, emotions, or nerves in the past 12 months?'
    - The response option: 'I was concerned about being exposed to COVID-19' was edited to 'I was worried about getting COVID-19.'
- Question No.35: '(If no) What are the reasons you are no longer receiving counseling or therapy?'
    - Removed response option of: 'Because of COVID-19.'
- Question No.38: '(If yes) In the past 12 months, what were the reasons you did not get the medicine(s) you had been prescribed to help with your mental health, emotions, or nerves?'
    - Removed response option of: 'Because of COVID-19.'

Section 4: Lifestyle

- Question No.43: 'These next questions are about your housing situation. In the past 12 months, has paying your rent or mortgage gotten easier, stayed the same, or gotten harder?'
    - 'I do not pay rent or a mortgage' was removed as a response option.
- Question No.48: 'When shopping for food, how often do you buy fresh fruits or vegetables that are not canned, frozen, or otherwise processed?'
    - 'I do not shop for food' was added as a response option.
- Question No.49: 'There are many reasons why people don't buy fresh fruits or vegetables every time they shop for food. How about you—what are the reasons why you don't always buy fresh fruits or vegetables?'
    - 'I do not like them' was added as a response option.
- Question No.50: 'In the past 30 days, did you or anyone in your household get free groceries or free meals from the following? Select all that apply.'
    - The response options of 'None of the above' was edited to 'I received free groceries or free meals but not from any of the above' and 'I did not receive free groceries or free meals' was added.
- Question No.53: 'How do you identify yourself? Select all that apply.'
    - Response of 'Other' was edited to 'I use a different term, please specify _____' with an open text box.

Section 5: About You

- Question No.55: 'Do you identify as Appalachian?'
    - Added question.
- Question No.56: 'My Appalachian identity is important to me.'
    - Added question.
- Question No.58: 'What is your current marital status?'
    - The response options of 'Married' was edited to 'Now married' and 'Living with a partner' was removed when asking about current marital status.
- Question No.60: '(If you live with someone else) How many members of your household, including yourself, are 18 years of age or older?'
    - Added question.
- Question No.62: Are you the primary caretaker of a child younger than 18 years of age that is not your own biological or adopted child? Select all that apply.
    - Question was expanded and edited to include any children that are not biological or adopted, including godchild, significant other's child, a sibling, cousin, or other relative's child, or a child who is not a relative.
- Question No.63: 'You indicated you are the primary caretaker of a child younger than 18 years of age that is not your own biological or adopted child. Was this set up through a formal arrangement with Child Protective Services (sometimes called being a "kinship provider")?'
    - Question was expanded to define Child Protective Services
- Question No.67: 'Are you currently..? Select all that apply.' (asking about employment status)
    - 'Student' was added as a response option.
- Question No.68: What is the highest degree or year of school you COMPLETED? If currently enrolled, mark the previous grade or highest degree received.'
    - The question was edited to include 'If currently enrolled, mark the previous grade or highest degree received.'
    - Response options were expanded to be more inclusive.
- Question No.69: 'How much do you disagree or agree with the following statements about in-home healthcare for children or adults?'
    - Added question.

Section 6: Household Finances

- Removed COVID-19 section.
- Replaced with Household Finances section to determine job status and financial security.
- Removed:
    - 'Have you or someone in your household experienced any of the following because of COVID-19?'
    - 'Because of the impact of COVID-19, have you or your household done any of the following?'
    - 'Have you experienced any long-term emotional or mental health effects that you think might be related to you having COVID-19?'
    - 'Have you experienced any long-term emotional or mental health effects that you think might be related to a family member or friend having COVID-19?'

- Question No.72: 'In the past 12 months, have you or your household done any of the following? ['Yes' or 'No': Used up all or most of your savings; cut back spending on food; increased your credit card debt; took money out of retirement, college, or long-term savings accounts; borrowed money from family or friends; pawned or sold possessions]'
    - Added question.
- Question No.73: 'In the past 12 months, has anyone in your household received any of the following public benefits?'
    - Question was moved from the 'About You' section to the last item of the new Household Finances section.

Section 7: Health Behaviors

- Expanded Health Behaviors section renamed from Substance Abuse to Health Behaviors.
- 'In the past 12 months, have you gotten sick or had flu-like symptoms when trying to cut down or stop using any of the following?'
    - Removed question.
- 'Thinking about the last time you used prescription opioids, or 'pills', in any way a doctor did not direct you to use it, what were the reasons you used it the last time?'
    - Removed question.
- 'Have you ever had an overdose that required you to go to the emergency room or receive medical attention right away?'
    - Removed question.
- Question No.74: 'In the past 30 days, on how many days have you had at least one drink of any alcoholic beverage such as beer, wine, a malt beverage, or liquor?'
    - "I have not had any drinks over the past 30 days' was removed as a response option.
    - Response is now an open numerical text box.
- Question No.77: 'Do you now use electronic "vaping" products or other e-cigarettes, including but not limited to JUUL, e-cigars, e-pipes, mods, personal vaporizers, vape pipes, vaping pipes, e-hookahs, and hookah pens?'
    - Added question.
- Question No.79: '(If you smoke cigarettes) Are you thinking about quitting smoking or tobacco use in the next six months?'
    - The phrase 'or tobacco use' was added to the question.
- Question No.85: 'In the past 12 months, have you taken prescription opioids, or 'pills,' while drinking alcohol or within a couple of hours of drinking?'
    - 'In the past 12 months' was added to question.
- Question No.86: 'In the past 12 months, have you taken prescription opioids, or "pills," while using benzodiazepines also called "downers" or "benzies"? Some examples of benzodiazepines are Xanax, Ativan, Klonopin, Valium, etc.'
    - In the past 12 months was added to question
- Question No.89: 'Have you overdosed in the past 12 months?'
    - Added question.

- Question No.90: 'In the past 12 months, have you received Narcan® (naloxone) for an overdose?'
    - Added question.
- Question No.91: 'In the past 12 months, have you had an overdose that required you to go to the emergency room?'
    - Added question.
- Question No.95: '(If you answered no to 94) Which of the following statements explain why you did not get the treatment or counseling you needed for your use of alcohol or drugs?'
    - The response options 'I was not getting better' was removed.
    - 'I did not have time' was removed.

Section 8: Other Topics

- 'Which of the following do you have access to?' [In reference to physical activity]
    - Removed question.
- Question No.106: 'What is your best estimate of you and your family members' total 2022 annual income before taxes and other deductions? Please select the category that applies.'
    - Removed the response option of total annual income below $15,001.
- Question No.107: To be more precise, please enter your best estimate of you and your family members' total 2022 annual income before taxes and other deductions?'
    - Added question with an open numerical text box.
- Question No.109: 'How did you hear about the Mountain State Assessment of Trends in Community Health (MATCH) survey?'
    - Added question.

9 Appendix D – Data Collection

9.1 Data Collection Figures

Figure 9-1: MATCH IVR Path

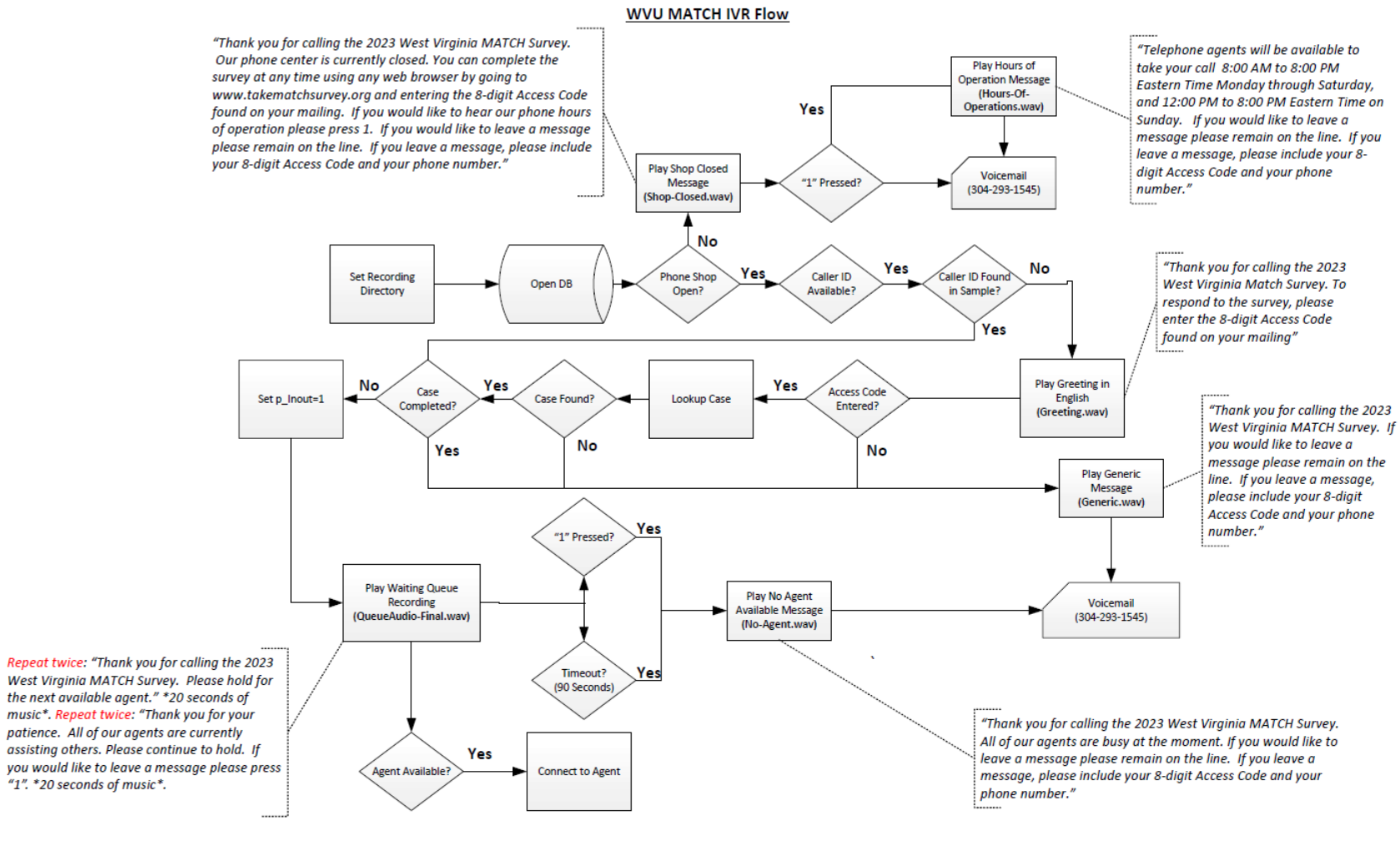


Figure 9-2: Automatic Call Distribution Module Welcome/Introduction Screen

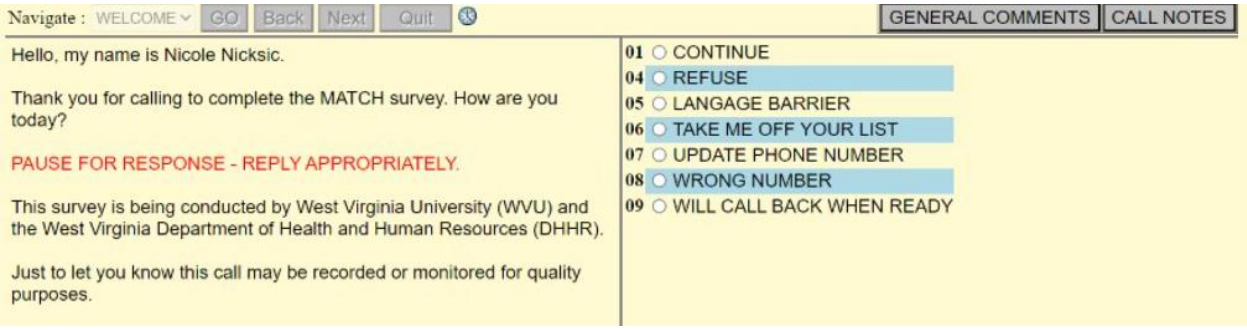


Figure 9-3: Automatic Call Distribution Module Consent Screen

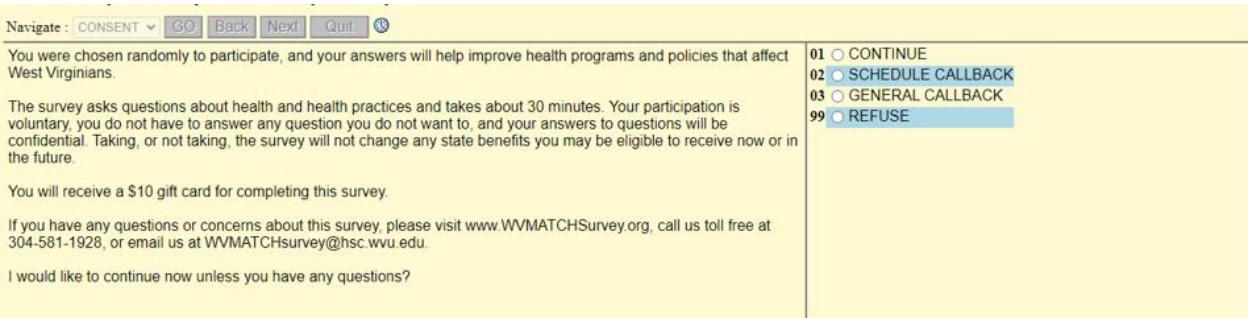


Figure 9-4: Automatic Call Distribution Module Survey Launch Screen

Figure 9-5: Survey Start Screen in Web Browser

Figure 9-6: Example Survey Question for DON'T KNOW and REFUSED

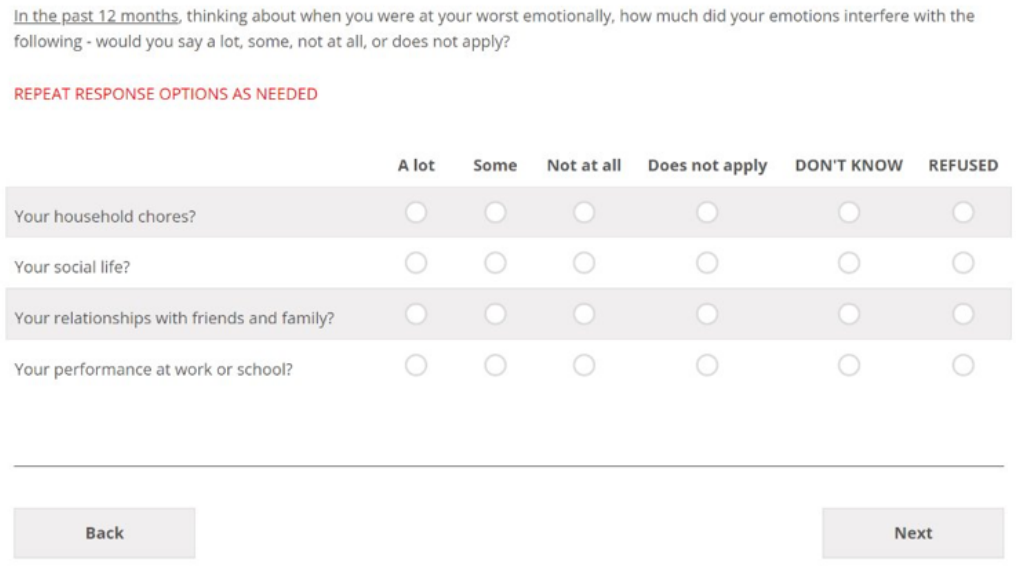
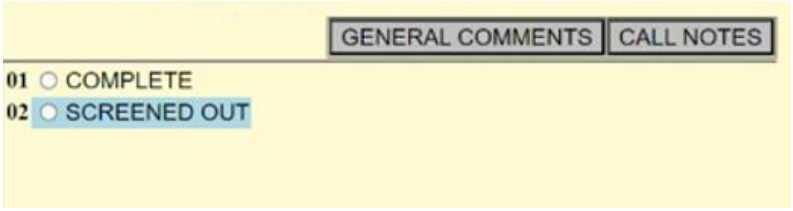
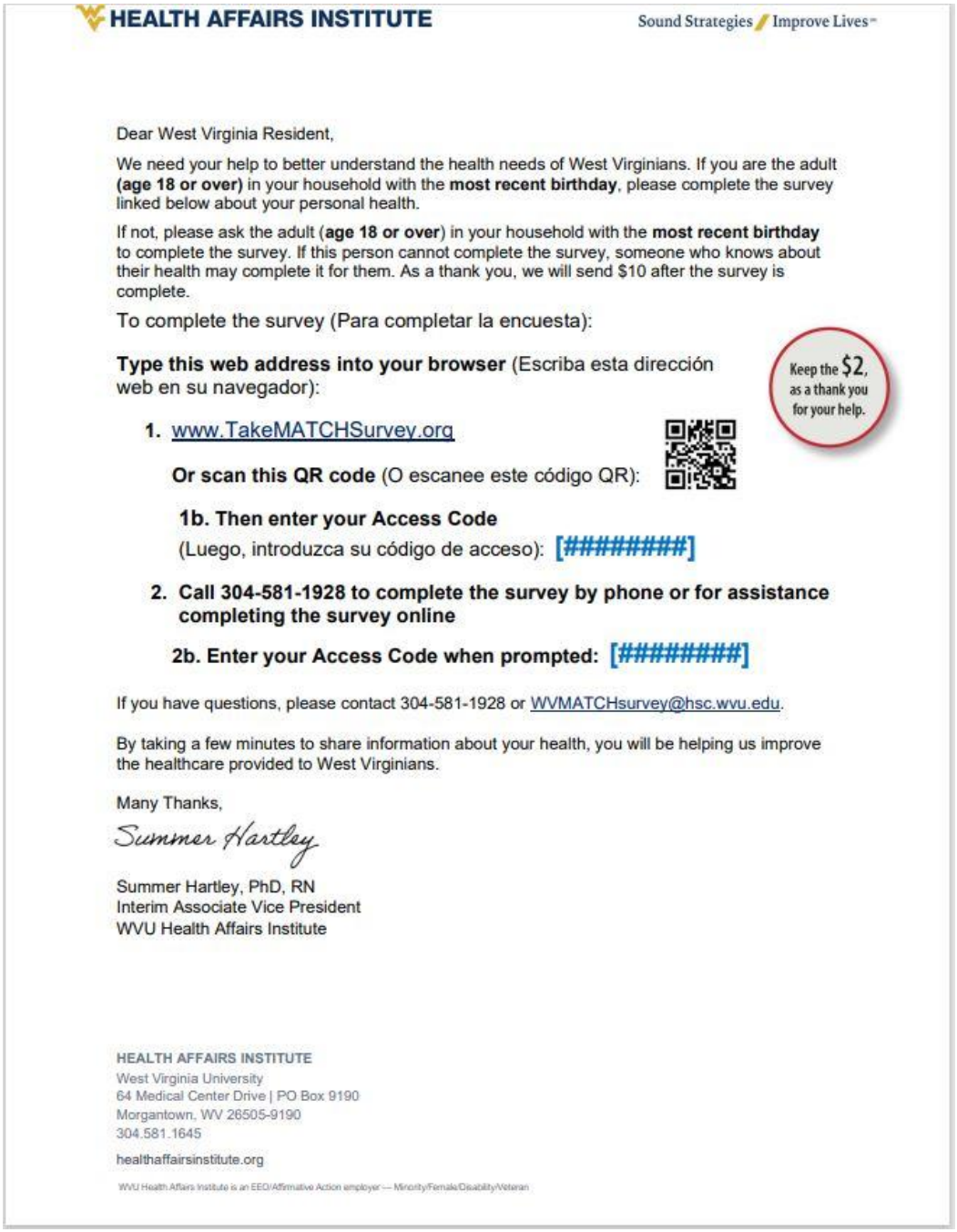


Figure 9-7: Automatic Call Distribution Module Selecting Complete or Screened Out



9.2 Mailing Materials

Figure 9-8: Mailing 1 Invitation Letter



HEALTH AFFAIRS INSTITUTE

Sound Strategies Improve Lives™

Dear West Virginia Resident,

We need your help to better understand the health needs of West Virginians. If you are the adult **(age 18 or over)** in your household with the **most recent birthday**, please complete the survey linked below about your personal health.

If not, please ask the adult (**age 18 or over**) in your household with the **most recent birthday** to complete the survey. If this person cannot complete the survey, someone who knows about their health may complete it for them. As a thank you, we will send $10 after the survey is complete.

To complete the survey (Para completar la encuesta):

Type this web address into your browser (Escriba esta dirección web en su navegador):

Keep the $2, as a thank you for your help.

1. www.TakeMATCHSurvey.org

   Or scan this QR code (O escanee este código QR):

   1b. Then enter your Access Code
   (Luego, introduzca su código de acceso): **[########]**

2. **Call 304-581-1928 to complete the survey by phone or for assistance completing the survey online**

   2b. Enter your Access Code when prompted: [########]

If you have questions, please contact 304-581-1928 or WVMATCHsurvey@hsc.wvu.edu.

By taking a few minutes to share information about your health, you will be helping us improve the healthcare provided to West Virginians.

Many Thanks,

Summer Hartley

Summer Hartley, PhD, RN
Interim Associate Vice President
WVU Health Affairs Institute

HEALTH AFFAIRS INSTITUTE
West Virginia University
64 Medical Center Drive | PO Box 9190
Morgantown, WV 26505-9190
304.581.1645

healthaffairsinstitute.org

WVU Health Affairs Institute is an EEO/Affirmative Action employer — Minority/Female/Disability/Veteran

Figure 9-9:Mailing 2 Reminder Postcard

Figure 9-10: Mailing 3 PAPI Invitation Letter

HEALTH AFFAIRS INSTITUTE

Sound Strategies / Improve Lives

Dear West Virginia Resident,

Our team sent you and other West Virginians a survey about health. If you or someone in your household already completed the survey, thank you very much!

If the adult (**age 18 or over**) in your household with the **most recent birthday** has not yet completed the survey, we wanted to again invite them to take the survey.

It is important we have a completed survey from each selected household to make sure every county in West Virginia is represented in the results. Please ask the adult (**age 18 or over**) in your household with the **most recent birthday** to take the survey soon. If that person cannot complete the survey, then someone who knows about their health may complete it for them.

Please complete the survey in one of two ways:

MAIL:

Complete the enclosed paper survey and mail it back in the postage paid envelope.

ONLINE:

1. Type this web address into your browser: www.TakeMATCHSurvey.org

(Escriba esta dirección web en su navegador: www.TakeMATCHSurvey.org)

Or scan this QR code (O escanee este código QR):

2. Then enter your Access Code (Luego, introduzca su código de acceso):
[########]

For assistance completing online or to answer the survey by phone, call us:* ***304-581-1928*

If you have any questions, please contact 304-581-1928 or WVMATCHsurvey@hsc.wvu.edu.
As a thank you, we will send $10 after receiving the completed survey.

Thank you for your help with this very important issue,

Summer Hartley

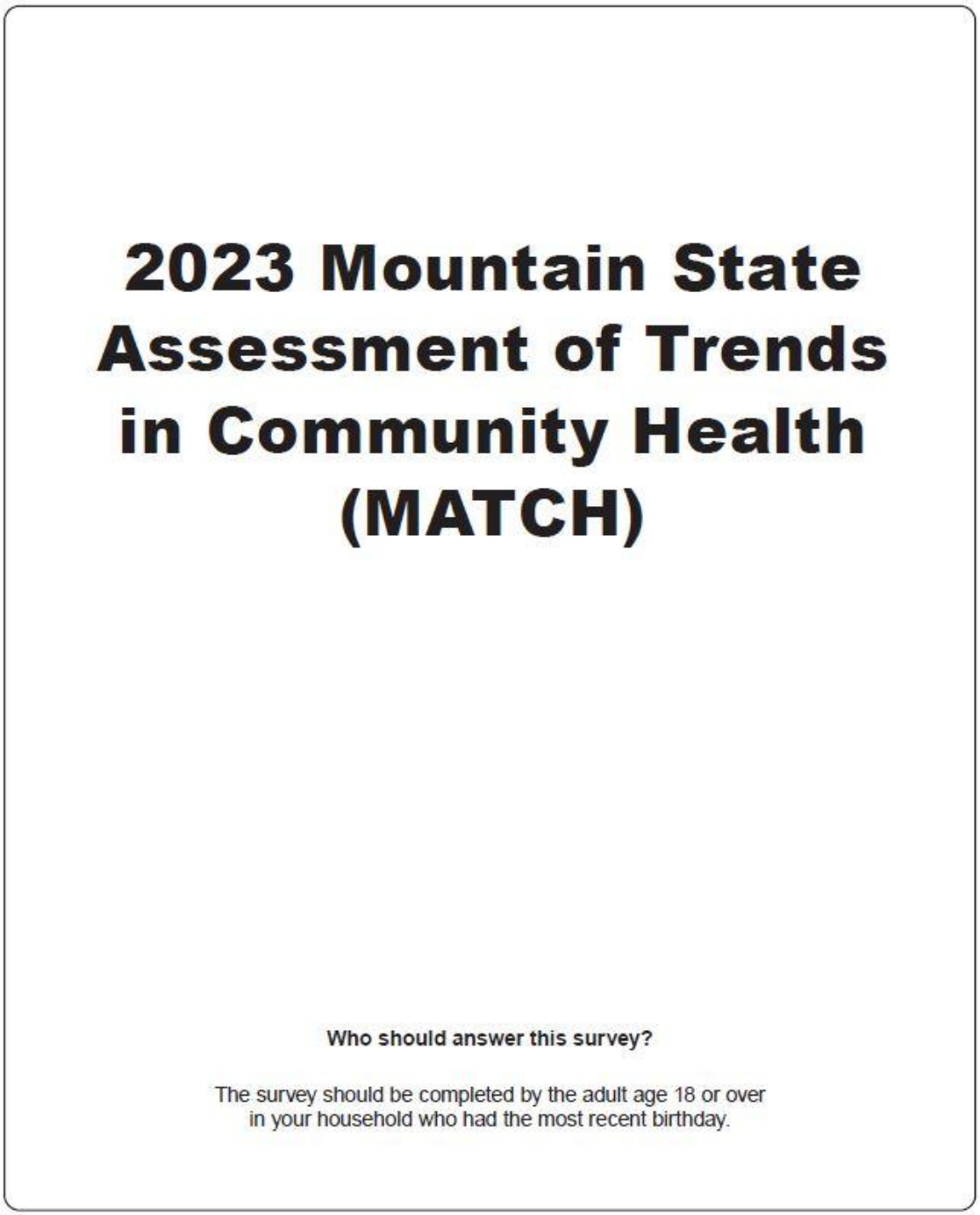
Summer Hartley, PhD, RN
Interim Associate Vice President
WVU Health Affairs Institute

HEALTH AFFAIRS INSTITUTE
West Virginia University
64 Medical Center Drive | PO Box 9190
Morgantown, WV 26505-9190
304.581.1645

healthaffairsinstitute.org

WVU Health Affairs Institute is an EEO/Affirmative Action employer — Minority/Female/Disability/Veteran

Figure 9-11: Mailing 3 PAPI (Cover Page)

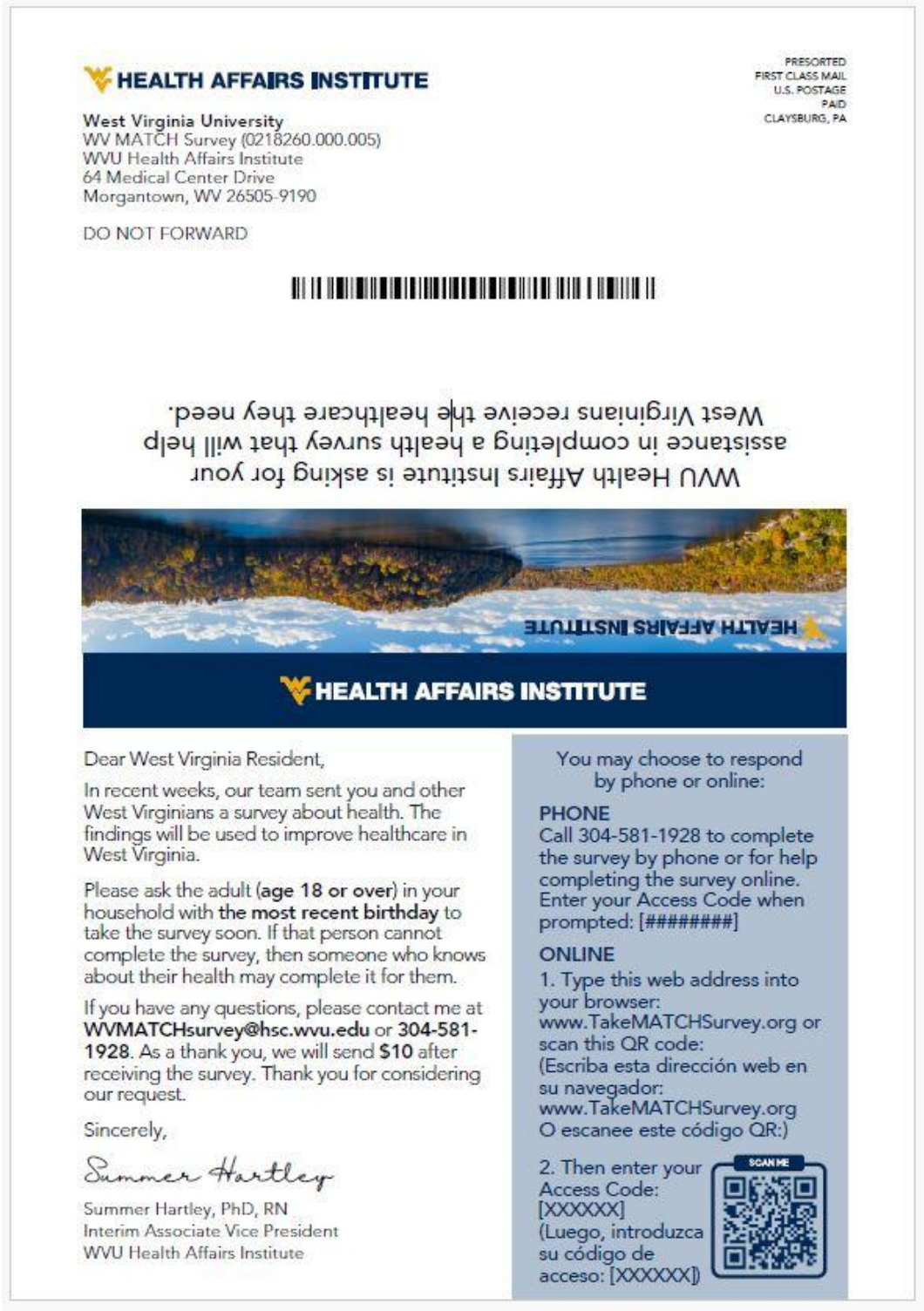


2023 Mountain State Assessment of Trends in Community Health (MATCH)

Who should answer this survey?

The survey should be completed by the adult age 18 or over in your household who had the most recent birthday.

Figure 9-12: Mailing 4 Second Reminder Postcard



HEALTH AFFAIRS INSTITUTE

PRESORTED
FIRST CLASS MAIL
U.S. POSTAGE
PAID
CLAYSBURG, PA

West Virginia University
WV MATCH Survey (0218260.000.005)
WVU Health Affairs Institute
64 Medical Center Drive
Morgantown, WV 26505-9190

DO NOT FORWARD

WVU Health Affairs Institute is asking for your assistance in completing a health survey that will help West Virginians receive the healthcare they need.

HEALTH AFFAIRS INSTITUTE

HEALTH AFFAIRS INSTITUTE

Dear West Virginia Resident,

In recent weeks, our team sent you and other West Virginians a survey about health. The findings will be used to improve healthcare in West Virginia.

Please ask the adult (**age 18 or over**) in your household with **the most recent birthday** to take the survey soon. If that person cannot complete the survey, then someone who knows about their health may complete it for them.

If you have any questions, please contact me at **WVMATCHsurvey@hsc.wvu.edu** or **304-581-1928**. As a thank you, we will send **$10** after receiving the survey. Thank you for considering our request.

Sincerely,

Summer Hartley

Summer Hartley, PhD, RN
Interim Associate Vice President
WVU Health Affairs Institute

You may choose to respond by phone or online:

PHONE
Call 304-581-1928 to complete the survey by phone or for help completing the survey online. Enter your Access Code when prompted: [########]

ONLINE
1. Type this web address into your browser: www.TakeMATCHSurvey.org or scan this QR code:
(Escriba esta dirección web en su navegador: www.TakeMATCHSurvey.org O escanee este código QR:)

2. Then enter your Access Code: [XXXXXX]
(Luego, introduzca su código de acceso: [XXXXXX])

SCAN ME

Figure 9-13: Front of Mailing 5 Targeted Postcard

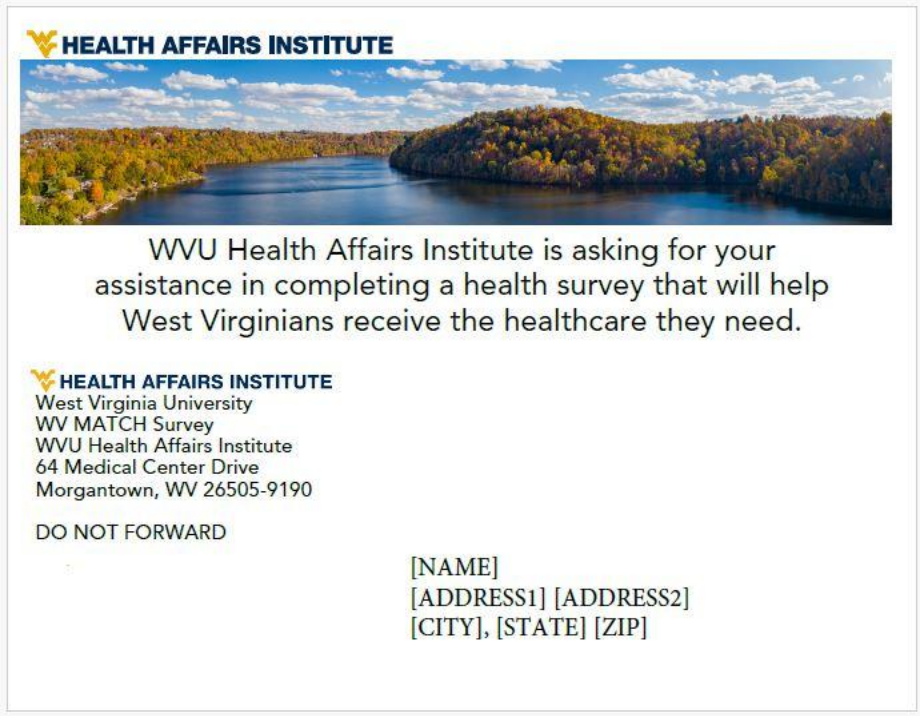
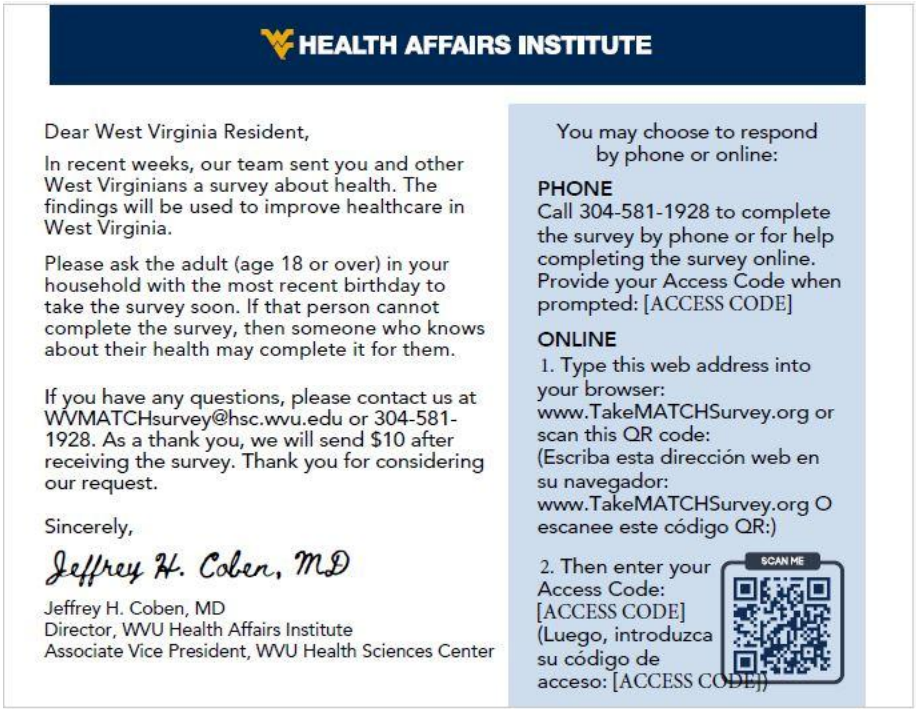


Figure 9-14: Back of Mailing 5 Targeted Postcard



9.3 Community Outreach Efforts by Month

October 2023:

- MATCH survey featured during MetroNews coverage of the 2023 high school and college football seasons.
- MATCH team held booths at the 2023 Appalachian Addiction & Prescription Drug Abuse Conference and the Appalachian Translational Research Network 2023 Annual Health Summit.

November 2023:

- MATCH survey continues to be featured during MetroNews coverage of the 2023 high school and college football seasons.
- MATCH team held a booth at the American Public Health Association Meeting and Expo and presented three posters and two presentations – one in a round table session and one in an oral session.

December 2023:

- MATCH survey continues to be featured during MetroNews broadcasts and through internet advertisements on their website.

January 2024:

- MATCH survey continues to be featured during MetroNews broadcasts and through internet advertisements on their website.
- County factsheets showcasing the 2021 MATCH survey data were posted to the MATCH website.

February 2024:

- MATCH survey continues to be featured during MetroNews broadcasts and through internet advertisements on their website.
- County factsheets showcasing the 2021 MATCH survey data were posted to the MATCH website. Thirty-three of the 55 factsheets were emailed to their corresponding health department. These factsheets also include information on the next MATCH survey fielding to aid in outreach efforts.
- Regional Planning and Development Council's (RPDC's) 1, 4, 6, and 8 were contacted and asked to share the MATCH survey message and digital flyer with their partners in their communities. Outreach examples include sharing on social media platforms, email blasts to members and/or incorporating information into digital monthly newsletter. Collectively, the RPDCs serve: McDowell, Mercer, Monroe, Raleigh, Summers, Wyoming, Cabell, Lincoln, Logan, Mason, Mingo, Wayne, Fayette, Greenbrier, Nicholas, Pocahontas, Webster, Grant, Hampshire, Hardy, Mineral, and

Pendleton counties and Economic Development Associations (EDA's) serving Logan, Lincoln, Mingo, Greenbrier, Grant, and Wyoming counties.

March 2024:

- State organizations were contacted to share the MATCH survey message and digital flyer with their partners in their communities. Organizations include The Governor's Office, State Family Resource Networks, County Commissioners Association, Generation WV, WV Association of Free Clinics, WV Association of Alcohol and Drug Abuse Counselors, WV Nurses Association, WV Association of REALTORS, WV Chamber of Commerce, WV Commerce and Comms Marketing Group, and WVU Extension Downtown Office.
- A presentation of the MATCH survey was given to Welch Memorial Hospital. Pocahontas and Logan Hospitals have been contacted to schedule a date and time for a presentation.

April 2024:

- Presentations of the MATCH survey were given to Pocahontas Memorial Hospital, Southern Highlands Community Health Center, and Rockview Clinic.
- Flyers were sent to Welch Memorial and Pocahontas Memorial Hospitals to be distributed in high tragic areas of their clinics.
- An article about the MATCH survey was featured in the WVU eNews and sent to all WVU staff and faculty.

May 2024:

- One hundred color flyers were mailed to six community health centers located in Oceana, Welch, Mullens, and Princeton, WV.
- The following health organizations agreed to share MATCH content on social media: Community Care of Flatwoods, New River Health, Minnie Hamilton Health Center, Roane County Family Health Care, Gilmer County Family Resource Network, and the WVU Downtown Extension Office.
- Grocery stores serving Braxton, Wyoming, Roane, and Gilmer counties agreed to display MATCH materials in their stores.
- Both Braxton County and Roane County USPS agreed to display MATCH materials in their stores.
- Sutton Public Library of Braxton County and Roane County Public Library agreed to display MATCH materials.

9.4 Response Rates

Table 9-1: Number of Completed Surveys by Frame and Mode and County

	Respondents		
County	**Web**	**Paper**	**Total**
Barbour	191	84	275
Berkeley	581	164	745
Boone	159	68	227
Braxton	134	74	208
Brooke	177	75	252
Cabell	498	151	649
Calhoun	141	79	220
Clay	148	72	220
Doddridge	165	76	241
Fayette	174	69	243
Gilmer	137	68	205
Grant	148	93	241
Greenbrier	183	62	245
Hampshire	186	64	250
Hancock	183	69	252
Hardy	168	67	235
Harrison	330	112	442
Jackson	187	88	275
Jefferson	244	72	316
Kanawha	977	303	1,280
Lewis	165	61	226
Lincoln	173	73	246
Logan	196	55	251
Marion	299	103	402
Marshall	196	73	269
Mason	172	75	247
McDowell	171	74	245
Mercer	296	113	409
Mineral	163	79	242
Mingo	178	65	243
Monongalia	516	99	615
Monroe	172	78	250
Morgan	177	68	245
Nicholas	160	66	226
Ohio	193	66	259
Pendleton	143	81	224
Pleasants	150	74	224

	Respondents		
County	**Web**	**Paper**	**Total**
Pocahontas	170	92	262
Preston	171	74	245
Putnam	233	95	328
Raleigh	323	107	430
Randolph	186	76	262
Ritchie	160	75	235
Roane	151	56	207
Summers	146	79	225
Taylor	174	55	229
Tucker	166	74	240
Tyler	161	68	229
Upshur	173	69	242
Wayne	170	75	245
Webster	149	80	229
Wetzel	159	72	231
Wirt	169	66	235
Wood	400	129	529
Wyoming	173	83	256

Table 9-2: Response Rates

ORDER	LEVEL	NAME	SAMPLED	COMPLETE	PARTIAL	ALLELIGIBLE	UNKNOWNELIGIBLE	NOTELIGIBLE	YIELDRATE	RR2	RR6
54	STATE	West Virginia	88,000	14,866	1,837	76,718	4,628	6,654	0.19	0.21	0.22
54001	COUNTY	Barbour	1,445	248	27	1,234	55	156	0.19	0.21	0.22
54003	COUNTY	Berkeley	4,125	663	82	3,926	84	115	0.18	0.19	0.19
54005	COUNTY	Boone	1,295	201	26	1,147	79	69	0.18	0.19	0.20
54007	COUNTY	Braxton	1,012	177	31	847	43	122	0.21	0.23	0.25
54009	COUNTY	Brooke	1,396	224	28	1,302	23	71	0.18	0.19	0.19
54011	COUNTY	Cabell	3,893	595	54	3,402	151	340	0.17	0.18	0.19
54013	COUNTY	Calhoun	1,107	194	26	971	49	87	0.20	0.22	0.23
54015	COUNTY	Clay	1,328	201	19	1,108	92	128	0.17	0.18	0.20
54017	COUNTY	Doddridge	1,198	213	28	1,009	33	156	0.20	0.23	0.24
54019	COUNTY	Fayette	1,477	221	22	1,249	77	151	0.16	0.18	0.19
54021	COUNTY	Gilmer	1,066	180	25	883	28	155	0.19	0.23	0.23
54023	COUNTY	Grant	1,162	207	34	1,068	55	39	0.21	0.21	0.23
54025	COUNTY	Greenbrier	1,152	216	29	1,022	68	62	0.21	0.22	0.24

54027	COUNTY	Hampshire	1,196	228	22	1,108	47	41	0.21	0.22	0.23
54029	COUNTY	Hancock	1,466	233	19	1,334	27	105	0.17	0.19	0.19
54031	COUNTY	Hardy	990	208	27	940	33	17	0.24	0.24	0.25
54033	COUNTY	Harrison	2,017	406	36	1,796	69	152	0.22	0.24	0.25
54035	COUNTY	Jackson	1,464	241	34	1,320	62	82	0.19	0.20	0.21
54037	COUNTY	Jefferson	1,590	286	30	1,536	14	40	0.20	0.20	0.21
54039	COUNTY	Kanawha	6,621	1126	154	5,742	269	610	0.19	0.21	0.22
54041	COUNTY	Lewis	1,092	196	30	969	52	71	0.21	0.22	0.23
54043	COUNTY	Lincoln	1,557	209	37	1,411	32	114	0.16	0.17	0.17
54045	COUNTY	Logan	2,053	212	39	1,498	279	276	0.12	0.14	0.17
54047	COUNTY	McDowell	2,092	209	36	1,229	480	383	0.12	0.14	0.20
54049	COUNTY	Marion	1,705	367	35	1,545	24	136	0.24	0.26	0.26
54051	COUNTY	Marshall	1,284	245	24	1,211	10	63	0.21	0.22	0.22
54053	COUNTY	Mason	1,344	216	31	1,219	37	88	0.18	0.20	0.20
54055	COUNTY	Mercer	2,250	366	43	1,954	138	158	0.18	0.20	0.21

54057	COUNTY	Mineral	1,071	222	20	999	24	48	0.23	0.24	0.24
54059	COUNTY	Mingo	2,323	199	44	1,810	300	213	0.10	0.12	0.13
54061	COUNTY	Monongalia	2,580	574	41	2,340	44	196	0.24	0.26	0.26
54063	COUNTY	Monroe	1,431	212	38	1,316	70	45	0.17	0.18	0.19
54065	COUNTY	Morgan	1,300	221	24	1,181	54	65	0.19	0.20	0.21
54067	COUNTY	Nicholas	1,081	202	24	866	83	132	0.21	0.24	0.26
54069	COUNTY	Ohio	1,260	229	30	1,134	11	115	0.21	0.23	0.23
54071	COUNTY	Pendleton	967	196	28	862	65	40	0.23	0.24	0.26
54073	COUNTY	Pleasants	1,068	203	21	999	34	35	0.21	0.22	0.22
54075	COUNTY	Pocahontas	1,252	230	32	1,040	63	149	0.21	0.24	0.25
54077	COUNTY	Preston	1,049	222	23	971	22	56	0.23	0.25	0.25
54079	COUNTY	Putnam	1,720	299	29	1,563	83	74	0.19	0.20	0.21
54081	COUNTY	Raleigh	2,144	384	46	1,838	114	192	0.20	0.22	0.23
54083	COUNTY	Randolph	1,260	225	37	1,116	87	57	0.21	0.22	0.23
54085	COUNTY	Ritchie	1,098	212	23	978	50	70	0.21	0.23	0.24

54087	COUNTY	Roane	1,026	180	27	898	36	92	0.20	0.22	0.23
54089	COUNTY	Summers	1,222	192	33	1,025	140	57	0.18	0.19	0.22
54091	COUNTY	Taylor	1,062	199	30	988	28	46	0.22	0.23	0.23
54093	COUNTY	Tucker	1,073	221	19	926	59	88	0.22	0.24	0.26
54095	COUNTY	Tyler	1,221	197	32	1,062	52	107	0.19	0.21	0.22
54097	COUNTY	Upshur	1,091	218	24	1,011	30	50	0.22	0.23	0.24
54099	COUNTY	Wayne	1,395	211	34	1,261	42	92	0.18	0.19	0.19
54101	COUNTY	Webster	1,103	201	28	899	73	131	0.21	0.24	0.25
54103	COUNTY	Wetzel	1,119	207	24	958	77	84	0.21	0.22	0.24
54105	COUNTY	Wirt	1,107	209	26	992	61	54	0.21	0.22	0.24
54107	COUNTY	Wood	2,401	487	42	2,229	73	99	0.22	0.23	0.24
54109	COUNTY	Wyoming	2,199	226	30	1,476	443	280	0.12	0.13	0.17
5400160	STRATUM	BARBOUR, No substrata	1,445	248	27	1,234	55	156	0.19	0.21	0.22
5400311	STRATUM	BERKELEY, NMW	1,326	241	29	1,249	35	42	0.20	0.21	0.22
5400312	STRATUM	BERKELEY, MW	793	128	9	749	18	26	0.17	0.18	0.18

54003 13	STRATUM	BERKELEY, NMNW	620	111	14	603	9	8	0.20	0.20	0.21
54003 14	STRATUM	BERKELEY, MNW	1,386	183	30	1,325	22	39	0.15	0.16	0.16
54005 60	STRATUM	BOONE, No substrata	1,295	201	26	1,147	79	69	0.18	0.19	0.20
54007 60	STRATUM	BRAXTON, No substrata	1,012	177	31	847	43	122	0.21	0.23	0.25
54009 51	STRATUM	BROOKE, NMW	739	128	15	675	15	49	0.19	0.21	0.21
54009 57	STRATUM	BROOKE, (NMNW u MW u MNW)	657	96	13	627	8	22	0.17	0.17	0.17
54011 31	STRATUM	CABELL, NMW	1,357	208	12	1,138	60	159	0.16	0.18	0.19
54011 32	STRATUM	CABELL, MW	1,007	161	18	898	33	76	0.18	0.19	0.20
54011 36	STRATUM	CABELL, (NMNW u MNW)	1,529	226	24	1,366	58	105	0.16	0.18	0.18
54013 60	STRATUM	CALHOUN, No substrata	1,107	194	26	971	49	87	0.20	0.22	0.23
54015 60	STRATUM	CLAY, No substrata	1,328	201	19	1,108	92	128	0.17	0.18	0.20
54017 60	STRATUM	DODDRIDGE, No substrata	1,198	213	28	1,009	33	156	0.20	0.23	0.24

5401951	STRATUM	FAYETTE, NMW	667	102	9	510	44	113	0.17	0.20	0.22
5401957	STRATUM	FAYETTE, (NMNW u MW u MNW)	810	119	13	739	33	38	0.16	0.17	0.18
5402160	STRATUM	GILMER, No substrata	1,066	180	25	883	28	155	0.19	0.23	0.23
5402360	STRATUM	GRANT, No substrata	1,162	207	34	1,068	55	39	0.21	0.21	0.23
5402551	STRATUM	GREENBRIER, NMW	586	112	13	495	42	49	0.21	0.23	0.25
5402557	STRATUM	GREENBRIER, (NMNW u MW u MNW)	566	104	16	527	26	13	0.21	0.22	0.23
5402751	STRATUM	HAMPSHIRE, NMW	663	131	7	599	33	31	0.21	0.22	0.23
5402757	STRATUM	HAMPSHIRE, (NMNW u MW u MNW)	533	97	15	509	14	10	0.21	0.21	0.22
5402951	STRATUM	HANCOCK, NMW	746	123	7	665	18	63	0.17	0.19	0.20
5402957	STRATUM	HANCOCK, (NMNW u MW u MNW)	720	110	12	669	9	42	0.17	0.18	0.18
5403160	STRATUM	HARDY, No substrata	990	208	27	940	33	17	0.24	0.24	0.25

54033 31	STRATUM	HARRISON, NMW	818	170	9	677	43	98	0.22	0.25	0.26
54033 32	STRATUM	HARRISON, MW	596	106	16	554	16	26	0.20	0.21	0.22
54033 36	STRATUM	HARRISON, (NMNW u MNW)	603	130	11	565	10	28	0.23	0.25	0.25
54035 51	STRATUM	JACKSON, NMW	691	117	20	593	43	55	0.20	0.22	0.23
54035 57	STRATUM	JACKSON, (NMNW u MW u MNW)	773	124	14	727	19	27	0.18	0.18	0.19
54037 11	STRATUM	JEFFERSON, NMW	655	143	10	634	8	13	0.23	0.24	0.24
54037 12	STRATUM	JEFFERSON, MW	266	48	5	257	2	7	0.20	0.20	0.21
54037 13	STRATUM	JEFFERSON, NMNW	198	32	2	195	0	3	0.17	0.17	0.17
54037 14	STRATUM	JEFFERSON, MNW	471	63	13	450	4	17	0.16	0.17	0.17
54039 11	STRATUM	KANAWHA, NMW	2,328	429	44	1,917	115	296	0.20	0.23	0.25
54039 12	STRATUM	KANAWHA, MW	1,639	261	44	1,488	53	98	0.19	0.20	0.20
54039 13	STRATUM	KANAWHA, NMNW	596	121	5	528	21	47	0.21	0.23	0.24
54039 14	STRATUM	KANAWHA, MNW	2,058	315	61	1,809	80	169	0.18	0.20	0.21
54041 60	STRATUM	LEWIS, No substrata	1,092	196	30	969	52	71	0.21	0.22	0.23

54043 51	STRATUM	LINCOLN, NMW	683	91	15	587	21	75	0.16	0.17	0.18
54043 57	STRATUM	LINCOLN, (NMNW u MW u MNW)	874	118	22	824	11	39	0.16	0.17	0.17
54045 51	STRATUM	LOGAN, NMW	960	72	17	539	193	228	0.09	0.12	0.17
54045 57	STRATUM	LOGAN, (NMNW u MW u MNW)	1,093	140	22	959	86	48	0.15	0.16	0.17
54047 51	STRATUM	MCDOWELL, NMW	1,180	77	10	445	397	338	0.07	0.10	0.20
54047 57	STRATUM	MCDOWELL, (NMNW u MW u MNW)	912	132	26	784	83	45	0.17	0.18	0.20
54049 31	STRATUM	MARION, NMW	677	154	13	596	14	67	0.25	0.27	0.28
54049 32	STRATUM	MARION, MW	484	93	9	447	5	32	0.21	0.23	0.23
54049 36	STRATUM	MARION, (NMNW u MNW)	544	120	13	502	5	37	0.24	0.26	0.26
54051 51	STRATUM	MARSHALL, NMW	691	140	12	648	4	39	0.22	0.23	0.23
54051 57	STRATUM	MARSHALL, (NMNW u MW u MNW)	593	105	12	563	6	24	0.20	0.21	0.21

54053 51	STRATUM	MASON, NMW	610	94	13	536	17	57	0.18	0.19	0.20
54053 57	STRATUM	MASON, (NMNW u MW u MNW)	734	122	18	683	20	31	0.19	0.20	0.20
54055 31	STRATUM	MERCER, NMW	658	102	12	531	67	60	0.17	0.19	0.21
54055 32	STRATUM	MERCER, MW	702	125	15	642	29	31	0.20	0.21	0.22
54055 36	STRATUM	MERCER, (NMNW u MNW)	890	139	16	781	42	67	0.17	0.19	0.20
54057 51	STRATUM	MINERAL, NMW	567	124	5	517	18	32	0.23	0.24	0.25
54057 57	STRATUM	MINERAL, (NMNW u MW u MNW)	504	98	15	482	6	16	0.22	0.23	0.23
54059 51	STRATUM	MINGO, NMW	1,146	83	19	731	249	166	0.09	0.10	0.14
54059 57	STRATUM	MINGO, (NMNW u MW u MNW)	1,177	116	25	1,079	51	47	0.12	0.12	0.13
54061 11	STRATUM	MONONGALIA, NMW	13,82	327	24	1,226	26	130	0.25	0.28	0.29
54061 12	STRATUM	MONONGALIA, MW	628	123	9	585	8	35	0.21	0.22	0.23
54061 13	STRATUM	MONONGALIA, NMNW	290	75	4	277	1	12	0.27	0.28	0.29

54061 14	STRATUM	MONONGALIA, MNW	280	49	4	252	9	19	0.19	0.20	0.21
54063 60	STRATUM	MONROE, No substrata	1,431	212	38	1,316	70	45	0.17	0.18	0.19
54065 51	STRATUM	MORGAN, NMW	766	137	16	671	43	52	0.20	0.21	0.23
54065 57	STRATUM	MORGAN, (NMNW u MW u MNW)	534	84	8	510	11	13	0.17	0.18	0.18
54067 51	STRATUM	NICHOLAS, NMW	506	92	10	362	44	100	0.20	0.25	0.28
54067 57	STRATUM	NICHOLAS, (NMNW u MW u MNW)	575	110	14	504	39	32	0.22	0.23	0.25
54069 31	STRATUM	OHIO, NMW	537	110	10	480	6	51	0.22	0.25	0.25
54069 32	STRATUM	OHIO, MW	326	54	11	290	3	33	0.20	0.22	0.22
54069 36	STRATUM	OHIO, (NMNW u MNW)	397	65	9	364	2	31	0.19	0.20	0.20
54071 60	STRATUM	PENDLETON, No substrata	967	196	28	862	65	40	0.23	0.24	0.26
54073 60	STRATUM	PLEASANTS, No substrata	1,068	203	21	999	34	35	0.21	0.22	0.22

54075 60	STRATUM	POCAHONTAS, No substrata	1,252	230	32	1,040	63	149	0.21	0.24	0.25
54077 51	STRATUM	PRESTON, NMW	534	118	8	484	13	37	0.24	0.25	0.26
54077 57	STRATUM	PRESTON, (NMNW u MW u MNW)	515	104	15	487	9	19	0.23	0.24	0.24
54079 31	STRATUM	PUTNAM, NMW	955	171	10	841	67	47	0.19	0.20	0.22
54079 32	STRATUM	PUTNAM, MW	478	76	11	444	13	21	0.18	0.19	0.20
54079 36	STRATUM	PUTNAM, (NMNW u MNW)	287	52	8	278	3	6	0.21	0.21	0.22
54081 31	STRATUM	RALEIGH, NMW	713	121	12	536	63	114	0.19	0.22	0.25
54081 32	STRATUM	RALEIGH, MW	583	123	15	547	16	20	0.24	0.25	0.25
54081 36	STRATUM	RALEIGH, (NMNW u MNW)	848	140	19	755	35	58	0.19	0.20	0.21
54083 51	STRATUM	RANDOLPH, NMW	579	108	19	483	66	30	0.22	0.23	0.26
54083 57	STRATUM	RANDOLPH, (NMNW u MW u MNW)	681	117	18	633	21	27	0.20	0.21	0.21
54085 60	STRATUM	RITCHIE, No substrata	1,098	212	23	978	50	70	0.21	0.23	0.24

54087 60	STRATUM	ROANE, No substrata	1,026	180	27	898	36	92	0.20	0.22	0.23
54089 60	STRATUM	SUMMERS, No substrata	1,222	192	33	1,025	140	57	0.18	0.19	0.22
54091 60	STRATUM	TAYLOR, No substrata	1,062	199	30	988	28	46	0.22	0.23	0.23
54093 60	STRATUM	TUCKER, No substrata	1,073	221	19	926	59	88	0.22	0.24	0.26
54095 60	STRATUM	TYLER, No substrata	1,221	197	32	1,062	52	107	0.19	0.21	0.22
54097 51	STRATUM	UPSHUR, NMW	479	102	13	426	18	35	0.24	0.26	0.27
54097 57	STRATUM	UPSHUR, (NMNW u MW u MNW)	612	116	11	585	12	15	0.21	0.21	0.22
54099 51	STRATUM	WAYNE, NMW	616	98	15	521	27	68	0.18	0.21	0.22
54099 57	STRATUM	WAYNE, (NMNW u MW u MNW)	779	113	19	740	15	24	0.17	0.17	0.18
54101 60	STRATUM	WEBSTER, No substrata	1,103	201	28	899	73	131	0.21	0.24	0.25
54103 60	STRATUM	WETZEL, No substrata	1,119	207	24	958	77	84	0.21	0.22	0.24
54105 60	STRATUM	WIRT, No substrata	1,107	209	26	992	61	54	0.21	0.22	0.24

5410731	STRATUM	WOOD, NMW	955	199	13	877	38	40	0.22	0.23	0.24
5410732	STRATUM	WOOD, MW	816	159	18	766	22	28	0.22	0.22	0.23
5410736	STRATUM	WOOD, (NMNW u MNW)	630	129	11	586	13	31	0.22	0.23	0.24
5410960	STRATUM	WYOMING, No substrata	2,199	226	30	1,476	443	280	0.12	0.13	0.17

10 Appendix E: Data Processing

10.1 Data Processing Tables

This section provides additional detailed tables for data processing methods.

Table 10-1: Imputation Variable Construction

Construction Variable	Construction
All variables for imputation	Recode missing value of '999' to '.' and keep non-missing value as it is.
RANDORDER	RANDORDER = UNIFORM("RANDOMIZATION SEED")
CASEIDN	CASEIDN = input(CASEID, best32.)
IMPUTEWT	IMPUTEWT = 1
DMI_AGE15YR	= 1, if DMI_AGE in (18-34) = 2, else if DMI_AGE in (35-49) = 3, else if DMI_AGE in (50-64) = 4, else if DMI_AGE in (65-120)
DMI_RACEG3	= 3, if DMI_RACEG5 in (3 4 5) = DMI_RACEG5, otherwise
DMI_RACEHG3	= 3, if DMI_HISPANC = 1 OR DMI_RACEG3 = 3 = DMI_RACEG3, if DMI_HISPANC = 2
DMI_MRTLG3	= 1, if DMI_MRTLSTA = 1 = 2, if DMI_MRTLSTA in (2, 3, 4) = 3, if DMI_MRTLSTA = 5 = DMI_MRTLSTA, otherwise
DMI_EDUCG3	= 1, if DMI_MAXEDUC in (1, 2, 3, 4) = 2, if DMI_MAXEDUC in (5, 6, 7) = 3, if DMI_MAXEDUC in (8, 9, 10, 11, 12)
TMP_LIVEWTH_ALNE	= DMS_LIVEWTH_ALNE, IF DMS_LIVEWTH_ALNE ^= 999 = 2, IF DMI_LIVEWTH_SPOU = 1 OR (2 <= DMD_RESADLT < 100) OR (1 <= DMD_RESCHLD < 100) = 999, OTHERWISE

Table 10-2: Step by Step Imputation Processing Streamline

Socio demographics	Source Variable for Imputation	Out Imputation Variable	Randomization Seed	Imputation Seed	Imputation Cell Variables	Sort Variables
Age	DMS_AGE	DMI_AGE	26185413	38900579	MODE REGIONBBH	GEO_IDBG RANDORDER
Sex	DMS_BRTHSEX	DMI_BRTHSEX	41760704	42890523	MODE REGIONBBH DMI_AGE15YR	GEO_IDBG RANDORDER
Hispanic	DMS_HISPANC	DMI_HISPANC	26825449	60064926	REGIONBBH DMI_AGE15YR DMI_BRTHSEX	MODE GEO_IDBG RANDORDER
Race	DMC_RACEG5	DMI_RACEG5	33674314	43375161	REGIONBBH DMI_HISPANC DMI_AGE15YR	MODE DMI_BRTHSEX GEO_IDBG RANDORDER
Marital status	DMS_MRTLSTA	DMI_MRTLSTA	50771452	48171440	DMI_AGE15YR DMI_BRTHSEX DMI_RACEG3	MODE GEO_IDBG RANDORDER
Education	DMS_MAXEDUC	DMI_MAXEDUC	41264403	59301971	DMI_AGE15YR DMI_BRTHSEX DMI_RACEHG3 DMI_MRTLG3 REGIONBBH	MODE GEO_IDBG RANDORDER
Live with spouse	DMS_LIVEWTH_SPOU	DMI_LIVEWTH_SPOU	38211265	24924122	DMI_MRTLG3 DMI_AGE15YR DMI_BRTHSEX	DMI_RACEHG3 MODE GEO_IDBG RANDORDER
Live alone	TMP_LIVEWTH_ALNE	DMI_LIVEWTH_ALNE	9240806	15275165	DMI_AGE15YR DMI_BRTHSEX DMI_EDUCG3	MODE GEO_IDBG RANDORDER

Table 10-3: County Level Unequal Weighting Effect and Kish Effective Sample Size

COUNTY	N	UWE	KESS
Overall	16703	2.02	8,270.47
Barbour	275	1.95	140.93
Berkeley	745	1.78	418.18
Boone	227	1.48	153.85
Braxton	208	2.32	89.54
Brooke	252	1.59	158.65
Cabell	649	1.77	365.8
Calhoun	220	1.68	131.18
Clay	220	1.79	123.22
Doddridge	241	1.76	136.63
Fayette	243	1.56	155.73
Gilmer	205	2.05	99.85
Grant	241	1.78	135.35
Greenbrier	245	1.47	167.06
Hampshire	250	1.6	155.97
Hancock	252	1.4	179.61
Hardy	235	1.66	141.86
Harrison	442	1.74	253.71
Jackson	275	1.58	174.05
Jefferson	316	1.59	198.3
Kanawha	1280	1.76	728.78
Lewis	226	1.75	129.09
Lincoln	246	1.52	161.92
Logan	251	1.55	161.68
Marion	402	1.73	232.07
Marshall	269	1.76	152.67
Mason	247	1.43	172.68
McDowell	245	1.64	149.48
Mercer	409	1.89	215.99
Mineral	242	1.65	146.96
Mingo	243	1.61	150.96
Monongalia	615	1.54	400.36
Monroe	250	1.72	145.29
Morgan	245	1.85	132.31
Nicholas	226	1.73	130.64
Ohio	259	1.62	159.67
Pendleton	224	1.68	133.47

COUNTY	N	UWE	KESS
Pleasants	224	1.61	139.14
Pocahontas	262	1.6	163.82
Preston	245	1.6	152.91
Putnam	328	1.5	218.3
Raleigh	430	1.69	254.65
Randolph	262	1.68	156.36
Ritchie	235	1.57	150.07
Roane	207	2.17	95.31
Summers	225	1.6	140.91
Taylor	229	1.92	119.13
Tucker	240	1.67	144.02
Tyler	229	1.73	132.49
Upshur	242	1.73	139.7
Wayne	245	1.56	156.83
Webster	229	1.85	124.08
Wetzel	231	1.6	144.33
Wirt	235	1.88	124.94
Wood	529	1.59	333.39
Wyoming	256	1.68	152.71

Table 10-4: List of Geographic Levels Supported

Geographic Level	Example GEO_ID Format	Notes
Block Group (BG)	1500000US240338060012	Used for most fine grain variables
Census Tract (CT)	1400000US24033806001	Roll-up level
County (CY)	0500000US24033	Used for substitution/imputing

Table 10-5: Data Sources and Fields Used in Constructing ACS Geographic Tables

Data Source	Extracted Field	Role in Calculation
ACS B09021	Total household population (CLASS = "TOTAL")	Used for counts at subcounty levels; used as denominator in some tables (e.g., poverty rate tables)
Characteristic ACS Table (e.g., B17021, B25003)	Numerator counts and matching denominators (e.g., B17021_002E and B17021_001E)	Provides characteristic counts and total counts for proportion calculation
GEO_ID	Geographic identifier	Links counts to block group, tract, and county levels

Table 10-6: Geographic 2022 5-year ACS Tables

TOPIC	Universe and Level	ACS Table Name	Breakdown Used for 2023 MATCH
Poverty status	Poverty status determined - CBG	B17021	Poverty rate < 0.0677, 0.0677 ≤ poverty rate ≤ 0.2346, poverty rate > 0.2346
Housing	Housing units - CBG	B25002	Housing occupied <= 0.7859, 0.7859 < Housing occupied < 0.9357, Housing occupied >= 0.9357 Housing vacant <= 0.0643, 0.0643 < Housing vacant < 0.2141, Housing vacant >= 0.2141
Tenure (owner vs renter occupied)	Occupied housing units - CBG	B25003	Occupied by owner <= 0.6674, 0.6674 < Occupied by owner < 0.8894, Occupied by owner >= 0.8894 Occupied by renter <= 0.1106, 0.1106 < Occupied by renter < 0.3326, Occupied by renter >= 0.3326
Building	Occupied housing units - CBG	B25032	I Unit in housing <= 0.8447, 0.8447 < 1 Unit in housing < 1.00, 1 Unit in housing > 1.00

TOPIC	Universe and Level	ACS Table Name	Breakdown Used for 2023 MATCH
			2-4 Units in housing <= 0.00, 0.00 < 2-4 Units in housing < 0.0545, 2-4 Units in housing > 0.0545 5+ Units in housing <= 0.00, 0.00 < 5+ Units in housing < 0.0832, 5+ Units in housing > 0.0832
Types of insurance coverage by age	Civilian non-institutionalized - CBG	B27010	With Medicaid aged 19-34 <= 0.0677, 0.0677< With Medicaid aged 19-34 < 0.3864, With Medicaid aged 19-34 > 0.3864 With Medicaid aged 35-64 <= 0.0732, 0.0732 < With Medicaid aged 35-64 < 0.2795, With Medicaid aged 35-64 > 0.2795 With Medicaid aged 65+ <= 0.0000, 0.1000 < With Medicaid aged 65+ < 0.000, With Medicaid aged 65+ > 0.1000 With Medicaid aged 19-64 <= 0.0983, 0.0983 < With Medicaid aged 19-64 < 0.2926, With Medicaid aged 19-64 > 0.2926 With Medicaid aged 19+ <= 0.0818, 0.0818 < With Medicaid aged 19+ < 0.2379, With Medicaid aged 19+ > 0.2379 With no insurance aged 19-34 <= 0.0000, 0.0000 < With no insurance aged 19-34 < 0.1835, With no insurance aged 19-34 > 0.1835 With no insurance aged 35-64 <= 0.0194, 0.0194 < With no insurance aged 35-64 < 0.1196, With no insurance aged 35-64 > 0.1196 With no insurance aged 65+ <= 0.00, 0.00 < With no insurance aged 65+ < 0.0832, With no insurance aged 65+ > 0.0832

TOPIC	Universe and Level	ACS Table Name	Breakdown Used for 2023 MATCH
			With no insurance aged 19-64 <= 0.0383, 0.0383 < With no insurance aged 19-64 < 0.1363, With no insurance aged 19-64 > 0.1363 With no insurance aged 19+ <= 0.0285, 0.0285 < With no insurance aged 19+ < 0.1014, With no insurance aged 19+ > 0.1014
Internet access	Households - CBG	B28002	No internet access < 0.0578, 0.0578 ≤ no internet access ≤ 0.1976, no internet access > 0.1976
Internet subscriptions	Households - CBG	B28003	Dial up usage < 0.00, 0.00 ≤ Dial up usage ≤ 0.00, Dial up usage > 0.00 Broadband usage <= 0.7414, 0.7414 < Broadband usage < 0.9081, Broadband usage >= 0.9081 Computer, no internet <= 0.0175, 0.0175 < Computer, no internet < 0.1062, Computer, no internet >= 0.1062 No computer <= 0.0397, 0.0397 < No computer < 0.1583, No computer >= 0.1583

11 References

1. Marketing Systems Group. Accessed 07/29/2024, 2024.
2. NORC. Accessed 07/29/2024, 2024.
3. Services USDoHaH. Poverty Guidelines. Office of the Assistant Secretary for Planning and Evaluation. 2024.
4. A Grammar of Data Manipulation - dplyr. 08/07/2024, https://dplyr.tidyverse.org/
5. Final Dispositions of Case Codes and Outcome Rates for Surveys. *American Association for Public Opinion Research*. 2023; (10th Edition) doi:https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf
6. *SAS software*. SAS Institute Inc.; 2023.
7. *R: A Language and Environment for Statistical Computing*. Version 4.5.1. R Foundation for Statistical Computing; 2025. https://www.R-project.org
8. *Stata Statistical Software: Release 18*. StataCorp LLC; 2023.
9. *IBM SPSS Statistics for Windows*. Version 31.0. IBM Corp.; 2025.
10. *survey: analysis of complex survey samples*. CRAN; 2024. https://cran.r-project.org/package=survey