2022 Health Center Patient Survey Data File User's Manual

Health Resources and Services Administration
Bureau of Primary Health Care
Parklawn Building, 5600 Fishers Lane
Rockville, MD 20857

HRSA
Health Resources & Services Administration

Contents

Section **Page**

List of Tables

1. Introduction

This manual provides documentation for users of the 2022 Health Center Patient Survey (HCPS) Public Use File (PUF). Information about the study design, survey instruments, data collection methods and processes, weighting procedures, and instructions on how to use the data is presented in this manual. This manual will also familiarize the user with the HCPS and provide information necessary for the appropriate use of the data. This chapter contains information on the purpose and significance of the HCPS and the confidentiality of the data.

1.1 Purpose and Significance of the Patient Survey

The purpose of the 2022 HCPS was to obtain nationally representative data about the patients of Section 330–funded health centers. The survey is unique in its effort to capture person-level data from patients of all types of Health Center Program awardees. With the current survey, HRSA aimed to

- Gather data about the patients of the Community Health Center (CHC), Migrant Health Center (MHC), Health Care for the Homeless (HCH), and Public Housing Primary Care (PHPC) programs and the services they obtain.
- Enable comparisons of care received by health center patients with care received by the general population, as measured by the National Health Interview Survey (NHIS) and other national surveys.
- Gather information that will assist policymakers and HRSA staff to
  - assess how well HRSA-supported health care sites are currently able to meet health care needs;
  - identify areas for improvement and guide planning decisions; and
  - complement data that are not routinely collected from other HRSA data sources.

Although the data will be used for the items previously mentioned, it is important to note that the data can only be used for research purposes and cannot be used to scrutinize individual awardee performance. Awardee and site participation were secured under this premise.

The survey builds on the successes of the 1995 CHC User/Visit Survey, the 2002 CHC and National Health Service Corps Site User/Visit Survey, the 2003 HCH User/Visit Survey, the 2009 Primary Health Care Patient Survey, and the 2014 HCPS. However, the current survey included an oversample of patients who identified themselves as Asian, American Indian/Alaska Native (AI/AN), and Native Hawaiian/Pacific Islander (NH/PI), aged 65 or older.

The data collected through the 2022 HCPS are needed to understand the health care needs of the populations served, to assess how well HRSA programs are able to meet those needs, and to guide planning decisions so that programs might be better able to do so. HRSA does not routinely collect this type of information from Section 330–funded sites, and these data are not available from the uniform data system (UDS) maintained by HRSA or any other source. Although a trend analysis between years is not presented and not encouraged, the data editing described in Section 5 was implemented to maintain continuity between survey years to the extent possible for variables that are included in both current and previous survey iterations.

The data cannot and should not be used to scrutinize individual awardee performance. Awardee and site participation were secured under this premise. Additionally, users of the HCPS PUF should not attempt to identify any individual respondent. The HCPS PUF is not suitable for analyzing respondent geography; rare medical conditions; or sensitive topics such as HIV, substance use, and mental health. These sensitive

topics are present in the full survey, but the PUF is a subset of the full survey, and not all survey items on these topics are available in the PUF.

RTI International conducted the 2022 HCPS. RTI is an independent nonprofit institute that provides research, development, and technical services to government and commercial clients worldwide.

1.2 Overview of the User's Manual

This Data File User's Manual provides the information necessary for most analytic purposes.

Information about the sample design is found in **Chapter 2**, while **Chapter 3** contains information about the data collection instruments. Data collection methods and processes are described in **Chapter 4**. Data editing and coding are discussed in **Chapter 5**. **Chapter 6** describes the weighting procedures, and **Chapter 7** explains how to use the PUF. The next section in **Chapter 1** provides a discussion of the methods used to create the PUF from the restricted-use file.

1.3 Confidentiality of Data

To protect the privacy of respondents, all variables that could be used to identify individuals have been treated in the PUF. Previously published estimates may not be exactly reproducible from the variables in the PUF because of the disclosure protection procedures that were implemented.

1.3.1 Data Disclosure Protection

Disclosure arises when respondents in the survey can be identified and correctly linked to individuals in the population. To protect data confidentiality for the 2022 HCPS, statistical disclosure avoidance procedures have been applied to the data to minimize disclosure risk of survey respondents from being identified while still maintaining analytic quality. Disclosure avoidance techniques include standard data deletion (dropping variables), data coarsening such as top or bottom recoding, variable recategorization, local suppression to reduce sample uniques (i.e., a single respondent in a cell with respect to one or more identifying variables) and probabilistic perturbation via random swapping so that the intruder is uncertain if the record is his or her true target. All direct identifiers including name, address, and phone number have been deleted from the file, and all geographic identifiers have been removed. These procedures ensure that the confidentiality of survey respondents is adequately protected.

In addition to controlling for disclosure risk, data quality was monitored during the disclosure treatment process by running multiple random swapping scenarios so that a best run with highest quality was selected. Data utility measures used in the assessment of data quality included estimates and their standard errors and correlations for certain key outcomes in the aspects of demographics, insurance coverage, health conditions, substance use, cancer screening, care for chronic diseases, and satisfaction with care. Regression models were also fit to assess multivariate data quality, before and after swapping, for chronic diseases such as hypertension and diabetes and for cancer screening outcomes such as Pap smear, mammogram, and colonoscopy and were regressed on sociodemographic characteristics.

Information loss was assessed at the global level[1] to make sure maximum data quality is preserved after treatment so that sound statistical inference can be drawn using the PUF.

[1] Global utility measures include Mean Relative Root Mean Square Error, Hellinger's distance, mean absolute relative difference of Cramer's V for measuring change of pair-wise association, and mean absolute relative bias for regression coefficients for all models (Dohrmann et al., 2009).

2. Study Sample Design

The 2022 HCPS applied a three-stage sampling design to reflect a nesting structure. The first stage sampling units were awardees; the second stage sampling units were eligible health center sites within awardees; and the third stage sampling units were eligible patients who had at least one visit in the past 12 months to an eligible health center site. A total of 103 unique awardees and 318 health center sites were recruited, and 4,414 patient interviews were completed. Among them, 2,915 patient interviews were completed for CHC, 473 for MHC, 597 for HCH, and 429 for PHPC. **Table 2-1** summarizes the samples at each of the three sampling stages. Awardees that receive funding from multiple programs and sites that were selected for multiple patient types (CHC, MHC, HCH, and PHPC) are included multiple times in **Table 2-1** under each of the applicable funding programs. For awardees with multiple funding programs, an independent site and patient sample was selected from each funding program; therefore, recruiting 103 awardees was equivalent to selecting a sample from 181 awardees. Of those, there were 100 for CHC, 28 for MHC, 32 for HCH, and 21 for PHPC. Data were collected from 102 of the 103 recruited awardees.

Table 2-1. Three-Stage Sampling Summary for the HCPS

Funding Program	First Stage		Second Stage		Third Stage		
	Number of				Patients Who Approached FIs	Eligible (Selected) Patients	Completed Patient Interviews
	Awardees		Sites				
	Recruited[a]	Participating[b]	Recruited[c]	Participating[d]			
CHC	100	99	248	247	—	3,440	2,915
MHC	28	27	65	64	—	604	473
HCH	32	32	71	70	—	658	597
PHPC	21	21	43	43	—	492	429
Total[e]	103	102	318	316	10,378	5,194	4,414

NOTE: CHC = Community Health Center Program; FI = field interviewer; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.

[a] Awardees that were successfully recruited.
[b] Awardees that had at least one completed patient interview.
[c] Sites that were successfully recruited.
[d] Sites that had at least one completed patient interview.
[e] Totals do not equal sum of counts by funding program because some awardees had multiple funding programs at the first stage and some sites were selected for multiple patient types at the second stage.

Because of the impact of COVID-19 on data collection, the number of recruited awardees and completed interviews fell short of the targeted numbers as shown in **Section 2.2**.

2.1 Target Population

The 2022 HCPS included people who resided in the 50 states and the District of Columbia and met the definition of a health center patient used in HRSA's UDS; that is, people receiving face-to-face services from a CHC, MHC, HCH, or PHPC awardee and from a clinical staff member who exercises independent judgment in the provision of services.[2] Clients of awardees within U.S. territories and possessions were excluded from the 2022 HCPS.

Because many of the questions in the survey ask about services received in the past year, people who received services in person or via telehealth through one of these awardees at least once in the 12 months

[2] To meet the criterion for "independent judgment," the provider must be acting on their own when serving the patient and not assisting another provider.

prior to the current visit were considered eligible for the survey. This eligibility criterion was also implemented in HRSA's 2014 HCPS and 2009 Primary Health Care Patient Surveys.

2.2 Target Sample Sizes

The study goal was to recruit 210 awardees and complete 9,000 interviews, among them 5,100 for the CHC funding program, 1,480 for the MHC funding program, 1,660 for the HCH funding program, and 760 for the PHPC funding program. The target sample sizes in three design domains, namely funding program, race/ethnicity, and age group, are shown in **Table 2-2**. To achieve the target sample sizes, we planned to oversample patients of MHC, HCH, and PHPC funding types; patients of AI/AN, NH/PI, and Asian race groups; and patients aged 65 or older. Because of the impact of COVID-19 on data collection, HRSA agreed to reduce the target sample size to 4,500.

Table 2-2. Target Sample Sizes for the HCPS

Domain	Target Sample Size	Revised Target Sample Size	Proportion (%)
Funding Type			
CHC	5,100	2,550	56.8
MHC	1,480	740	16.4
HCH	1,660	830	18.4
PHPC	760	380	8.4
Race/Ethnicity			
Hispanic	3,170	1,585	35.2
Non-Hispanic White	2,250	1,125	25.0
Non-Hispanic Black	1,920	960	21.3
Non-Hispanic AI/AN	670	335	7.5
Non-Hispanic Asian	650	325	7.2
Non-Hispanic NH/PI	200	100	2.2
Non-Hispanic Others	140	70	1.6
Age Group			
0–17	2,130	1,065	23.7
18–64	5,770	2,885	64.1
65 or older	1,100	550	12.2
Total	9,000	4,500	

NOTE: AI/AN = American Indian/Alaska Native; CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; NH/PI = Native Hawaiian/Pacific Islander; PHPC = Public Housing Primary Care Program.

2.3 First Stage Sample Design

The first stage sample design involved the selection of a nationally representative sample of awardees. This section discusses the sampling frame construction, sample allocation, and sample selection procedures for the first stage sample design.

2.3.1 Sampling Frame

Data collection was initially planned to start in 2019, and awardee samples for the 2022 HCPS were selected in 2019. Thus, the 2018 HRSA UDS was used to construct the sampling frame for the first stage of selection. The UDS was compiled each year from annual data submissions by each Section 330–funded awardee. The UDS contained data on the number of patients served, awardee characteristics (such as the

types of grant funding received), state, urban/rural location,[3] and number of sites. The awardee characteristics were used in stratification.

The 2018 UDS data were collected from 1,362 awardees. Some awardees were excluded from the sampling frame, including the following:

- 31 awardees located in U.S. territories or possessions (i.e., Puerto Rico, the Virgin Islands, and the Pacific Basin);
- 1 awardee funded through the CHC program that only operated school-based sites;
- 1 awardee with fewer than 300 patients; and
- 8 awardees that received MHC funding only and that served clients through a voucher program.

A total of 1,321 eligible awardees reporting in 2018 were included in the awardee sampling frame. Some key awardee characteristics are shown in **Table 2-3**. In the awardee sampling frame, 928 awardees had a single funding program, whereas 393 awardees received funding from multiple programs. Of 1,248 awardees, roughly 94% received CHC funding, either solely or in combination with other funding programs.

Table 2-3. Awardee Characteristics in the Sampling Frame

Domain Category	Number of Awardees	Percentage Distribution
Funding Program Received		
CHC	864	65.4
HCH	52	3.9
MHC	2	0.2
PHPC	10	0.8
CHC, HCH	164	12.4
CHC, MHC	117	8.9
CHC, PHPC	32	2.4
MHC, HCH	1	0.1
PHPC, HCH	8	0.6
CHC, MHC, HCH	24	1.8
CHC, MHC, PHPC	5	0.0
CHC, PHPC, HCH	33	2.5
CHC, MHC, PHPC, HCH	9	0.7
Total	**1,321**	**100.0**
Region[a]		
Northeast	229	17.3
Midwest	260	19.7
South	449	34.0
West	383	29.0
Total	**1,321**	**100.0**

(continued)

[3] Urban/rural location was defined in the UDS.

Table 2-3. Awardee Characteristics in the Sampling Frame (continued)

Domain Category	Number of Awardees	Percentage Distribution
Urbanicity[b]		
Urban	741	56.1
Rural	580	43.9
Total	**1,321**	**100.0**
Number of Sites		
1	107	8.1
2	155	11.7
3	144	10.9
4–6	307	23.2
7–9	210	15.9
10–14	193	14.6
15–19	83	6.3
≥20	122	9.3
Total	**1,321**	**100.0**

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.
[a] "Region" refers to census region
[b] Urban/rural flag reported in the UDS.

The number of sites within an awardee ranged from 1 to 111, and 1,059 awardees had at least 3 sites, with an average of about 8.7 sites per awardee. The South had 449 awardees, the most in four regions. The West had 383 awardees, and the Northeast and Midwest had roughly the same number of awardees each: 229 and 260, respectively. More awardees were in urban areas than in rural areas.

Another important awardee characteristic is the number of patients served in 2018 (**Table 2-4**). Among the 1,321 eligible awardees in the awardee sampling frame, the number of patients receiving at least one face-to-face encounter for services during 2018 varied among the awardees, ranging from 427 to 232,430 and averaging 21,075 patients. The total number of patients was approximately 27.8 million. **Table 2-5** displays the patient distributions of race/ethnicity and age group. Comparing the percentages shown in **Table 2-2**, it clearly shows that patients in AI/AN, Asian, and NH/PI race/ethnicity categories and patients aged 65 or older need to be oversampled to achieve the target sample sizes.

Table 2-4. Distribution of Patients Served in 2018

Patient Distribution	Number of Patients
Range of Number of Patients	
Minimum	427
25th percentile (Q1)	6,319
Median	12,822
75th percentile (Q3)	25,480
Maximum	232,430
Mean Number of Patients per Awardee	21,075
Total Number of Patients Across All Awardees	27,839,622

Table 2-5. Race/ethnicity, Age Group and Veteran Status Distribution of Patients Served in 2018

Domain Category	Number of Patients	Percent Distribution
Race/Ethnicity		
Hispanic	9,477,226	34.0
Non-Hispanic White	9,916,043	35.6
Non-Hispanic Black	5,103,973	18.3
Non-Hispanic AI/AN	253,507	0.9
Non-Hispanic Asian	987,119	3.6
Non-Hispanic NH/PI	136,301	0.5
Non-Hispanic Others/ Unreported	1,965,453	7.1
Age Group		
0–17	8,581,195	30.8
18–64	16,719,388	60.0
65 or older	2,539,039	9.2
Veteran Status		
Veterans	382,988	1.4
Non-Veterans	27,456,634	98.6
Total	27,839,622	100.0

NOTE: AI/AN = American Indian/Alaska Native; NH/PI = Native Hawaiian/Pacific Islander.

2.3.2 Stratification

As shown in **Table 2-3**, the majority of awardees received CHC funding, while relatively few awardees received PHPC or MHC funding. Randomly selecting awardees without stratification would have resulted in very small awardee sample sizes for MHC and PHPC funding programs. To meet the target of completed interviews for each funding program, we have to complete a large number of interviews for the PHPC and MHC funding programs, which has two implications: (1) it is difficult to recruit many patients from PHPC and MHC awardees within a short period of data collection because of the low patient volume in PHPC or MHC awardees; and (2) the design effect[4] is inflated as the number of completed interviews per awardee increases, and consequently, the estimates will have low precision while the statistical power of comparison is reduced.

Stratification was needed to achieve target sample sizes for four funding programs, age group, and race/ethnicity, with relatively small cluster sizes.[5] We grouped awardees into four exclusive strata according to the types of funding they receive. These four groups served as the first-level strata and are defined in **Table 2-6**.

[4] The design effect is a measure of the precision gained or lost by the use of a more complex design instead of a simple random sample with the same sample size. For a multistage cluster sample like the 2022 HCPS, *deff* is a function of the clustering effect and the unequal weighting effect (*UWE)* and can be defined as *deff = UWE*(1 + (m-1)*ICC)*, where *m* is the number of interviews within an awardee; *ICC* is the intra-cluster correlation coefficient that measures the degree of similarity among respondents within an awardee; and *UWE* measures variation in the sample weight.

[5] Cluster size is measured as the number of completed interviews within an awardee for a funding program.

Table 2-6. Definition of First-Level Stratification

First Stage Strata	Awardee Funding Type	Number of Awardees in Sampling Frame
Stratum 1: Awardees received PHPC funding solely or in combination with other programs.	P; CP; PH; CMP; CPH; CMPH	97
Stratum 2: Awardees received MHC funding solely or in combination with CHC or HCH.	M; CM; MH; CMH	144
Stratum 3: Awardees received HCH funding solely or in combination with CHC.	H; CH	216
Stratum 4: Awardees received CHC funding solely.	C	864
Total		**1,321**

NOTE: C = Community Health Center (CHC) program; H = Healthcare for the Homeless (HCH) program; M = Migrant Health Center (MHC) program; P = Public Housing Primary Care (PHPC) program; multiple acronyms used together indicates that funding was received from multiple programs (e.g., CMH = an awardee received CHC, HCH, and MHC funding; CMP = an awardee received CHC, MHC, and PHPC funding).

AI/AN, Asian, and NH/PI patients were not evenly distributed among all awardees. They tended to be clustered in a few awardees: 975 awardees (75%) had fewer than 100 AI/AN patients, 1,023 awardees (84%) had fewer than 100 NH/PI patients, and 620 awardees (47%) had fewer than 100 Asian patients. The 20 awardees with the highest proportion of AI/AN patients accounted for 27.1% of total AI/AN patients in all 1,321 awardees; 20 awardees with the highest proportion of NH/PI patients accounted for 43.8% of total NH/PI patients; and 20 awardees with the highest proportion of Asian patients accounted for 29.6% of total Asian patients. Thus, to achieve target sample sizes in three race/ethnicity categories, awardees with concentrated patients in those three race/ethnicity categories must be obtained and selected at the first stage selection. Awardees with more than 20% of patients in one of the three race/ethnicity categories were considered patient-concentrated awardees. Stratum 4 (CHC funding solely) had over 87% of such awardees, and very few such awardees were from Strata 1, 2, and 3. Therefore, to effectively select awardees with concentrated patients in three race/ethnicity categories, Stratum 4 was further divided into two second-level strata according to whether an awardee has concentrated patients (more than 20%) in any one of the three race/ethnicity categories.

Although some awardees had a high proportion of patients aged 65 or older, older patients were distributed more evenly than the patients in the three race/ethnicity categories. The 20 awardees with the highest proportion of patients aged 65 or older only accounted for 1.26% of total patients aged 65 or older. Similarly, there were no awardees with concentrated veteran patients. As a result, oversampling awardees with concentrated patients aged 65 or older or veteran patients at the first stage of selection was not as effective as oversampling awardees with concentrated patients in the three race/ethnicity categories. Thus, we decided not to oversample awardees with concentrated patients aged 65 or older, or veteran patients. The plan was to oversample patients aged 65 or older or veteran patients at the third stage of selecting patients.

In Stratum 1, the awardees with only PHPC funding have fewer patients than the awardees with multiple funding types. A probability proportional to the size (PPS) sample in Stratum 1 will yield very few PHPC-only awardees. To overcome this problem, we further divided Stratum 1 into three substrata according to the patient volume and the proportion of PHPC patients in an awardee. There were seven final awardee strata, shown in **Table 2-7**.

Table 2-7. Awardee Sample Final Stratification

First Stage and Second Stage Strata	Awardee Funding Type	Final Stratum	Number of Awardees in Sampling Frame
Stratum 1: Awardees received PHPC funding solely or in combination with other programs	P; CP; PH; CMP; CPH; CMPH		
Stratum 1.1: Awardees with 25% or more than 25% of PHPC patients		1	36
Stratum 1.2: Awardees with less than 25% of PHPC patients			
Stratum 1.2.1: Large awardees		2	31
Stratum 1.2.2: Small awardees		3	30
Stratum 2: Awardees received MHC funding solely or in combination with CHC or HCH	M; CM; MH; CMH	4	144
Stratum 3: Awardees received HCH funding solely or in combination with CHC	H; CH	5	216
Stratum 4: Awardees received CHC funding solely	C		
Stratum 4.1: Awardees with more than 20% of AI/AN, or 20% Asian, or 20% of NH/PI patients	C	6	71
Stratum 4.2. Other awardees in Stratum 4	C	7	793
Total			**1,321**

NOTE: AI/AN = American Indian/Alaska Native; C = Community Health Center (CHC) program; H = Healthcare for the Homeless (HCH) program; M = Migrant Health Center (MHC) program; NH/PI = Native Hawaiian/Pacific Islander; P = Public Housing Primary Care (PHPC) program; multiple acronyms used together indicates that funding was received from multiple programs (e.g., CMH = an awardee received CHC, HCH, and MHC funding; CMP = an awardee received CHC, MHC, and PHPC funding).

2.3.3 Sample Allocation

Before selecting an awardee sample from each final stratum, we determined the awardee sample allocation for each final stratum. Oversampling awardees who received funding from PHPC, MHC, or HCH programs and awardees with concentrated patients in three oversampling race/ethnicity categories introduces more variation in sample weights, thus increasing unequal weighting effects (UWE). To minimize the variation in sample weights we allocated the awardee sample using a nonlinear optimization procedure, OPTMODEL in SAS (SAS, n.d.), which minimizes the UWE with the following constraints of the original target sample sizes:

- recruit 210 awardees;
- complete 9,000 interviews;
- complete 5,100 CHC interviews, 1,480 MHC interviews, 1,660 HCH interviews, and 760 PHPC interviews;
- complete interviews per awardee: 26 for CHC, 25 for MHC, 25 for HCH, and 16 for PHPC; and
- select at least one awardee from each awardee type.[6]

The optimum sample allocation to each awardee type is presented in **Table 2-8**. After aggregating awardee allocations to the seven final strata, the awardee sample allocation to the seven strata along with the sampling rates in each stratum are shown in **Table 2-9**. Assuming a 75% awardee recruitment rate, we

[6] Awardee type is defined according to what funding program(s) an awardee participated in or received funding from.

selected 280 awardees. The sampling rates for Strata 1, 2, 3, 4, and 6 are much higher than the overall sampling rate (21.2%), indicating that we oversample awardees in these strata.

Table 2-8. Optimum Awardee Sample Allocation

Domain Category	Awardees	
	Number	Sample Allocation
Funding Program Received		
C	864	90
H	52	3
M	2	2
P	10	7
CH	164	23
CM	117	29
CP	32	11
MH	1	1
PH	8	8
CMH	24	13
CMP	5	5
CPH	33	9
CMPH	9	9
Total	1,321	210

NOTE: C = Community Health Center (CHC) program; H = Healthcare for the Homeless (HCH) program; M = Migrant Health Center (MHC) program; P = Public Housing Primary Care (PHPC) program; multiple acronyms used together indicates that funding was received from multiple programs (e.g., CMH = a awardee received CHC, HCH, and MHC funding; CMP = a awardee received CHC, MHC, and PHPC funding).

Table 2-9. Awardee Sample Allocation and Sampling Rates in Final Awardee Strata

First Stage and Second Stage Strata	Final Stratum	Number of Awardees in Sampling Frame	Awardee Sample		Sampling Rate (%)	Recruited Awardees
			Selected	Released		
Stratum 1: Awardees received PHPC funding solely or in combination with other programs						
Stratum 1.1: Awardees with 25% or more than 25% of PHPC patients	1	36	27	27	75.0	7
Stratum 1.2: Awardees with less than 25% of PHPC patients						
Stratum 1.2.1: Large awardees	2	31	21	21	67.7	5
Stratum 1.2.2: Small awardees	3	30	17	17	56.7	6
Stratum 2: Awardees received MHC funding solely or in combination with CHC or HCH	4	144	60	60	41.7	26

(continued)

Table 2-9. Awardee Sample Allocation and Sampling Rates in Final Awardee Strata (continued)

First Stage and Second Stage Strata	Final Stratum	Number of Awardees in Sampling Frame	Awardee Sample		Sampling Rate (%)	Recruited Awardees
			Selected	Released		
Stratum 3: Awardees received HCH funding solely or in combination with CHC	6	216	35	35	16.2	18
Stratum 4: Awardees received CHC funding solely						
Stratum 4.1: Awardees with more than 20% of AI/AN, or 20% Asian, or 20% of NH/PI patients	6	71	71	71	100.0	21
Stratum 4.2. Other awardees in Stratum 4	7	793	49	49	6.2	20
Total		**1,321**	**280**	**280**	**21.2**	**103**

NOTE: AI/AN = American Indian/Alaska Native; CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; NH/PI = Native Hawaiian/Pacific Islander; PHPC = Public Housing Primary Care Program.

2.3.4 Sample Selection

As mentioned in **Section 2.3.1**, the awardees differed widely in the number of patients served. PPS sampling is a commonly used method of unequal probability sampling to handle the large variation in patients served among awardees. In this method, the probability of an awardee being sampled is proportional to a size measure. The size measure was the number of patients who visited the awardee for services from the 2018 UDS file. We selected a PPS awardee sample from each final stratum.

A PPS awardee sample was selected using the SAS SURVEYSELECT (SAS, n.d.) procedure with predetermined sample allocation in **Table 2-10** for each final stratum. During the selection, in addition to the seven strata for awardee sample selection discussed as shown in **Table 2-9**, we sorted the sampling frame by region (Northeast, Midwest, South, and West), urban/rural location, and awardee size (large, medium, small) when applying Chromy's (1981) probability minimal replacement sequential PPS selection procedure. Sorting the sampling frame by these key awardee characteristics and then applying the PPS sequential procedure induced implicit stratification according to the order of the units in a stratum. Therefore, the selected awardee samples were distributed among various regions, urban/rural locations, and awardee sizes to ensure that a representative awardee sample is selected.

Table 2-10 displays the awardee sampling frame and awardee sample distribution by region, urban/rural area, and awardee size. In the distribution of regions, the West has a higher proportion in the awardee sample, while the proportions of the South and Midwest in the awardee sample are lower compared to the awardee sampling frame. This difference is mainly a result of oversampling awardees with concentrated AI/AN and NH/PI patients; the majority of these awardees are in the West region (Alaska and Hawaii). The awardee sample has higher proportions in urban areas compared with the awardee sampling frame; the reason for this difference is that we oversample PHPC awardees and they are mainly in urban areas. The awardee sample has lower proportions of small- and medium-size awardees than the awardee sampling frame. This disparity occurs because of the PPS sampling method employed in awardee sample selection, which gives awardees with large patient volume a better chance of being selected than awardees with small patient volume.

Table 2-10. Awardee Sample Distribution by Region, Urban/Rural Location, and Awardee Size

Domains	Awardee			
	Frame		Sample	
	n	%	n	%
Region	1,321	100.00	103	100.00
Northeast	229	17.34	19	18.45
Midwest	260	19.68	14	13.59
South	449	33.99	23	22.33
West	383	28.99	47	45.63
Urbanicity	1,321	100.00	103	100.00
Urban	741	56.09	68	66.02
Rural	580	43.91	35	33.98
Awardee Size	1,321	100.00	103	100.00
Large	450	34.07	68	66.02
Medium	436	33.01	17	16.50
Small	435	32.93	18	17.48

If an awardee received funding from multiple programs, an independent site and patient sample was selected from each funding program. Thus, recruiting 103 awardees is equivalent to selecting a sample from 181 awardees (see **Table 2-1**). Of those, 100 served CHC patients, 28 served MHC patients, 32 served HCH patients, and 21 served PHPC patients. Because of low patient volume (i.e., the absence of patients in the funding program at an awardee) or language barriers, patients could not be selected and interviews could not be conducted at some MHC, HCH, and PHPC awardees. Data were collected from a total of 102 awardees and 179 funding programs (99 for CHC, 27 for MHC, 32 for HCH, and 21 for PHPC).

2.4 Second Stage Sample

Although some awardees provided services through a single site, most provided services at two or more sites. Therefore, the second stage sample design entailed selecting sites within awardees.

2.4.1 Sampling Frame

The 2018 UDS did not provide detailed site-level information about funding programs and patient volume. Therefore, to prepare the second stage sampling frame, sampling information was collected about each site when the awardee recruiters solicited awardee participation. Once an awardee was recruited and agreed to have the study conducted at its sites, recruiters worked with the awardee's administration to identify eligible sites. The following eligibility criteria were used, and the HRSA Project Officer was consulted to determine site eligibility on a case-by-case basis whenever necessary:

- The site should participate in at least one of the four specific funding programs and must have been operating under the awardee for at least 1 year.
- The site is not a school-based health center.
- The site is not a specialized clinic, except clinics providing OB/GYN services or pediatric care.
- The site does not provide services only through the migrant and seasonal farm worker voucher screening program.

- The site serves at least 100 patients for a funding type.

After the eligible sites were identified, the following information was collected from or verified with each participating awardee:

- Number of eligible sites serving each patient type (i.e., migrant and seasonal farm workers; homeless, public housing, and general patients).
- Address and contact information for each eligible site.
- Number of patients served during the previous year at each eligible site, overall and by type of patient (CHC, MHC, HCH, and PHPC).
- Sites with concentrated patients (more than 20%) in one of the three race/ethnicity categories (AI/AN, Asian, or NH/PI).

In most cases, one Field Interviewer (FI) was hired to collect data for each participating awardee. Therefore, selected sites must be within manageable distances for the FI(s). The awardees tend to operate sites in relatively localized areas. We evaluated distances between the administrative office/central site and the associated sites. For a specific funding program, the site with the largest patient volume was used as the central site. Typically, sites were excluded if they were located more than 100 miles from the central site.

2.4.2 Sample Selection

Sites were selected independently from the site sampling frame for each funding program if the awardee received funding from multiple programs.

If there were three or fewer sites for a patient type (i.e., migrant and seasonal farmworkers; homeless, public housing, and general patients) and they were within a manageable distance for one FI, all of the sites were included in the study. If one site was far from the other sites and the other sites were close to one another, the two sites that were close to each other were selected. However, if all three sites were far from one another, we selected the site with the largest patient volume. Similarly, when two sites for a specific funding program were far from each other, the one with the largest number of patients was selected.

For awardees with more than three sites for a patient type, we used a PPS sampling method similar to the one for awardees discussed in **Section 2.3.4** to select three sites from the sites within a manageable distance. The number of patients served by each site under a specific funding program served as the size measure in the PPS sampling.

To achieve our target sample sizes of AI/AN, Asian, and NH/PI patients, we not only oversampled awardees with concentrated patients in these three race groups at the first stage of selection, but we also identified sites with concentrated patients in at least one of the three targeted race/ethnicity categories. These sites were selected with higher probabilities than sites without concentrated patients. Sites with concentrated patients aged 65 or older were also selected with higher probabilities than sites without concentrated patients.

As shown in **Table 2-1**, 318 sites were recruited. Of those, 248 served CHC patients, 65 served MHC patients, 71 served HCH patients, and 43 served PHPC patients. Data were collected from a total of 316 sites, including 247 for CHC, 64 for MHC, 70 for HCH, and 43 sites for PHPC.

2.5 Third Stage Sample

The third stage sample design involved selecting patients for the study. Because some of the target populations of this study are quite mobile, a random sample of patients was selected for interview as they entered the site and registered with the receptionist for services. Because of the COVID-19 pandemic, the survey design switched from in person only to mixed mode. For in-person data collection, an FI visited a

selected site for a predetermined number of days and time slots in the sampling period to conduct interviews. For telephone data collection, patients who called in for an interview and were eligible after screening were interviewed. Data collection started in January 2021with telephone interviewing only, and switched to a mixed-mode design (telephone and in person) in July 2021.

2.5.1 Patient Interview Allocation to Awardees

To achieve the near self-weighting sample of patient interviews within each awardee stratum, the same number of patient interviews was desired from the awardees in each funding program. The interview quota for each awardee was determined by evenly allocating the targeted number of completed interviews to all participating awardees for a funding program, then inflating this target number to produce a production goal. The production goal assigned to each awardee was slightly inflated because some awardees were anticipated to have difficulty in achieving the goal because of low patient volume, particularly for MHC, HCH, and PHPC awardees. By doing so, the awardees with high patient volume could compensate for production challenges faced by the low-volume awardees. **Table 2-11** shows the quota per awardee for each funding type.

Table 2-11. Patient Interview Quota per Awardee

Funding Program	Patient Interview Quota/Awardee
CHC	27
MHC	27
HCH	27
PHPC	18

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.

2.5.2 Patient Interview Allocation to Sites Within an Awardee

Within each awardee, we used different methods to allocate patient interviews to multiple sites for awardees with three or fewer sites in a funding program and awardees with more than three sites in a funding program. For awardees with three or fewer sites, the number of patient interviews within that awardee were allocated proportionally to the patient size of the sites. That is,

$$n_{fij} = n_{fi} \frac{s_{fij}}{\sum_j s_{fij}},$$

where n_{fi} is the number of patients selected and S_{fij} is the number of patients in j^{th} site from an awardee for funding program *f*. For awardees with more than three sites that were selected through PPS, the number of selected patients was divided equally among three selected sites. Doing so will help to reduce the UWE.

2.5.3 Patient Screening and Selection in a Mixed-Mode Design

To oversample patients in the three race/ethnicity categories, patients aged 65 or older, and veterans, we designed a screening sheet that receptionists could use to screen and select patients when a patient entered the site and registered for service. A patient would be considered eligible if they had received service through one of the awardees supported by HRSA funding programs at least once in the 12 months prior to the current visit.

Our original plan was that receptionists could ask eligible patients questions about their race/ethnicity and age to determine whether they belonged to the oversampling groups. If a patient was not in an oversampling group, the receptionist selected the first eligible patient registered after the FI informed the receptionist that they were ready for the next interview. The receptionist read a brief script about the study to the selected patient and directed the patient to the FI for questions or participation. If a patient belonged to one of the oversampling groups, the receptionist selected the patient and sent the patient to the FI if they were available. However, this screening and oversampling procedure was not fully implemented because of the impact of COVID-19. In health sites where in-person data collection was applied, this

oversampling plan was implemented. For health sites where phone data collection was carried out, it was impossible to implement the oversampling plan.

The receptionist was asked to track the number of patients who enter the site, the number of patients who were eligible, and the number of patients selected while the FI was at the site to conduct data collection. The receptionist used tally marks to count patients as they entered or completed a table based on the sign-in sheet or appointment list before the FI left the site. The patient tally sheets for each day the FI visited the site were sent to RTI for data entry, and counts were planned to calculate the analysis weights for the study (see **Section 6** for more details). For sites that have more than one receptionist, all receptionists tracked the number of patients that visited, even though only one receptionist was selected to recruit patients. This process was also affected by COVID-19, where the information on number of patients visited and referred was missing from many health center sites.

The original plan was that if a site was chosen for data collection in multiple funding programs, the FI screened participating patients to determine patient population type (i.e., homeless, migrant and seasonal farmworkers, public housing, or low income) and used the appropriate questionnaire to conduct the patient interview. This plan was implemented for some sites when in-person data collection was used. For sites where phone data collection was conducted, the receptionist handed out a study brochure and a flyer (see **Section 4.3.1** for more details) to every eligible patient. The flyer had the field supervisor's (FS's) phone number that patients needed to call to participate. When a patient called, the FS confirmed the patient's eligibility status (i.e., the patient received a service at least once in the past year from the site). Then, the FS scheduled an appointment for an interview with the FI. The FI called the patient at the scheduled appointment time and conducted screening and interviewing using the same procedure as in person.

At the early stage of data collection, interviewers screened and selected patients who were in the patient type that a site was selected for, and the quota for that funding type had not been met. This was the same procedure that was implemented in the 2014 HCPS. However, this requirement was relaxed in the late stage of data collection. As long as the patient was eligible, the patient was selected and interviewed regardless of funding type. For example, if a site was selected for CHC funding type only, an eligible patient who was a migrant or farmworker, living in public housing, or homeless was also selected and interviewed.

As shown in **Table 2-1**, 5,194 patients were classified as eligible (selected): 3,440 for CHC, 604 for MHC, 658 for HCH, and 492 for PHPC, and 4,414 patient interviews were completed: 2,915 for CHC, 473 for MHC, 597 for HCH, and 429 for PHPC. **Table 2-12** displays the patient sample distribution. The patient sample had a higher proportion in the West and a lower proportion in the Midwest. The difference of the proportion in region was corrected by poststratification in calculating sample weights, is discussed in **Section 6.8**. The patient sample distribution for urbanicity and awardee size was very similar to the patient population. The patient sample had a higher proportion of large awardees than the patient population and a lower proportion of medium awardees than the patient population.

Table 2-12. Patient Sample Distribution by Region, Urban/Rural Location, and Awardee Size

Domains	Patient Population[a]		Patient Sample	
	n	%	n	%
Region	27,839,622	100.0	4,414	100.0
Northeast	5,477,225	19.7	748	17.0
Midwest	5,023,946	18.1	427	9.7
South	8,348,508	30.0	1,128	25.6
West	8,989,943	32.3	2,111	47.8
Urbanicity	27,839,622	100.0	4,414	100.0
Urban	18,520,051	66.5	3,001	68.0
Rural	9,319,571	33.5	1,413	32.0
Awardee Size	27,839,622	100.0	4,414	100.0
Large	20,187,793	72.5	3,674	83.2
Medium	5,694,021	20.5	421	9.5
Small	1,957,808	7.0	319	7.2

[a] Patient population was based on the patient counts from 1,321 awardees in the awardee sample frame in the preliminary 2018 UDS.

Some limitations of the patient selection in the 2022 HCPS should be noted: (1) The planned patient selection process could not be implemented in sites where phone data collection was implemented. Instead, all eligible patients were informed about the survey and interested patients called the phone number provided to participate in the study. Therefore, this patient selection process was not entirely random. (2) The data collection time spent at each awardee or site to achieve a predetermined awardee patient quota was different. It is likely that the patient sample in the study could overrepresent certain types of patients, such as patients with seasonal flu or COVID-19, or miss certain types of patients, such as patients who visited the site at longer intervals than the data collection period. Consequently, patient characteristics of the patient sample might be different from the patient population, such as age, race/ethnicity, and medical condition. Some patient characteristics can be corrected in the poststratification adjustment of sample weight calculation, such as age, and race/ethnicity(discussed in **Section 6.8**). However, whether differences existed and the magnitude of any difference remained unexamined for some patient characteristics (e.g., patient medical conditions).

3. Data Collection Instruments

3.1 Questionnaire Development

HRSA and RTI collaborated on the development of the 2022 HCPS instrument. Questionnaire development began in October 2017. Initial review of the questionnaire focused on efforts to streamline the questionnaire. Questions were combined where possible, open-ended responses were reviewed, and new response categories added in an effort to remove open-ended items. Items with low response in 2014 were identified as possible candidates to be removed.

A meeting with a Technical Advisory Panel focusing on reviewing the first draft of the questionnaire was held in March 2018. The questionnaire was subsequently streamlined and new questions added to meet the data needs of HRSA, and a final version was prepared for cognitive interviewing. Two rounds of cognitive interviewing were conducted, using the programmed version of the survey instrument, with respondents in each language. Following cognitive testing the computer-assisted personal interviewing (CAPI) instruments were prepared for data collection activities. As a result of the pandemic that began in March 2020, RTI adapted the questionnaire to allow for telephone data collection. In addition, some questions were revised and new questions added to collect data on COVID-19.

The data elements included in the survey instrument aimed to gather information related to patients'

- Care-seeking behaviors
- Sociodemographic characteristics
- Reasons for seeking care
- Health status
- Use of services
- Satisfaction with care
- Unmet health care needs
- Perceived quality of care

The translated instruments from 2014 were revised for the 2022 HCPS. New questions were translated and incorporated into the instrument. In 2014, we had a Korean version of the survey instrument, but determined that this should be removed as the Korean instrument was not used in 2014. A decision was made to replace the Korean instrument with Tagalog as we missed a large portion of potential Filipino respondents in 2014 because they did not speak English. RTI's in-house language methodologist supervised the translation of the edits made to the English questionnaire into the various languages using a multistep translation process.

3.2 Questionnaire Modules

To meet the 2022 HCPS research goals, the final HCPS instrument included 19 modules. Each of the following modules was administered to patients: introduction, access to care, routine care, conditions, follow-up conditions, cancer screening, health center services, health insurance, prescription medication, dental, mental health, substance use, prenatal care/family planning (females aged 15–49), HIV testing (all respondents aged 18+), living arrangements, neighborhood characteristics (new module in 2022), income and assets, and demographics. Most items applied to all sample members. However, some sections were only applicable to a subset of sample members (i.e., questions on pregnancy were only asked of women of child-bearing age).

Table 3-1 lists all modules, topics, number of questions, and types of questions included in each. Respondents were routed through the questionnaire using skip logic so they were not asked questions that did not apply to their situation. Note that the PUF has gone through the disclosure process described in **Section 1.3.1**. Therefore, all variables and topics listed below in the instrument description may not appear on the data file. This is consistent with the need to protect respondent confidentiality.

Table 3-1. Description of HCPS Instrument

Module	Topic	Number of Questions	Types of Questions
S	Patient screening	10	1. Age category 2. Race 3. Eligibility questions (health services received in the past 12 months, farm work, homelessness, public housing status) 4. Consent for interview and audio recording
A	Introduction	14	1. Date of birth 2. Language spoken 3. Race Note: age was asked here to also help navigate the skip logic within the instrument.
B	Access to care	7	1. Access to care 2. Reason for inability to get or delay in getting medical care needed (similar questions on dental care, prescription medicines, counseling/mental health treatment, and prenatal care/family planning are in other modules)
C	Routine care	26	1. Health care providers seen in past 12 months 2. Vaccinations received 3. Reasons why not received recent checkup
D	Conditions	85	1. Height and weight 2. Weight management/exercise 3. Medical history and conditions (pregnancy, hypertension, blood cholesterol, hepatitis, tuberculosis, asthma, diabetes, cancer, hearing, vision, and other health conditions) 4. Questions about COVID-19 5. Activities of daily living
E	Conditions follow-up	60	1. More specific questions on care received for health conditions (high blood pressure, asthma, diabetes, blood cholesterol, and other health conditions)
F	Cancer screening	49	1. Cancer screening services received (pap smear, human papilloma virus, mammogram, colonoscopy/sigmoidoscopy exam, blood stool test)
G	Health center services	56	1. Usual source of care 2. Telehealth 3. Referrals 4. Language assistance received 5. Help received to access social programs 6 Distance/mode of transportation to health center 7. Health center services experience and satisfaction with a wide range of health center characteristics 8. Information technology available at health center

(continued)

Table 3-1. Description of HCPS Instrument (continued)

Module	Topic	Number of Questions	Types of Questions
H	Health insurance	25	1. Current insurance coverage 2. Reasons for lack of insurance 3. Coverage provided by insurance
I	Prescription medication	23	1. Prescription services experience 2. Ease of getting prescription 3. Satisfaction level 4. Skipped/delayed/other reasons to save money
J	Dental	38	1. Reason for inability to get or delay in getting dental care 2. Where dental treatment was received 3. Condition of teeth 4. Dental problems
K	Mental health	53	1. Questions about feelings 2. Autism/Developmental Delay 3. Questions about suicide thoughts and intentions 4. Reason for inability to get or delay in getting mental health care 5. Where mental health treatment or counseling was received (e.g., prescription medication, group or individual counseling, inpatient treatment)
L	Substance use	116	1. Use of substances (vaping, tobacco, alcohol, illicit drugs), substance abuse treatment 2. Medication-assisted treatment
M	Prenatal care/ family planning (females aged 15–49)	37	1. Reason for inability to get or delay in getting prenatal care and family planning services 2. Rating of prenatal care and family planning services
N	HIV testing	15	1. Whether HIV test was received; if not, reason why
O	Living arrangements	19	1. Type of place where living currently 2. Crowding 3. Extent of past homelessness experience 4. Food insecurity
P	Neighborhood characteristics	8	1. Condition of neighborhood 2. Communication with neighbors 3. Quality of schools
Q	Income and assets	13	1. Family income 2. Receipt of income support and public assistance
R	Demographics	35	1. Place of birth 2. Education 3. Marital status 4. Veteran status 5. A series of questions on employment status, participation in employer-sponsored health insurance, moves in past 12 months; questions for migrant seasonal farm worker respondents on farm work experience

4. Data Collection Methods

4.1 Awardee Recruitment, Site Contacts, and Training

Recruitment of the sampled awardees and sites began in January 2020. The recruiting activity included contacting each sampled awardee and site; describing the nature and purpose of the study; identifying and establishing rapport with the key stakeholders; addressing concerns and answering questions; and encouraging stakeholders to allow participation in the study.

4.1.1 Preparations for Awardee Recruitment and Site Contacts

RTI developed a team of nine recruiters (who also served as FSs) selected for their field experience and their skills in gaining cooperation required to secure health center participation. The team received a 2-day in-person training on study goals, design, and processes to secure participation.

The recruiters were also responsible for drafting, initiating, and obtaining written notification of approval to participate for each of the sites. Recruiters tailored the Letter of Agreement (LOA) to each site, which included information regarding obtaining informed consent from minors (aged 13–17) and proxies, designated site contacts, and protocols for distressed subjects and other suspected problems requiring intervention.

The recruiters used a recruiting information system (RIS) to track and document all communications with the awardees and sites. This proprietary software supported recruiters through each stage of contact with the awardee and site staff, tracking each milestone in this communication, and providing standardized reports on recruitment status. The system allowed recruiters access to all stored data, including contact histories, approval/permission status, sample status, and reference materials.

4.1.2 Awardee Recruitment Procedures

RTI sent advance packages to the selected awardees, which marked the start of the recruiting process. The advance package contained the following:

- Personalized RTI lead letter from the Project Director
- Lead letter from HRSA
- RTI brochure

The first contact with the awardee office was a brief telephone call in which the recruiters confirmed the name, title, and correct mailing address for the awardee chief executive officer or decision maker who would receive the introductory packet. The recruiters also captured the telephone number and email address of the awardee contact. All updates were entered into the RIS and used to generate the introductory packet mailing.

The advance packages were sent on a flow basis via UPS overnight or as email attachments, depending on the awardee's preference. The recruiters made the first follow-up call approximately 2 to 3 days after the mailing.

During the follow-up calls, the recruiter discussed in detail the study's objectives, its operational components, the project schedule, and what would be expected of staff at the awardee and site facilities. The goal of the telephone and email contacts was to identify requirements and secure approval for participation at the awardee level, identify potential barriers to participation, and collect information needed to draw the site sample. The recruiter also discussed the study's protection of human subjects policies and any local requirements or policies concerning research with children and adolescents.

Recruiters obtained permission/approval from all applicable awardee administrators and applicable review boards, such as local Institutional Review Boards (IRBs), prior to any awardee's participation.

Each recruiter assisted in preparing for the local review process, as appropriate. The awardees were provided with all the documentation required for study review and approval. These materials included the RTI IRB protocol application for this study, all RTI IRB-approved consent and assent forms, the RTI IRB approval letter, recruitment and data collection materials, and the survey instrument.

During the awardee recruitment process, a few awardees were found to be ineligible or refused to participate. Ineligibility occurred when the awardee no longer received funding or when a high volume of patients did not speak English or any one of the languages the instrument was translated to.

Prior to the COVID-19 pandemic, the most commonly cited reasons for refusal were privacy concerns and key decision makers' perception of an excessive burden on awardee and site staff and resources. The onset of the COVID-19 pandemic created many challenges with the awardee and site recruitment process. Awardees who initially agreed to participate refused or asked to delay data collection at their sites because of a reduction in workforce, high COVID-19 rates in their areas, and the implementation of COVID-19-related activities, including vaccinations. Some sites had to temporarily shut down whereas others had changes in management. Recruiters had to re-recruit many of the sites, determine new points of contact, and reintroduce the study. At this stage of recruitment, the recruiters informed the sites of the change in survey design from in person only to mixed mode. The first batch of sites that were recruited were for telephone mode only for a pilot data collection. After the pilot, more sites were recruited for both in-person and telephone data collection.

Once the recruiters secured cooperation from the awardees, an Awardee Information Sheet (AIS) was sent to them to complete. Awardees had the option to complete the AIS online or to send the AIS to the recruiter. When the awardees completed the AIS, the recruiters entered this information into the RIS and marked the awardee as ready for sampling. This prompted the sampling statistician to select the sites and assign quotas for each site. Each week, sites were selected according to the procedures described in **Section 2.3.4**.

The recruiters made final contact with the awardees after the sites were selected and informed them on which sites had been selected and the patient interview quota at those sites. At that time, the recruiters also encouraged awardees to serve as advocates for the 2022 HCPS and personally contact the sites to encourage support for data collection.

4.2 Site Contacting Procedures

The awardee organization and each associated site were usually recruited independently; that is, some sites required permissions and approvals beyond those obtained at the awardee level. In a few cases, the contact person at the awardee level was the same contact at the service site level; however, most sites had a separate contact person. The recruiters did not contact any site without first obtaining permission from the awardee contact and not until after the awardee first spoke with the site contacts about the study. Once recruiters were allowed to contact the site staff, they did so immediately and quickly provided study information to the appropriate contacts.

Site-level contacts were sent an advance package that contained the same materials as the one sent to the awardees, except that the site lead letter was directed to the sites and also indicated that RTI had received permission from the awardees to contact them. The site advance package contained the following documents:

- Personalized RTI lead letter from the Project Director
- Lead letter from HRSA
- RTI brochure

During the initial and follow-up telephone contacts, the recruiter identified requirements and approvals for participation at the site level, identified potential barriers to participation, and identified critical

information required by the data collection team. Such information included each site's hours of operation, identification of space at the site for interviews, and identification of the two site contact persons who would assist the FI with site logistics and act as a contact person for critical incidents. As with the awardee recruitment, guidelines were used for the site contacting effort.

After *all* approvals were obtained at the site level, recruiters initiated an LOA with the appropriate administrator at each participating site. This letter outlined all tasks required for the study, specified any restrictions imposed by the local IRB or other review committees, identified all contact people, and specified the type of remuneration that the site preferred to be used with their participating patients. Study activities did not commence at any site until all required approvals were obtained and the LOA was signed and returned. The signed LOAs and the site profile were provided to the appropriate field staff.

4.2.1 Additional IRB/Committee Approvals

RTI anticipated that many awardees and sites would have boards of governance or advisory boards and securing approval from these boards would be necessary before the study could proceed. In such cases, RTI cooperated in whatever way possible and provided study materials whenever requested.

Among participating awardees, five needed their local IRB approval before they could participate in the 2022 HCPS. Because of COVID-19, these awardees had challenges obtaining their local IRB approval and, hence, were unable to participate in study.

Awardees in Alaska needed approval from the Alaska Area Institutional Review Board (AAIRB) before they could participate. The application for the AAIRB was completed and approved. Even after receiving approval, several awardees were difficult to reach and others refused for reasons related to COVID-19.

4.3 Respondent Recruitment

Respondent selection was conducted through onsite sampling. Although the FIs were provided the days they were to work at the site during the data collection period and received training on the sampling process, they were not directly involved in sampling because of patient confidentiality issues. The FI's job was to be at the site while sampling occurred to recruit all sampled patients who expressed an interest in the study.

As each patient arrived at the site on a day the FI was present, the receptionist would register the patient to receive health services. The receptionist would then determine whether each arriving patient met the initial eligibility criteria to be considered for the 2022 HCPS (i.e., had received services at least once in the past 12 months and was not an unaccompanied 13- to 17-year-old). If the patient met the initial eligibility criteria, the receptionist selected the first patient who registered after the FI informed the receptionist that they were available in the waiting room and ready to administer the next interview. The receptionist would read the brief receptionist/respondent recruitment script to the patient (or to their parent or guardian, for selected children) and give them a copy of the 2022 HCPS brochure.

If the selected patient was interested in participating in the 2022 HCPS or had questions about the study, they were directed to approach the FI, who was waiting in a designated area at the site. The FI gave a short description of the 2022 HCPS interview using the interviewer recruitment script and answered any questions. If the patient was interested in participating in the study, the FI would take them to a designated private location at the site to begin screening, obtain verbal informed consent, and start the interview process. The FI asked the patient some initial screening questions to confirm eligibility for the study before the actual interview began. If the patient was eligible to participate in the 2022 HCPS, the FI either continued with the interview or scheduled an appointment if the respondent could not begin the interview right away. For scheduled appointments or breakoffs, the FI asked respondents for contact information (first name and phone number where they could be reached). A breakoff occurred if respondents were unable to complete the interview in one sitting (e.g., if they needed to leave for their doctor's appointment) and wished to complete the interview at a later date. For appointments and

breakoffs, the FI and respondent agreed on a location and time to meet and complete the 2022 HCPS interview.

4.3.1 Telephone Interview Respondent Recruitment

Patient sampling and recruitment procedures for phone interviews was comparable to the procedures of having an interviewer at the site. As each patient entered the site during the sample selection period, the receptionist(s) registered them to receive health services and recorded a tally mark on the patient tally sheet. The receptionist determined whether each arriving patient met the initial eligibility criteria to be considered for the 2022 HCPS (i.e., had received services at least once in the past year and was not an unaccompanied 13- to 17-year-old). If the patient was 13–17 years old and not accompanied by a parent or legal guardian, the patient was automatically ineligible. The receptionist also asked eligible patients questions about their race/ethnicity, veteran status, and age to determine whether they belonged to the oversampling groups. The receptionist introduced the study to the patient (or to the parent/legal guardian if an adolescent of 13–17 years) and encouraged them to participate. The receptionist also handed out a study brochure and a flyer. The study brochure highlighted the importance of the study, the types of questions asked, privacy, why the patient was selected, and the incentive they will get after completing the interview. The flyer had the FS's phone number they needed to call to participate. To ensure legitimacy of the respondent, the flyer had a unique identifier that the FS asked for when the patient called. Flyers for patients age 13–17 were printed in a different color. This helped receptionists determine which type of flyer to hand out to the parent. The unique ID printed on the flyer was different from the sequential IDs printed on the flyers for the adult and proxy interviews.

Protocol for Interviewing Adults and Proxies for Children

When a patient called, the FS first asked for the unique ID printed on the flyer. The FS then confirmed the patient's eligibility (i.e., the patient received services at least once in the past year). If deemed eligible, the FS scheduled an appointment for an interview with the interviewer. The interviewer called the patient at the scheduled appointment time and attempted to complete the interview. The interviewer obtained verbal consent from the respondent and recorded this in the CAPI, which is how we capture consent even when in person. The interviewer sent a PDF of the showcards via email if available or read the response options to the respondent. Patients who completed the interview by phone received a $25 gift card or check.

Protocol for Interviewing Adolescents 13 to 17 Years Old

When the parent/guardian called the FS to set up an interview appointment, the FS first asked for the unique ID number printed on the flyer. This helped the FS determine that the potential respondent was an adolescent. The FS conducted an initial screening for eligibility. To be eligible, the adolescent must have received services from the health center the past 12 months. The FS also answered any questions the parent/guardian or the adolescent had. The FS informed the parent/guardian that they needed to be present for the first part of the interview and that the FS will obtain the parent's consent first before interviewing their child. The FS also let the parent and the adolescent know that they needed to be in a private space in the home where they would not be overheard by others. The parent completed their portion of the interview first and then left the room, allowing the adolescent to complete the remainder of the survey.

The FS then scheduled an appointment for the interview with the FI. Parents/guardians received a reminder call the day before the appointment. The reminder was sent via a text message only if the parent/guardian and the adolescent gave permission to do so. Texting was used only for scheduling an appointment or for appointment reminders. The FS also mentioned that the adolescent would receive a $25 gift card incentive upon completion of the interview.

Adolescents who called in to set up an interview were told that we needed their parent/guardian to set up the appointment. When the adolescent called, the FI first determined that the parent/guardian was present and available. If the parent/guardian was present, the FI administered the screener, which asked demographic and patient type (e.g., funding type of the patient) questions. Responses to these questions

not only determined the adolescent's eligibility but was also used to route the questions in the main interview. The parent/guardian had to complete the screening questions, rather than the adolescent.

If eligible, the FI confirmed that the parent and adolescent were in a private space in their home where they would not be overheard by others. A question about being in a private space was added to the instrument that the FI read to both parent and adolescent.

The FI then started the consent process, obtaining parent consent first. Parent consent included consent for self (parents are asked questions about income, health insurance, and demographics) and consent for their child. After the parent consented, the FI conducted the parent's portion of the interview. The adolescent was asked to leave so that the parent could answer the questions in private. The parent portion of the interview took about 10 minutes. After completing the interview, the FI thanked the parent for their participation. If the parent did not have time to complete these questions, the FI broke off and rescheduled for another time. If the parent refused to answer the questions, the FI advanced the survey to the adolescent portion and the parent questions were documented as nonresponse.

Before starting the assent process for the adolescent, the FI again confirmed that the adolescent was in a private area in the home. After the adolescent confirmed that they were in a private setting, the FI obtained the adolescent's assent and their permission to record portions of the interview. Even if the parent/guardian consented to the interview, the adolescent had the option to refuse to participate. Refusal conversion was not conducted with adolescent participants.

At the end of the interview, the adolescent was asked for their mailing address, which was used to mail the incentive. The adolescent's mailing address information was not combined with their survey responses. Mailing address information was deleted after 3 months following the completion of the interview.

4.4 Recruitment Materials

Each recruitment document went through extensive review and revision by RTI project staff before being finalized. Each document was also translated into Spanish, Tagalog, Vietnamese, and Chinese. Both English and other language versions of the documents were provided to the certified bilingual FIs with an anticipated high percentage of monolingual non-English speakers in their assignment. The following is a description of each document that was used.

4.4.1 Receptionist Respondent Recruitment Script

Once the receptionist determined that the patient met the basic eligibility requirements for the 2022 HCPS, they read a brief script to the patient (or to their parent or guardian for selected children). This script gave a very brief introduction to the survey and informed patients that if eligible to complete the full interview, they would be provided with remuneration valued at $25 as a token of appreciation for participating. The receptionist/respondent recruitment script was read to patients in conjunction with providing the brochure.

4.4.2 Question/Answer Brochure

The trifold brochure included an introduction to the survey and answers to some FAQs, such as why the respondent was selected, respondent consent and privacy, and purpose of the study. The brochure also contained the project's toll-free telephone number for patients to call to verify the legitimacy of the study or to ask any additional questions. The toll-free telephone number to RTI's Human Protections Office (i.e., IRB) was also included. Receptionists provided this brochure to all selected patients.

4.4.3 Interviewer Recruitment Script

Each time FIs were approached by a referred patient, they read a short script to the patient. The interviewer recruitment script briefly described the 2022 HCPS interview process. The script gave an introduction on the FI, RTI, and the purpose of the 2022 HCPS.

4.5 Oversampled Subgroups

The 2022 HCPS oversampled patients who are Asian, AI/AN, or NH/PI; veterans; and aged 65 or older. This oversample contains several subgroups with native English speakers (e.g., Hawaiians, American Indians) and patients who may be recent immigrants and have limited English skills (e.g., Hispanic or Asian participants). Our basic approach was to make every reasonable effort to include these patients in the study.

The 2022 HCPS instrument was programmed in five languages: English, Spanish, Chinese, Tagalog, and Vietnamese. Interviews were conducted in English only if respondents indicated that they spoke English "very well." If patients did not speak English very well, they were more likely to misunderstand the question and provide an incorrect answer, affecting the quality of the data. Before visiting the site, FSs determined which language a majority of the patients at the site spoke using the Site Profile Sheet. Sites with high concentrations of patients who spoke a language other than English were assigned to a bilingual interviewer. Monolingual FIs were to contact their FS immediately if they were assigned a site where a majority of the patients spoke a language other than English. About 73% of the interviews were conducted in English and 27% in Spanish. None of the interviews were conducted in Chinese, Tagalog, or Vietnamese.

The process for interviewing Spanish-speaking patients was identical to that for English-speaking patients. Consent and assent forms, brochures, and scripts were translated into Spanish. All auxiliary materials, such as showcards, were also translated. It was important that interviews be conducted uniformly, regardless of interview language used. Interpreters were not used on this survey.

In addition to language issues, cultural sensitivity practices are also important when interviewing patients from other cultures. Patients from other cultures may not be as familiar with surveys and may be more suspicious about why they were selected to be in the study. FIs took special care in explaining the study and why patients were selected and to assure them that their responses would be kept strictly confidential. FIs were sensitive to differences in culture and interacted with patients in a manner that respected these differences while maintaining a professional attitude toward the study and its administration.

As with other interviewing situations, FIs maintained a neutral interviewing style and were attuned to nonverbal cues when conducting in-person interviews, indicating that a respondent was uncomfortable with the interview questions. For phone interviews, FIs were trained to identify signs of distress by listening to verbal and nonverbal indicators such as shakiness in the voice, changes in volume, or crying. If a patient seemed uncomfortable during an interview with particular questions, FIs reminded the patient that they did not have to answer the question.

4.6 Conducting the Interview

4.6.1 Consent Procedures

Table 4-1 provides a summary of the types of permission required to conduct interviews with respondents from each age group and the type of interview conducted. Interviews were not conducted unless the permission and consent requirements had been met.

Table 4-1. Summary of Permission, by Type of Sampled Patient

Age of Sampled Patient	Permission Required	Type of Interview
0–12 years old (accompanied by parent/guardian)	Adult proxy informed consent	Proxy interview with parent/guardian
13–17 years old (accompanied by parent/guardian)	Parental permission form Youth assent form	Self-interview with adolescent
18+ years old	Adult informed consent	Self-interview with adult

The informed consent form was read aloud to each participant, either on the phone or in person. Spanish, Chinese, Vietnamese, and Tagalog versions of the consent forms were available for use by RTI certified bilingual FIs only for respondents who preferred to conduct the interview in these languages. They were also printed on different colors of paper for easy identification by the FI. For in-person interviews, a copy of the consent form was given to respondents for their records.

Once the respondent agreed to participate after being read the appropriate consent/assent form, the FI immediately attempted to begin the interview. In-person interviews were conducted at the site or at a location chosen by the respondent either before or after the respondent's medical appointment. If conducted at the site, interviews were administered in a private location, such as an unoccupied office, treatment room, or conference room.

Migrant and seasonal farm workers were encouraged to begin the interview process on site (either before or after their doctor's appointment) because it was anticipated it may be difficult for them to arrange to meet the FI at a later time and date. For safety and logistical reasons, project protocol required that all homeless respondents be interviewed at the site, either at the time of screening or at a later date. All patients were encouraged to begin the interview process immediately, but some respondents found it more convenient to schedule an appointment with the FI for a later time and date.

Table 4-2 shows the interview mode by patient type. The majority of the interviews were conducted in person (63%). About 37% of the interviews were completed by telephone.

Table 4-2. Interview Mode by Funding Type

Awardee Funding Programs	In person (health center or off site)		Telephone	
	n	%	n	%
CHC	1,606	55.7	1,279	44.3
MHC	329	69.6	144	30.4
HCH	556	93.0	42	7.0
PHPC	268	62.3	162	37.7
Total	2,759	62.9	1,627	37.1

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.

The 2022 HCPS interview was administered using a CAPI instrument. FIs read each question aloud and recorded the respondent's answers. For questions with a long list of response options, respondents who were interviewed in person were provided with a showcard from which to select their answer. For respondents interviewed by telephone, the FI read the response options to the respondent.

Once the interview was completed, all respondents received remuneration for participating: $25 cash or a gift card if conducted in person and a check or a gift card if conducted by telephone. The type of remuneration provided to the respondents was determined by the site during the recruitment phase of the

study. For proxy interviews for child respondents aged 12 and younger, the remuneration was provided to the parental/guardian who responded on behalf of the child.

4.7 Data Collection Results

4.7.1 Funding Type

The target interview goal for the 2022 HCPS was 4,500 completed interviews. The targets by funding type were 2,550 for CHC, 740 for MHC, 830 for HCH, and 380 for PHPC. **Table 4-3** provides a breakdown of each funding type's interview targets and final completion figures.

Table 4-3. Completed Interviews, by Funding Type

Funding Type	Target Interview Goal	Completed	
		Interviews	Percent of Interview Goal
CHC	2,550	2,915	114.3
MHC	740	473	63.9
HCH	830	597	71.9
PHPC	380	429	112.9
Total	4,500	4,414	98.1

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.

Based on the data from **Table 4-4**, MHC sites had the highest proportion of completed interviews with proxies for young children across the funding types. Overall, approximately 89% of interviews were completed with adults, approximately 2% of cases were completed with youth, and approximately 10% of cases were completed by proxy for young children. Interviews with youth (self) were more common among the MHC than the other population types. Interviews with youth (proxy) were more common among the MHC funding type.

Table 4-4. Distribution of Completed Interviews, by Interview Type and Funding Type

Funding Type	Adult Self (18 years and older)	Youth	
		Self (13–17 years)	Proxy (12 years and younger)
CHC	2,571 (88.2%)	50 (1.7%)	294 (10.1%)
MHC	357 (75.5%)	27 (5.7%)	89 (18.8%)
HCH	591 (99.0%)	8 (1.9%)	6 (1.0%)
PHPC	393 (91.6%)	0 (0.0%)	28 (6.5%)
Total completed interviews	3,912 (88.6%)	85 (1.9%)	417 (9.5%)

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.

4.7.2 Oversampled Subgroups

The target interview goals by oversampled subgroups were 325 for Asian participants, 335 for AI/AN participants, 100 for NH/PI participants, 550 for patients aged 65 or older, and 150 for veterans. **Table 4-5** provides a breakdown of each oversampled group's interview targets and final completion numbers. Interview goals were met for patients aged 65 and older but not for Asian participants (62.5%), AI/AN participants (91.9%), NH/PI participants (85.0%), and veteran participants (79.3%).

Table 4-5. Completed Interviews, by Oversampled Subgroup

Type	Target Interview Goal	Completed	
		Interviews	Percentage of Interview Goal
Asian	325	203	62.5
American Indian/Alaska Native	335	308	91.9
Native Hawaiian/Pacific Islander	100	85	85.0
Aged 65 or older	550	651	118.4
Veterans	150	119	79.3
Total	1,460	1,366	93.6

Table 4-6 provides the distribution of the oversampled subgroups by funding type. A majority of Asians, AI/ANs, NH/PIs, those 65 years and older and veterans received CHC funding. Next to CHC, HCH was the second source of funding for Asians (7.4%), AI/ANs (20.5%), NH/PI (5.9%) and veterans (21.9%).

Table 4-6. Distribution of Completed Interviews, by Oversampled Groups and Funding Type

Funding Type	Asians	American Indian/ Alaska Natives	Native Hawaiians/ Pacific Islanders	Aged 65 Years or Older	Veterans
CHC	175 (86.2%)	200 (64.9%)	73 (85.9%)	501 (77.0%)	78 (65.6%)
MHC	4 (2.0%)	18 (5.8%)	1 (1.2%)	34 (5.2%)	2 (1.7%)
HCH	15 (7.4%)	63 (20.5%)	5 (5.9%)	42 (6.5%)	26 (21.9%)
PHPC	9 (4.4%)	27 (8.8%)	6 (7.1%)	74 (11.4%)	13 (11.0%)
Total completed interviews	203 (100.0%)	308 (100.0%)	85 (100.0%)	651 (100.0%)	119 (100.0%)

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.

Table 4-7 shows a screening response rate of 82.9% and an interviewing response rate of 85.3%. **Table 4-8** shows final cooperation rates by funding type. Of respondents who agreed to complete the screener and who were determined to be eligible, cooperation rates ranged from a low of 78.3% for MHC patients to a high of 90.7% for HCH patients.

Table 4-7. Final Response Rate for the 2022 HCPS

Sample Category	Number	Percent, %	
		Total Referred	Confirmed Eligible Participants
Screening			
Total sample referred	6,268	—	—
Total sample approached FI	5,278	84.3	—
Approached FI but refused screener	84	1.3	—
Total completed screeners	5,194	82.9	—
Interviewing			
Ineligible cases	19	0.3	—
Eligible cases	5,175	82.6	—
Refusals, breakoffs, and other nonresponses	761	12.1	14.7
Total completed interviews	4,414	70.4	85.3

Table 4-8. Final Cooperation Rate for 2022 HCPS

Sample Category		Number	% of Eligible Participants
CHC	Confirmed eligible	3,440	—
	Refusals, breakoffs, and other nonresponses	525	15.3
	Total completed interviews	2,915	84.7
PHPC	Confirmed eligible	604	—
	Refusals, breakoffs, and other nonresponses	131	21.7
	Total completed interviews	473	78.3
HCH	Confirmed eligible	658	—
	Refusals, breakoffs, and other nonresponses	61	9.3
	Total completed interviews	597	90.7
MHC	Confirmed eligible	492	—
	Refusals, breakoffs, and other nonresponses	63	12.8
	Total completed interviews	429	87.2

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.

4.8 Quality Control Procedures

RTI employed various techniques to ensure high-quality survey data collection. The procedures put into place are highlighted below and detailed in the sections that follow.

- Performed extensive survey data reviews to ensure and validate data quality.
- Instituted practices to ensure that FIs were adhering to all project protocols.
- Used computer audio-recorded interviewing (CARI) technology to monitor FI performance and other quality measures.
- Distributed documentation to reiterate concepts that were particularly important for quality.

4.8.1 Survey Data Review

Interview data quality was monitored closely throughout data collection. RTI reviewed interview data during the testing phases of the project and while data collection was in process. Staff examined specific instrumentation characteristics and data, including the following:

- questions with a larger-than-expected proportion of "don't know," "other," "not applicable," or "refused" responses;
- routing patterns of completed cases to ensure logic accuracy and consistency;
- lengths of interview sections;
- any evidence of interviewer "shortcutting" or falsification;
- data timing to ensure that interviews were completed in an efficient and reasonable time;
- time-per-case and cost-per-case data to ensure efficiency in travel time and effectiveness in time management; and
- refusal rates for the CARI recordings.

4.8.2 Standard Administration Procedures

Strict adherence to general 2022 HCPS questionnaire administration procedures and following the CAPI instrument were two items emphasized during interviewer training and the field period to ensure high-quality data collection. Interviewers were trained to:

- Read all questions and answer choices verbatim and not suggest answers or in any way bias respondents' interpretation of a question or their answer to a question.
- Use the probing techniques described in the interviewer manual and reviewed during training to elicit accurate and complete responses and not to appear to pass judgment or agree or disagree with respondents' comments.
- Administer every item that CAPI displays and trust the CAPI program to accurately guide them through the interview and provide the appropriate question.

4.8.3 CARI Verification

CARI is a laptop computer application developed by RTI to audio record interviews. It provides a means for both verifying the interview and monitoring the quality of the interview.

Immediately after obtaining consent for the interview, the FIs obtained consent from the respondent for recording portions of the interview. If consent was provided, up to 10 sections were recorded, depending on the instrument skip patterns, totaling approximately 14 minutes and 30 seconds of each interview. In all, 91% of respondents consented to CARI recording: 4,007 of 4,447 total participants. CARI recordings were reviewed for at least 10% of each FI's completed interviews, with 567 interviews reviewed.

At the start of the data collection period, one of the first two completed interviews was reviewed for each FI in addition to one randomly selected interview within the first 10% completed. The subsequent cases reviewed were either selected randomly or chosen for review because they were completed after the most recent feedback was provided by that FI's supervisor to track performance over time.

The data collection team also monitored each FI's respondent refusal rate to the initial CARI consent. Interviewers with higher than normal CARI refusal rates were subject to increased scrutiny by RTI project staff, and FSs took action based on high percentage rates. These actions included initiating retraining or disciplinary action.

Review of CARI files also served as a useful tool for correcting mistakes made during the interview process. An interviewer performance and feedback tracking process was implemented and used throughout the data collection period. The data quality manager was responsible for providing FSs feedback from the CARI reviews that were conducted, which the FSs in turn used to provide accurate feedback to FIs. This tool was also used to track FI performance over time. In addition, the data collection team held weekly telephone meetings with the Regional Supervisors (RS) to address any issues found in CARI or data reviews and to allow the RSs to report back any problems noted by the FSs and FIs. The RSs conducted weekly calls with their FSs, and the FSs conducted weekly calls with the FIs in their region to review production, field costs, schedule, and any data quality or procedural issues. Any issues identified were addressed during these calls.

5. Data Editing and Coding

5.1 Data Cleaning and Editing

At the start of data processing, all partial interviews were flagged and removed from the data file. A total of 11 partial interviews were removed.

There were few data cleaning issues to resolve because of the CAPI program's built-in skip logic and range and consistency checks. During postprocessing, edit programs were written to evaluate the skip patterns and edit checks. This process was used to confirm the CAPI edits, resolve any residual inconsistencies, and apply codes to indicate legitimate skips. Nested questionnaire items were compared to "gate" items for confirmation of skip logic paths.

Frequency distributions for all items were reviewed to confirm that all responses were within the expected range. In addition, responses were cross-referenced to identify inconsistent data. The following are a few examples of consistency checks employed during this process:

- Extremely low weight values were cross-referenced with the height values to identify bad data. For adults, the minimum height was set at 4 feet and the minimum current weight was set at 75 pounds. For children, programmers set the minimum current weight at 3 pounds. Children were required to have a minimum weight 1 year ago of at least 1 pound.
- Years of residency in the United States was also cross-referenced with age, and a bad data code was applied when years of residency exceeded years of age.
- Date of first visit was cross-referenced with the patient's date of birth to ensure that no respondents reported a health care visit before birth.
- Current age and age at last lead blood test were compared for consistency.
- Sex was cross-referenced with pregnancy to ensure that skip patterns were effective.

Some high values for income were reported. Because of the population characteristics, staff created a ceiling for adult income at $500,000 and youth income at $100,000.

Respondents were asked about their ethnicity ("Are you of Hispanic, Latino, or Spanish origin?") and also about their race ("What race or races do you consider yourself to be?"). The tables in Volume 2 use a combination of race and ethnicity. This was done in the 2014 survey and is common in large, national surveys. In the 2022 HCPS, this was an especially complex variable to create because of the oversampling and analytical focus on certain groups (e.g., NH/PIs, AI/AN, Asians). When a person identified as one of the key groups of interest, they were assigned to that group, even if they selected multiple races or indicated they were Hispanic.

Many individuals who identify as Hispanic do not distinguish race from ethnicity. In fact, arguably fewer Americans in general make this distinction, which means that the traditional coding has lost its original, intended precision (Gonzalez-Barrea & Lopez, 2015). For example, the term "Hispanic Black" historically referred to persons from Cuba. In this survey, those who reported Hispanic Black origin were more likely to have a Black parent, rather than familial ties to Cuba. Therefore, all individuals who reported being of Hispanic origin in any way were grouped together unless they specified belonging to a racial group like NH/PI, AI/AN, or Asian.

The Race/Ethnicity variable contained the following groups: Asian, Native Hawaiian/Pacific Islander, Native American/Alaska Native, White Non-Hispanic, Black Non-Hispanic, Hispanic, Other, and Unknown. As described above, people in the Asian, Native Hawaiian/Pacific Islander, Native American/Alaska Native groups are people who did not specify that they are Hispanic.

Aside from the race and Hispanic origin variables, item nonresponse was low in this survey. There were very few instances of refused or missing data to individual questionnaire items. However, missing values for education and income were imputed to improve data analysis. Hot deck imputation was used to eliminate missing values (n = 440) for the Education variable. Education was used as an imputation class variable to impute Income. Because Federal Poverty Levels were a key interest, especially regarding the Affordable Care Act analysis, missing income values (n = 768) were unacceptable. Imputation classes were defined by the following variables: income above/below $35,000, education, race/ethnicity. The imputed values for education and income are provided on the analytic dataset. Imputed records have imputation flags for easy identification. The income variable was then used in the Federal Poverty Level categorical variable.

Finally, a review of all verbatim responses was conducted to remove any recorded information (such as name or location of the health center) that might lead to the identification of the health center or the interviewee.

5.2 Open-Ended Question Coding

Another important step in data processing was the coding of open-ended responses. This section outlines the coding procedures implemented. The code frames contained in this document were developed after analyzing the verbatim responses recorded in the open-ended question fields found throughout the survey instrument.

New codes were created and the open responses were categorized. These codes and their descriptive labels were determined by sorting the database of verbatim responses and identifying clusters of similar responses. When a cluster of at least 10 similar responses could be identified, a meaningful and descriptive label to append to the original code frame was developed.

Although every effort was made to develop additional codes that could accommodate all the responses, there was occasionally an item that may have elicited such a wide variety of responses that it could not be coded back into the existing code frame or meaningfully clustered into one of the newly created codes. In these rare instances, such responses were assigned under a more general code such as “Other.”

6. Weighting

The goal of the 2022 HCPS is to produce estimates of the characteristics of all members of the target population, not only the individuals who completed surveys. Sample weights allow the results to be extended from the survey respondents to the entire target population. Therefore, when unweighted totals or percentages are provided throughout the analysis tables, these represent the actual number or percentage of the respondents from the sample who fall into a specific category or responded in a specific way. In addition, when weighted totals or percentages are provided, these represent the estimated number or percentages of the national population from the same domain.

As part of the post-survey data processing activities, analysis weights were calculated for the 2022 HCPS data that followed the standard procedures described in Korn and Graubard (1997). The final weight for each patient consisted of eight components, and each component represented the probability for a sampling unit being selected at one sampling stage, a nonresponse adjustment, a poststratification adjustment, or other type of adjustment. These components are listed in **Table 6-1**, and each component is discussed in detail in the following sections.

Table 6-1. Summary of Patient Survey Sample Weight Components

The First Stage—Awardee Selection	
#1	Inverse Probability of Awardee Selection
#2	Adjustment for Percentage of Awardees Released
#3	Awardee Nonresponse Adjustment
The Second Stage—Site Selection	
#4	Inverse Probability of Site Selection
#5	Site Nonresponse Adjustment
The Third Stage—Patient Selection	
#6	Inverse Probability of Patient Selection
#7	Patient Nonresponse Adjustment
#8	Patient Poststratification Adjustment

6.1 Weight Component #1: Inverse Probability of Awardee Selection

Weight component #1 reflected the probability of awardee selection at the first stage of the sample design. The selection probabilities for awardees in sampling Stratum 6 were 1 because all awardees in those strata were selected (see **Table 2-9**). In other sampling strata, the selection probability for the i^{th} awardee within the h^{th} stratum was given by

$$G_{hi} = n_h \frac{S_{hi}}{\sum_i S_{hi}} ,$$

where h stands for the sampling strata (h = 1, 2, …, 7, corresponding to 7 awardee sampling strata); i is the awardee index (sequential number that is applied after each stratum is sorted) on the frame within a sampling stratum; n_h is the number of awardees selected in the h^{th} sampling stratum; and S_{hi} is the size measure, which is the number of patients served by each awardee from the 2018 UDS data. The weight component weight #1 was calculated as

$$wt1 = \frac{1}{G_{hi}}.$$

6.2 Weight Component #2: Adjustment for Percentage of Awardees Released

As shown in **Table 2-9**, 280 awardees (262 unique awardees selected) were all released to the field, thus the percentage of awardee released adjustment (*wt2*) was set to 1.

6.3 Weight Component #3: Awardee Nonresponse Adjustment

This adjustment accounted for failure to recruit an awardee and was calculated as

$$wt3 = \frac{N_h}{n_h}$$

where N_h is the number of awardees released and n_h is the number of awardees recruited in sampling stratum *h*.

6.4 Weight Component #4: Inverse Probability of Site Selection

Weight component #4 reflected the site probability of selection within an awardee for a specific funding program. The selection probability for the j^{th} site within the i^{th} awardee for funding program *f* was given by

$$C_{fij} = \begin{cases} 1, & if\ 3\ or\ fewer\ sites\ were\ all\ selected\ or \\ \dfrac{3s_{fij}}{\sum_j S_{fij}}, & if\ 3\ sites\ were\ selected\ through\ PPS\ sampling \end{cases}$$

where s_{fij} is the number of patients in site *j* within awardee *i* for funding program *f*. When all sites were selected for a funding type, the selection probability was 1. When three sites were selected through PPS sampling method, the selection probability was calculated as in the formula above. Sometimes, three sites with the largest patient volume were selected. When this happened, the selection probability was 3 divided by total number of sites for a specific funding program.

Weight component #4 was calculated as

$$wt4 = \frac{1}{C_{fij}}.$$

6.5 Weight Component #5: Site Nonresponse Adjustment

Weight component #5 accounted for failure to recruit a site within an awardee for a specific funding program. Weight component #5 was calculated as

$$wt5 = \frac{N_{fi}}{n_{fi}},$$

where N_{fi} is the number of sites selected, and n_{fi} is the number of sites recruited in i^{th} awardee for funding program *f*.

6.6 Weight Component #6: Inverse Probability of Patient Selection

Weight component #6 reflected the patient selection probability. The patient selection probability was calculated as

$$P_{fijk} = \frac{m_{fij}}{s_{fij}}$$

where m_{fij} is the number of patients selected, and s_{fij} is the estimated number of patients in the j^{th} site within the i^{th} awardee for funding program f in the survey year. We planned to estimate s_{fij} in the formula below as was done in the 2014 HCPS:

$$s_{fij} = \frac{total\ operation\ hours\ for\ the\ site\ in\ a\ year}{number\ of\ hours\ FI\ was\ in\ the\ site} \times r_{fij}$$

where r_{fij} is the estimated proportion of patients for funding type f in that site according to the number of patients the site served in the past year as reported during awardee recruitment; it was 1 if the site served only one patient type.

Because data collection was interrupted, in-person data collection was conducted only in some sites so the information on the number of patients visited collected by site receptionists was incomplete and not accurate. Thus, for the 2022 HCPS, the number of patients served for each funding type in the past year collected during the awardee recruitment was used to estimate the S_{fij}. Because the phone mode was introduced, it was difficult to track the number of patients who were selected, so m_{fij} was the number of patients who completed screening instead of number of patients selected.

Weight component #6 was calculated as

$$\text{wt6} = \frac{1}{P_{fijk}}.$$

6.7 Weight Component #7: Patient Nonresponse Adjustment

The product of weight components #1 to #6 was considered as the design-based weights (w_{fijk}).

Weight component #7 adjusted the design-based weights to account for the failure to complete a patient interview to reduce nonresponse bias. The weight component #7 was calculated

$$\text{wt7} = \sum_s w_{fijk} / \sum_r w_{fijk} \ ,$$

where s is for all screened patients and r is for respondents. The nonresponse adjustment was done at the site level for each funding type.

6.8 Weight Component #8: Patient Poststratification Adjustment

To reduce coverage bias and nonresponse bias left unaddressed after patient nonresponse adjustment in the study estimates, a poststratification adjustment was applied to the nonresponse-adjusted weights (the product of *wt1*…*wt7*) to calibrate the weight sums to patient counts derived from 1,339 eligible awardees in the 2021 UDS. The final 2021 UDS had patient counts for MHC, HCH, and PHPC funding programs. The patient counts for CHC were estimated by subtracting the patient counts of MHC, HCH, and PHPC

from overall UDS patient counts. The variables considered in the poststratification adjustment are summarized in **Table 6-2**.

Table 6-2. Proposed Variables in Poststratification

Variable	Number of Levels	Category
Census Region	4	Northeast; Midwest; South; West
Urbanicity	2	Urban; Rural
Age Group	9	0–4; 5–12; 13–19; 20–24; 25–34; 35–44; 45–54; 55–64; 65+
Race	5	White; Black; Native American/Alaska Native; Asian/Native Hawaiian and Pacific Islanders; Others
Hispanic	2	Hispanic; Non-Hispanic
Insurance Status	5	Private; Medicare; Medicaid; Public; None
Poverty Level	4	≤100% FPL; 101%–200% FPL; >200% FPL; Unknown

The poststratification adjustment factor was calculated using general exponential model (GEM; Folsom & Singh, 2000). Because of the oversampling for PHPC, MHC awardees, and awardees with concentrated patients in three race categories, there were large weights in each funding type. Large weights or extreme weights can inflate variance of estimates, so they need to be adjusted. GEM has the feature to control extreme weight while performing poststratification adjustment by applying tight bounds to the respondents with large weights. Within each funding program, the nonresponse-adjusted patient weights were defined extreme weights if they were larger than median weights + 2.5*Interquartile Range (IQR), where IQR is the difference between the 75th percentile and 25th percentile. A separate poststratification adjustment via GEM was conducted for each funding program. As a result, the sum of the poststratified weights matched the patient counts from 2021 UDS for each funding program. **Table 6-3** summarizes the variables that were controlled in the GEM.

In fitting GEM, some variables were dropped or collapsed because of a model convergence problem or because they inflated the UWE if they were included in the model. For example, nine age groups were collapsed to four levels (≤44; 45–54; 55–64; 65+) for PHPC, and five levels of insurance status were collapsed to four levels (Private + Public; Medicare; Medicaid; None) in the poststratification adjustment for HCH. The poverty variable could not be kept in any GEM model.

Table 6-3. Variable Summary in Poststratification Adjustment via GEM

Variables	CHC	MHC	HCH	PHPC
Census Region	All[a]	West; non-West[e]	All	All
Urbanicity	All	All	All	All
Age Group	0–19; 20–24; 25–34; 35–44; 45–54; 55–64; 65+[b]	0-19; 20–24; 25–34; 35–44; 45–54; 55–64; 65+	≤34; 35–44; 45-54; 55-64; 65+[f]	≤44; 45–54; 55–64; 65+[h]
Race	White; Black; Others[c]	White; Black; Others	White; Black; Others	White; non-White[i]
Hispanic	All	All	All	All
Insurance Status	All	All	Private + Public; Medicare; Medicaid; None[g]	Private + Public + None; Medicare; Medicaid[j]
Poverty Level	None[d]	None	None	None

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.
[a] All means all levels were kept in the GEM model.
[b] Nine age groups were collapsed to seven age groups.
[c] Five race categories were collapsed to three race categories.
[d] None means no level was kept in the GEM model.
[e] Northeast, Midwest, and South were collapsed.
[f] Nine age groups were collapsed to five age groups.
[g] Private and public insurance were collapsed.
[h] Nine age groups were collapsed to four age groups.
[i] Five race categories were collapsed to two race categories.
[j] Private + public insurance and no insurance were collapsed.

6.9 Final Analysis Weights

The final analysis weights (ANALWT) are the product of eight weight components described above, *ANALWT=wt1*...*wt8*. **Table 6-4** displays the distribution of the ANALWT and the nonresponse-adjusted weights and UWE for each funding program. The sum of ANALWT matched the total number of Health Center Program patients, which is approximately 29.7 million reported by all 1,339 eligible awardees in their final 2021 UDS reports.

Table 6-4. Weight Distribution

	CHC		MHC		HCH		PHPC		Overall	
Statistics	NR Adjusted Weight[a]	ANALWT[b]	NR Adjusted Weight	ANALWT	NR Adjusted Weight	ANALWT	NR Adjusted Weight	ANALWT	NR Adjusted Weight	ANALWT
N	2,915	2,915	473	473	597	597	429	429	4,414	4,414
Sum	20,585,491	27,365,945	1,167,966	849,245	897,334	608,642	1,229,253	870,108	23,879,740	29,693,940
Mean	7,062	9,388	2,469	1,795	1,503	1,020	2,865	2,028	5,410	6,727
Minimum	21	4	12	3	32	6	21	3	12	3
Median	2,921	2,870	437	549	572	429	847	511	1,715	1,628
Maximum	218,263	135,951	53,485	20,932	94,614	13,888	39,576	26,568	218,263	135,951
UWE	4.15	3.42	6.27	3.59	13.81	3.27	5.29	4.50	5.10	4.48

NOTE: CHC = Community Health Center Program; HCH = Healthcare for the Homeless Program; MHC = Migrant Health Center Program; PHPC = Public Housing Primary Care Program.
[a] NR adjusted weight is the weights before poststratification, the product of wt1*wt2*wt3*wt4*wt5*wt6*wt7.
[b] ANALWT is the final analysis weights after poststratification, the product of wt1*wt2*wt3*wt4*wt5*wt6*wt7*wt8.

The UWE shown in **Table 6-4** is defined as $(1 + [CV_{analwt}]^2)$, where $[CV_{analwt}]$ is the coefficient of variation of the ANALWT. Thus, the UWE is a measure for the variability of weights. UWE would have a value of 1 if the weights were equal. In the HCPS, the different sampling rates for awardees at the first design stage, varying numbers of selected sizes at the second design stage, different patient selection probability because of varying patient sizes, and different adjustment factors all attributed to the UWE. The UWEs of nonresponse-adjusted weights were high, particularly for HCH and MHC. After poststratification adjustment we were able to bring UWEs down for all funding types as we applied extreme weight control features in GEM.

As shown in **Table 6-4**, the weights within HCH, MHC, and HCH funding programs had relatively higher variation before poststratification adjustment; the UWE varied from 5.29 to 13.81. The UWEs were reduced after poststratification adjustment for each funding type. On average, each CHC patient represented about 9,388 patients in the CHC patient population, each MHC represented 1,795 patients in the MHC patient population, each HCH patient represented 1,020 patients in the HCH patient population, and each PHPC patient represented 2,028 patients in the PHPC population. Thus, when data were combined for all four funding programs, the weight variation was anticipated to be greater. The UWE for combined HCPS data was 4.48.

The formulas and data sources used for calculating sample weights are listed in **Table 6-5**.

Table 6-5. Description and Data Source of Terms in Formulas Calculating Sample Weights

Formula	Terms	Description	Data Source
$G_{hi}=n_h*S_{hi}/\sum_i s_{hi}$	G_{hi}	Selection probability for the i^{th} awardee within h^{th} stratum	Output from PROC SURVEYSELECT in SAS
	n_h	Prespecified number of awardees selected for the study in h^{th} stratum	RTI calculates the sampling rates and allocates awardee samples into each stratum (see example in **Table 2-9**)
	s_{hi}	Number of patients served in the year prior to the survey year in i^{th} awardee within h^{th} stratum	BPHC's 2018 UDS
	$\sum_i s_{hi}$	Total number of patients the awardees served in the year prior to the survey year in h^{th} stratum	BPHC's 2018 UDS
$C_{fij} = \begin{cases} 1, \text{ or} \\ \frac{3s_{fij}}{\sum_j S_{fij}}, \end{cases}$	C_{fij}	Selection probability for j^{th} site within i^{th} awardee for funding program f; equals to 1 if 3 or fewer sites are selected, or is calculated if 3 sites are selected using PPS	Output from PROC SURVEYSELECT in SAS, or equals to 1
	S_{fij}	Number of patients served in the year prior to the survey year from j^{th} site within i^{th} awardee for funding program f	RTI recruiters collect this information from the awardee or site in recruiting process
	$\sum_j S_{fij}$	Total number of patients served in the year prior to the survey year from all sites within i^{th} awardee for funding program f	Sum of s_{fij} within the awardee for a specific funding program

(continued)

Table 6-5. Description and Data Source of Terms in Formulas Calculating Sample Weights (continued)

Formula	Terms	Description	Data Source
$P_{fijk} = \frac{m_{fij}}{s_{fij}}$	P_{fijk}	Selection probability of patient *k* from awardee *i*, site *j* for funding program *f*	Calculate from the formula
	m_{fij}	Number of selected patients to yield n_{fij} complete interview from awardee *i*, site *j* for funding program *f*	Field interviewer keeps track of the number of selected patients sent by a receptionist for each funding program
	S_{fij}	Number of patients served in the year prior to the survey year from j^{th} site within i^{th} awardee for funding program *f*	RTI recruiters collect this information from the awardee or site in recruiting process
$wt1 = 1/G_{hi}$	$wt1$	Inverse of probability of awardee selection	Inverse of G_{hi}
$wt2 = \frac{M_h}{N_h}$	$wt2$	Percentage of awardee released adjustment, where M_h is the number of awardees selected and N_h is the number of awardees released in sampling stratum *h*	Calculate from the formula, equals to 1
$wt3 = \frac{N_h}{n_h}$	$wt3$	Awardee nonresponse adjustment, where Nh is the number of awardees released and nh is the number of awardees recruited in sampling stratum h	Calculate from the formula
$wt4 = \frac{1}{C_{fij}}$	$wt4$	Inverse of probability of site selection	Inverse of Cfij
$wt5 = \frac{N_{fi}}{n_{fi}}$	$wt5$	Site nonresponse adjustment, where N_{fi} is the number of sites selected and n_{fi} is the number of sites recruited in i^{th} awardee for funding program *f*	
$wt6 = \frac{1}{P_{fijk}}$	$wt6$	Inverse of probability of patient selection	Inverse of P_{fijk}
$w_{fijk} = wt1*wt2*wt3*wt4*wt5*wt6$	w_{fijk}	Design weights for each selected patient	Product of six design-based weight components corresponding to three selection stages
$wt7 = \frac{\sum_s w_{fijk}}{\sum_r w_{fijk}}$	$wt7$	A simple ratio nonresponse adjustment	Calculate the nonresponse adjustment within each site for a funding program
	$\sum_s w_{fijk}$	Sum of the design weights of all selected patients within a site for a specific funding program	Sum of w_{fijk} of all selected patients within a site
	$\sum_s w_{fijk}$	Sum of the design weights of completed interview within a site	Sum of w_{fijk} of completed interviews within a site
$wt8$	$wt8$	Poststratification adjustment done by each funding program; adjusts weights to BPHC's 2021 UDS total number of patients for various demographic domains	GEM developed at RTI; control totals are from BPHC's 2021 UDS
$ANALWT_{fijk}$ $wt1*wt2*wt3*wt4*wt5*wt6*wt7*wt8$	$ANALWT_{fiik}$	Final analysis weight	Product of design weight, nonresponse, and poststratification adjustments

7. Electronic Codebook for the PUF

The electronic codebook for the PUF is a PDF document containing a table of contents with the variables available on the file and the distribution of frequency of the variables. The codebook contains variable information such as the variable name, variable type (e.g., numeric, character), variable length, formatted levels of response, weighted and unweighted counts or frequencies, and weighted percentages. Only a portion of the questionnaire item is included in this document. A researcher should cross-reference this document with the questionnaire to ensure that the question matches the researcher's question of interest. There is no special software for using the PUF codebook.

7.1 Using the Data Files

The data are provided in several data file formats: SAS, SPSS, Stata, R, and ASCII. Users should download the file corresponding with their statistical software preference. Users of other software will be able to use these data files or use other software to modify the data for use. The data files have variable formats applied. Therefore, if a researcher is using the SAS data file (file with extension .sas7bdat), they must also save the SAS formats catalog (file with extension .sas7bcat) and include the following text at the top of their program to ensure that the formats are applied and the data can be read correctly:

libname loc "C:\Users\researcher\Documents\My SAS Files\Formats\"; /* This should be the file location where the user saves the formats file */

options nofmterr fmtsearch=(loc.formats); /* This will prevent format related errors and apply the formats from the location specified above in the SAS Library called "loc" */

7.1.1 Analyzing the Data—Accounting for the Complex Survey Design

As noted in **Chapter 2**, the 2022 HCPS is based on a complex survey design. This must be accounted for in any statistical analysis. For the convenience of users, some sample code with the complex design is provided below in SAS and in SUDAAN. Users wishing to use other software packages should review the code provided and the complex design description to ensure proper use.

SAS:

```
title "Example for Patient Survey 2022";

proc surveyfreq data=PS_Data;
        tables INT4a; strata
        final_strata; cluster granteeid;
        weight analwt;
    run;
```

SUDAAN:

```
proc crosstab data=indata filetype=sas design=wr deff; nest final_strata
granteeid;
weight analwt;
```

R:

```
library(survey)

ps_data <- readRDS('ruf.Rds')

ps_svy <- svydesign(
 data = ps_data,
 ids = ~GranteeID,
 strata = ~Final_strata,
 weights = ~analwt
)

svytotal(~INT4, design = ps_svy)
```

Stata:

```
Svyset [pweight=ANALWT], strata(Final_strata) vce(linearized)

Then use the svy prefix for analysis commands.
```

7.2 Alternative to Statistical Analysis with PUF Data Files

An alternative to statistical programming with the HCPS PUF data files is to use the 2022 HCPS Dashboard located on data.hrsa.gov. The dashboard is an interactive analysis tool that allows users to analyze HCPS PUF data. Users can tailor and download results by survey question and population. All dashboard estimates and confidence intervals are weighted and properly account for the complex survey design.

References

Chromy, J. R. (1981). Variance estimations for a sequential sample selection procedure. In D. Krewski, R. Platek, & J.N.K. Rao, eds. *Current topics in survey sampling*. New York: Academic Press, Inc.

Dohrmann, S., Krenzke, T., Roey, S., & Russell, J. N. (2009). *Evaluating the impact of data swapping using global utility measures*. Retrieved from https://fcsm.sites.usa.gov/files/2014/05/2009FCSM_Dohrmann_III-A.pdf

Folsom, R. E., & Singh A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *Proceedings of the Survey Research Methods Section of the American Statistical Association* (pp. 598–603). Retrieved from http://www.amstat.org/sections/srms/Proceedings, 2000 edition.

Gonzalez-Barerra, A., & Lopez, M. H. (2015, June 15). *Is being Hispanic a matter of race, ethnicity, or both?* Pew Research Center. Retrieved from https://www.pewresearch.org/fact-tank/2015/06/15/is-being-hispanic-a-matter-of-race-ethnicity-or-both/

Korn, E. L., & Graubard, B. I. (1997). *Analysis of health surveys.* New York: John Wiley & Sons.

SAS. (n.d.). SAS/STAT(R) 9.2 *User's guide, second edition*. Retrieved from http://support.sas.com/documentation/cdl/en/ormpug/59679/HTML/default/viewer.htm#optmodel.htm