2014 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

Final: April 20, 2016

HRSA
Health Resources & Services Administration

Contents

List of Tables

Number **Page**

1. Introduction

This manual provides documentation for users of the 2014 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 are 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 2014 HCPS, sponsored by the Health Resources and Services Administration (HRSA), aimed to collect data on patients who use health centers funded under Section 330 of the Public Health Service Act. Results from the Patient Survey will guide and support the Bureau of Primary Health Care (BPHC) in its mission to improve the health of the nation's underserved communities and vulnerable populations by assuring access to comprehensive, culturally competent, quality primary health care services. The Patient Survey collected data from the clients of health centers funded through four BPHC grant programs: the Community Health Center (CHC) Program, the Migrant Health Center (MHC) Program, the Health Care for the Homeless (HCH) Program, and the Public Housing Primary Care (PHPC) Program.

In addition to collecting data from health center clients in the four grant programs, the HCPS also oversampled patients within these programs who identified themselves as Asian, American Indian or Alaska Native (AIAN), or Native Hawaiian or Pacific Islander (NHPI) and patients aged 65 or older. To accommodate health center patient populations with limited English proficiency, the survey was translated and conducted in five languages. The survey was developed in English and then translated into Spanish, Chinese (Mandarin and Cantonese), Korean, and Vietnamese.

The survey is unique in its effort to capture person-level data from patients of all types of health center program grantees. With the current survey, BPHC aimed to

- gather data about the patients of the CHC, MHC, HCH, and 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; and
- gather information that will assist policymakers and BPHC 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 BPHC 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. Appropriate use of the HCPS PUF data includes

- estimating the rate of selected medical conditions (e.g., hypertension, diabetes, asthma) in the health center patient population;
- estimating the rate of a type of health care service (e.g., screening for cervical cancer, screening for elevated blood lead levels, recommendation to reduce salt intake) in the health center patient population;

- estimating the proportion of health center patients with certain characteristics (e.g., homeless health center patients who are missing teeth); and
- using a regression model to identify which factors or characteristics are associated with respondent's self-reported use of the emergency department in the last year or some other outcome.

The data cannot and should not be used to scrutinize individual grantee performance. Grantee 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.

The 2014 survey builds upon the successes of the CHC User/Visit Survey conducted in 2003, the 1995 CHC User/Visit Survey, the 2002 CHC and National Health Service Corps Site User/Visit Survey, and the 2009 Primary Health Care Patient Survey. While a trend analysis between years is not presented, 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. As mentioned earlier, the current survey included an oversample of patients who identified themselves as Asian, AIAN, or NHPI and patients 65 or older; these groups were not oversampled in the previous surveys. The oversample target was only achieved for the AIAN group.

RTI International conducted the 2014 HCPS. RTI is an independent nonprofit institute that provides research, development, and technical services to government and commercial clients worldwide. More information on RTI is available at http://www.rti.org.

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 public-use file. The next section in Chapter 1 provides discussion of the methods used to create the PUF from the restricted-use file (RUF).

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 patient care survey, 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 re-categorization, and local suppression to reduce sample uniques (i.e., a single respondent in a cell with respect to one or more identifying variables) as well as 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 socio-demographic 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 (RRMSE), 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, 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

2. Study Sample Design

The HCPS applied a three-stage sampling design to reflect a nesting structure. The first-stage sampling units were grantees, the second-stage sampling units were eligible health center sites within grantees, 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. There were 169 unique grantees recruited, and 3,965 patient interviews were completed for CHC, 1,217 for MHC, 1,230 for HCH, and 590 for PHPC. **Table 2-1** summarizes the samples at each of the three sampling stages. Grantees 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 grantees with multiple funding programs, an independent site and patient sample was selected from each funding program; therefore, recruiting 169 grantees was equivalent to selecting a sample from 306 grantees. Of those, there were 163 for CHC, 50 for MHC, 55 for HCH, and 38 for PHPC. Data were collected from all 169 recruited grantees.

Table 2-1. Three-Stage Sampling Summary for Patient Survey

Funding Program	First Stage		Second Stage			Third Stage		
	Number of Recruited Grantees[a]	Number of Participating Grantees[b]	Number of Recruited Sites[c]	Number of Participating Sites[d]	Patients Referred by Receptionist	Referred Patients Who Approached FIs	Eligible (Selected) Patients	Completed Patient Interviews
CHC	163	163	403	401	—	—	4,451	3,965
MHC	50	49	124	118	—	—	1,402	1,217
HCH	55	54	115	107	—	—	1,299	1,230
PHPC	38	37	72	69	—	—	629	590
Total[e]	169	169	521	520	11,852	10,378	7,781	7,002

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

[a] Grantees that were successfully recruited.

[b] Grantees 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 grantees had multiple funding programs at the first stage and some sites were selected for multiple patient types at the second stage.

2.1 Target Population

The HCPS included people who met the definition of a health center patient used in BPHC's Uniform Data System (UDS; i.e., people receiving face-to-face services from a CHC, MHC, HCH, or PHPC grantee and from a clinical staff member who exercises independent judgment in the provision of services[2]). Clients of grantees located within the 50 U.S. states and the District of Columbia were included; clients of grantees within U.S. territories and possessions were excluded.

Because many of the questions in the survey ask about services received in the past year, only people who received services through one of these grantees at least once in the 12 months prior to the current visit were considered eligible for the survey. This eligibility criterion was also implemented in BPHC's 2009 Primary Health Care Patient Survey, the 2002 Community Health Center Survey, and the 2003 Healthcare for Homeless Survey.

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

2.2 Target Sample Sizes

The study goal was to recruit 165 grantees and complete 6,600 interviews, among them 3,630 for the CHC funding program, 1,210 for the MHC funding program, 1,210 for the HCH funding program, and 550 for the PHPC funding program. The target sample sizes in three design domains, namely funding program, race or ethnicity, and age group, are shown in **Table 2-2**. To achieve the target sample sizes, patients of MHC, HCH, and PHPC funding types; patients of AIAN, NHPI, and Asian race groups; and patients aged 65 or older were oversampled.

Table 2-2. Target Sample Sizes for the 2014 Health Center Patient Survey

Domain	Target Sample Size	Proportion
Funding Type		
CHC	3,630	55.0%
MHC	1,210	18.3%
HCH	1,210	18.3%
PHPC	550	8.3%
Race/Ethnicity		
Hispanic	2,044	31.0%
Non-Hispanic White	1,558	23.6%
Non-Hispanic Black	1,618	24.5%
Non-Hispanic AIAN	409	6.2%
Non-Hispanic Asian	647	9.8%
Non-Hispanic NHPI	251	3.8%
Non-Hispanic Others	73	1.1%
Age Group		
0–17	2,200	33.3%
18–64	3,200	48.5%
65 or older	1,200	18.2%
Total	6,600	

NOTE: AIAN = American Indian/Alaska Native; CHC = Community Health Center Program; HCH = Healthcare for Homeless Program; MHC = Migrant Health Center Program; NHPI = 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 grantees. This section discusses the sampling frame construction, sample allocation, and sample selection procedures for the first-stage sample design.

2.3.1 Sampling Frame

To construct the sampling frame, we used the 2012 BPHC UDS (the most recent UDS data available at the time) for the first stage of selection. The UDS is compiled each year from annual data submissions by each Section 330–funded grantee. The UDS contained data on the number of patients

served, grantee characteristics (such as the types of grant funding received), state, urban or rural location,[3] and number of sites. The grantee characteristics were used in stratification.

The 2012 UDS data were collected from 1,198 grantees. Some grantees were excluded from the sampling frame, including

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

The grantee sampling frame included 1,149 eligible grantees reporting in 2012.

2.3.2 Stratification

Most grantees received CHC funding, while relatively few grantees received PHPC funding or MHC funding. Randomly selecting grantees without stratification would have resulted in very small grantee sample sizes for MHC and PHPC funding programs. To meet the target of completed interviews for each funding program, we have to complete many interviews for the PHPC and MHC funding programs, which has two implications: (1) the difficulty in recruiting enough patients from PHPC and MHC grantees within a short period of data collection because of the low number of patients in PHPC or MHC grantees and (2) the design effect[4] is inflated as the number of completed interviews per grantee increases, and consequently, the estimates will have low precision and the statistical power of comparison is reduced.

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

To achieve target sample sizes for three racial or ethnic groups—AIAN, Asian, and NHPI—we adopted substrata. These patients were not evenly distributed among all grantees. They tended to be clustered in a few grantees: on the basis of the 2012 UDS, 889 grantees (77%) had fewer than 100 AIAN patients, 1,000 grantees (87%) had fewer than 100 NHPI patients, and 650 grantees (57%) had fewer than 100 Asian patients. The 20 grantees with the highest proportion of AIAN patients accounted for 37.1% of total AIAN patients in all 1,149 grantees; the 20 grantees with the highest proportion of NHPI patients accounted for 51.4% of total NHPI patients; and the 20 grantees with the highest proportion of Asian patients accounted for 36.2% of total Asian patients. Thus, to achieve target sample sizes in three racial or ethnic categories, patient-concentrated grantees—those with more than 20% of their patients being AIAN, Asian, or NHPI—must be obtained and selected at the first-stage selection. Stratum 4 (CHC funding solely) had over 89% of such grantees, and very few such grantees were from Strata 1, 2, and 3. Therefore, to effectively select grantees with concentrated patients in three racial or ethnic categories,

---

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

[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 multi-stage cluster sample like the Health Center Patient Survey, *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 a grantee, *ICC* is the intra-cluster correlation coefficient that measures the degree of similarity among respondents within a grantee, and *UWE* measures variation in the sample weight.

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

Stratum 4 was further divided into four substrata according to whether a grantee has concentrated patients (more than 20%) in one of the three racial or ethnic categories.

Although some grantees had a high proportion of patients aged 65 or older, these older patients were distributed more evenly than the patients in the three racial or ethnic categories. The 20 grantees with the highest proportion of patients aged 65 or older only accounted for 2.04% of total patients aged 65 or older. As a result, oversampling grantees with concentrated patients aged 65 or older at the first stage of selection was not as effective as oversampling grantees with concentrated patients in the three racial or ethnic categories. Thus, we decided not to oversample grantees with concentrated patients aged 65 or older. The plan was to oversample patients aged 65 or older at the third stage of selecting patients.

In Stratum 1, the grantees with only PHPC funding have fewer patients than the grantees with multiple funding types. A probability proportional to the size (PPS) sample in Stratum 1 will yield very few PHPC-only grantees. To overcome this problem, we further divided Stratum 1 into four substrata according to the patient volume and the proportion of PHPC patients in a grantee. There were 10 final grantee strata.

2.3.3 Sample Allocation

Before selecting a grantee sample from each final stratum, we determined the grantee sample allocation for each final stratum. Oversampling grantees who received funding from PHPC, MHC, or HCH programs and grantees with concentrated patients in three racial or ethnic categories introduces more variation in sample weights, thus increasing unequal weighting effects (UWE). To minimize the variation in sample weights, we allocated the grantee sample using a nonlinear optimization procedure, OPTMODEL in SAS,[6] which minimizes the UWE with the following constraints:

- Select 165 grantees
- Complete 6,600 interviews distributed as 3,630 CHC interviews, 1,210 MHC interviews, 1,210 HCH interviews, and 550 PHPC interviews
- Complete interviews per grantee: 22 for CHC, 25 for MHC, 25 for HCH, and 15 for PHPC
- Select at least one grantee from each grantee type[7]

The optimum sample allocation to each grantee type is presented in **Table 2-3**. The grantee sample allocation to the 10 strata along with the sampling rates in each stratum are shown in **Table 2-4**. Assuming a 70% grantee recruitment rate, we selected 246 grantees. The sampling rates for Strata 1, 2, 3, 4, 5, 7, 8, and 9 are much higher than the overall sampling rate (21.4%), indicating that we oversample grantees in these strata.

[6] http://www.support.sas.com/documentation/cdl/en/ormpug/59679/HTML/default/viewer.htm#optmodel.htm

[7] Grantee type is defined according to what funding program(s) a grantee participated in or received funding from.

Table 2-3. Optimum Grantee Sample Allocation

Domain Category	Number of Grantees	Grantee Sample Allocation
Funding Program Received		
C	780	76
H	59	1
M	3	1
P	3	1
CH	117	16
CM	101	25
CP	25	11
MH	1	1
PH	6	1
CMH	26	10
CMP	4	4
CPH	19	12
CMPH	5	7
Total	**1,149**	**166***

NOTE: The optimum grantee sample allocation results in 166 grantees instead of 165 due to rounding.
C = Community Health Center (CHC) program; H = Healthcare for Homeless (HCH) program; M = Migrant Health Center (MHC) program; P = Public Housing Primary Care (PHPC) program; multiple acronyms used together indicate that funding was received from multiple programs (e.g., CMH = a grantee received CHC, HCH, and MHC funding; CMP = a grantee received CHC, MHC, and PHPC funding).

Table 2-4. Grantee Sample Allocation and Sampling Rates in Final Grantee Strata

First-Stage and Second-Stage Strata	Final Stratum	Number of Grantees in Sampling Frame	Grantee Sample Selected	Grantee Sample Released	Sampling Rate	Recruited Grantees
Stratum 1: Grantees received PHPC funding solely or in combination with other programs.						
Substratum 1.1: Grantees with < 25% PHPC patients and patient volume is < 75th percentile of the total patient volume in Stratum 1	1	31	22	21	70.0%	16
Substratum 1.2: Grantees with ≥ 25% PHPC patients and patient volume is < 75th percentile of the total patient volume in Stratum 1	2	15	15	15	100%	11
Substratum 1.3: Grantees with < 25% PHPC patients and patient volume is ≥ 75th percentile of the total patient volume in Stratum 1	3	15	15	14	93.3%	11
Substratum 1.4: Grantees with ≥ 25% PHPC patients and patient volume is ≥ 75th percentile of the total patient volume in Stratum 1	4	1	1	1	100%	1
Stratum 2: Grantees received MHC funding solely or in combination with other programs.	5	131	62	61	47.3%	47
Stratum 3: Grantees received HCH funding solely or in combination with other programs.	6	176	26	25	14.8%	22
Stratum 4: Grantees received CHC funding solely.						
Substratum 4.1: Grantees with more than 20% of AIAN patients	7	31	31	31	100%	15
Substratum 4.2. Grantees with more than 20% of Asian patients	8	16	16	16	100%	9
Substratum 4.3. Grantees with more than 20% of NHPI patients	9	10	10	10	100%	6
Substratum 4.4: All remaining grantees in Stratum 4	10	723	53	52	7.3%	31
Total		**1,149**	**251**	**246**	**21.4%**	**169**

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

2.3.4 Sample Selection

The grantees 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 grantees. In this method, the probability of a grantee being sampled is proportional to a size measure. The size

measure was the number of patients who visited the grantee for services as indicated in the 2012 UDS file. We selected a PPS grantee sample from each final stratum.

A PPS grantee sample was selected using the SAS SURVEYSELECT[8] procedure with predetermined sample allocation in **Table 2-5** for each final stratum. During the selection, in addition to the 10 strata for grantee sample selection discussed above, we sorted the sampling frame by region (Northeast, Midwest, South, or West); location type (urban or rural); and grantee size (large, medium, or small) when applying Chromy's (1981) probability minimal replacement sequential PPS selection procedure. Sorting the sampling frame by these key grantee characteristics and then applying the PPS sequential procedure induced implicit stratification according to the order of the units in a stratum. Therefore, the selected grantee samples were distributed among various regions, location types, and grantee sizes to ensure a representative grantee sample is selected.

Table 2-5. Grantee Sample Distribution by Region, Location Type, and Grantee Size

Domains	Grantee Frame		Grantee Sample	
	n	%	n	%
Region	**1,149**	**100.00**	**169**	**100.00**
Northeast	207	18.02	30	17.75
Midwest	225	19.58	25	14.79
South	405	35.25	47	27.81
West	312	27.15	67	39.64
Location Type	**1,149**	**100.00**	**169**	**100.00**
Urban	615	53.52	109	64.50
Rural	534	46.48	60	35.50
Grantee Size	**1,149**	**100.00**	**169**	**100.00**
Large	391	34.03	111	65.68
Medium	379	32.99	32	18.93
Small	379	32.99	26	15.38

Table 2-5 displays the grantee sampling frame and expected sample distribution by region, location type, and grantee size from the illustrative example. In the distribution of regions, the West has a higher proportion in the grantee sample, while the proportions of the other three regions in the grantee sample are lower compared to the grantee sampling frame. This difference is mainly due to oversampling grantees with concentrated AIAN and NHPI patients; the majority of these grantees are in the West (in Alaska and Hawaii). The grantee sample has higher proportions in urban areas compared to the grantee sampling frame; the reason for this difference is that we oversample PHPC grantees, which are mainly in urban areas. The grantee sample has lower proportions of small- and medium-sized grantees compared to the grantee sampling frame. This disparity occurs because of the PPS sampling method employed in grantee sample selection, which gives grantees with large patient volumes a better chance of being selected than grantees with small patient volumes. A best practice is to balance the sample to ensure the grantee sample represents grantees of different sizes.

[8] http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#surveyselect_toc.htm

If a grantee received funding from multiple programs, an independent site and patient sample was selected from each funding program. Thus, the 169 grantees recruited is equivalent to selecting a sample from 306 grantees (see **Table 2-1**). Of those, 163 served CHC patients, 50 served MHC patients, 55 served HCH patients, and 38 served PHPC patients. Because of low patient volume (i.e., the absence of patients in the funding program at a grantee) or language barriers, patients were not able to be selected nor interviewed at some MHC, HCH, and PHPC grantees. Data were collected from a total of 169 grantees and 303 programs (163 for CHC, 49 for MHC, 54 for HCH, and 37 for PHPC).

2.4 Second-Stage Sample Design

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

2.4.1 Sampling Frame

The preliminary 2012 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 grantee recruiters solicited grantee participation. Once a grantee was recruited and agreed to have the study conducted at its sites, recruiters worked with the grantee's administration to identify eligible sites. The following eligibility criteria were used, and the BPHC Project Officer was consulted to determine site eligibility on a case-by-case basis whenever necessary:

- The site participates in at least one of the four specific funding programs and must have been operating under the grantee for at least 1 year.
- The site is not a school-based health center.
- The site is not a specialized clinic, unless it is an OB/GYN or pediatric care clinic.
- 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 grantee:

- 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 funding type (CHC, MHC, HCH, and PHPC)
- sites with concentrated patients (more than 20%) in one of the three racial or ethnic categories (AIAN, Asian, or NHPI)

In most cases, one field interviewer (FI) was hired to collect data for each participating grantee. Therefore, selected sites must be within manageable distances for the FIs. The grantees tend to operate sites in relatively localized areas. We evaluated distances between the administrative office or 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 grantee received funding from multiple programs.

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

For grantees with more than three sites for a patient type, we used a PPS sampling method similar to the one for grantees discussed in **Section 2.3.1** 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 AIAN, Asian, and NHPI patients, we not only oversampled grantees with concentrated patients in these three racial or ethnic groups at the first stage of selection, but we also identified sites with concentrated patients in at least one of the three targeted racial or ethnic categories. These sites were selected with higher probabilities than sites without concentrated patients.

As shown in Table 2-1, 521 sites were selected. Of those, 403 served CHC patients, 124 served MHC patients, 115 served HCH patients, and 72 served PHPC patients. Because of low patient volumes, language barriers, or extensive traveling, patients were not able to be selected nor interviewed at some MHC, HCH, and PHPC sites. Data were collected from 520 sites, including 401 for CHC, 118 for MHC, 107 for HCH, and 69 for PHPC.

2.5 Third-Stage Sample Design

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. An FI visited a selected site for a predetermined number of days and time slots in the sampling period to conduct interviews.

2.5.1 Patient Interview Allocation to Grantees

To achieve the near self-weighting sample of patient interviews within each grantee stratum, the same number of patient interviews was desired from the grantees in each funding program. The interview quota for each grantee was determined by evenly allocating the targeted number of completed interviews to all participating grantees for a funding program, then inflating this target number to produce a production goal. The production goal assigned to each grantee was slightly inflated since some grantees were anticipated to have difficulty in achieving the goal because of low patient volumes, particularly for MHC, HCH, and PHPC grantees. By doing so, the grantees with high patient volumes could compensate for production challenges faced by the low-volume grantees. **Table 2-6** shows the quota per grantee for each funding type.

Table 2-6. Patient Interview Quota per Grantee

Funding Program	Patient Interview Quota/Grantee
CHC	23
MHC	28
HCH	28
PHPC	18

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

2.5.2 Patient Interview Allocation to Sites within a Grantee

Within each grantee, we used different methods to allocate patient interviews to multiple sites for grantees with three or fewer sites in a funding program and grantees with more than three sites in a funding program. For grantees with three or fewer sites, the number of patient interviews within that grantee 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 jth site from a grantee for funding program *f*. For grantees with more than three sites that were selected through PPS, the number of selected patients were divided equally among three selected sites. Doing so will help to reduce the UWE.

2.5.3 Patient Screening and Selection

To oversample patients in the three racial or ethnic categories and patients aged 65 or older, we designed a screening sheet that receptionists could use to screen and select patients when a patient entered the site and registered for service. Patients would be considered eligible if they had received service through one of the grantees supported by BPHC 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 or 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 he or she was 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 sent that patient to the FI if they were available. If the FI was working on an interview or was unavailable, the receptionist gave the selected patient a yellow laminated card and instructed them to wait in a designated area. When the FI was available and ready, the FI would look for a person holding a yellow laminated card. However, this screening and oversampling procedure was not approved by RTI's institutional review board (IRB) and the grantees. Thus, no screening and oversampling were implemented at the patient level. Instead, for all patients, we asked the receptionist to select the first eligible patient registered once the FI informed that he or she was ready for the next interview.

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 using the sign-in sheet or appointment list before the FI left the site. The patient count sheets for each FI data

collection visit were sent to RTI for data entry, and counts were used to calculate the analysis weights for the study (see **Section 6.1** 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.

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.

As shown in Table 2-1, 7,781 patients were classified as eligible (selected): 4,451 for CHC, 1,402 for MHC, 1,299 for HCH, and 629 for PHPC, and 7,002 patient interviews were completed: 3,965 for CHC, 1,217 for MHC, 1,230 for HCH, and 590 for PHPC. **Table 2-7** displays the patient sample distribution. The patient sample had a higher proportion in the West and a lower proportion in the Northeast. The difference of the proportion in region was corrected by poststratification in calculating sample weights. The patient sample distributions for location type and grantee size were very similar to the patient population.

The patients' selection method described previously was not entirely random. A selection method that selects every nth patient for an interview was not feasible for those sites in which the volumes of targeted patients were very low (such as a site serving multiple types of patients, including a very low volume of public-housing patients or homeless patients). At a low-volume site, sampling patients at an interval could result in wait times of several days before the nth patient arrived.

The patient data collection started on October 8, 2014, and ended on April 17, 2015. The sampling period covered more than 6 months, which is a reasonable sampling window; however, data collection time spent at each grantee to achieve a predetermined grantee patient quota was different, varying from 4 to 45 days, with 17 days as an average. Because of the short data collection time in some grantees, it was likely that the patient sample in the study could overrepresent certain types of patients, such as patients with seasonal flu, or miss certain types of patients, such as patients who visited the site regularly at a longer interval than the data collection period. Consequently, patient characteristics of the patient sample, such as age, race or ethnicity, gender, and medical condition, might be different from the patient population. Some patient characteristics can be corrected in the poststratification adjustment of sample weight calculation, such as age, race or ethnicity, and gender (discussed in **Section 6.1.8**). However, whether a difference is present and the magnitude of any difference remains unexamined for some patient characteristics (e.g., patient medical conditions).

Table 2-7. Patient Sample Distribution by Region, Location Type, and Grantee Size

Domains	Patient Population[a]		Patient Sample	
	n	%	n	%
Region	**20,602,711**	**100.0%**	**7,002**	**100.0%**
Northeast	4,207,695	20.4%	1,169	16.7%
Midwest	3,883,679	18.9%	884	12.6%
South	6,263,148	30.4%	2,007	28.7%
West	6,248,189	30.3%	2,942	42.0%
Location Type	**20,602,711**	**100.0%**	**7,002**	**100%**
Urban	13,408,183	65.1%	4,750	67.8
Rural	7,194,528	34.9%	2,252	32.2
Grantee Size	**20,602,711**	**100.0%**	**7,002**	**100.0%**
Large	14,656,050	71.1%	5,202	74.3%
Medium	4,430,778	21.5%	1,075	15.3%
Small	1,515,883	7.4%	725	10.4%

[a] Patient population was based on the patient counts from 1,034 grantees in the grantee sample frame in the preliminary 2012 Uniform Data System.

3. Data Collection Instruments

3.1 Questionnaire Development

BPHC, a Technical Advisory Committee (TAC), and RTI collaborated on the development of the HCPS instrument. The instrument builds on the previous periodic User Surveys, which provided valuable information on the process and outcomes of care in CHC and HCH programs. In addition, the HCPS included interviews of patients drawn from migrant populations and residents of public housing, populations included in the 2009 Patient Survey. The original questionnaires for the previous patient surveys drew heavily from questions in the NHIS conducted by National Center for Health Statistics (NCHS). Conformance with the NHIS allowed comparisons between these NCHS surveys and the previous health center patient surveys. The current HCPS, and the 2009 Patient Survey, included questions from not only the NHIS but also from the Medical Expenditure Panel Survey, the National Health and Nutrition Examination Survey, and numerous other surveys. Thus, comparisons between the HCPS results and current national survey data are possible, in addition to previous Patient Survey data.

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, and
- perceived quality of care.

The instrument was modified per the cognitive testing outcomes; translated from English into Spanish, Chinese, Korean, and Vietnamese; and programmed to create five language versions of the computer-assisted personal interviewing (CAPI) instrument. The rationale for translation into Chinese, Korean, and Vietnamese was due to the oversampling of these populations and to ensure that language barriers would not limit completion of these interviews.

3.2 Questionnaire Modules

To meet the Patient Survey research goals, the final HCPS instrument included 18 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, 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. Note that the public use file 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 Health Center Patient Survey Instrument

Module	Topic	Number of Questions	Types of Questions
S	Patient screening	10	1. Age category 2. Gender 3. Race 4. Eligibility questions (health services received in the past 12 months, farm work, homelessness, public housing status) 5. Consent for interview and audio recording
A	Introduction	14	1. Date of birth 2. Gender 3. Language spoken 4. Race
B	Access to care	10	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	40	1. Health care providers seen in past 12 months 2. Vaccinations received 3. Reasons why the patient has not received a recent checkup, etc.
D	Conditions	103	1. Height and weight 2. Weight management/exercise 3. Medical history and conditions (pregnancy, high blood pressure, hepatitis, tuberculosis, asthma, diabetes, cancer, hearing, vision, and other health conditions)
E	Conditions follow-up	69	1. More specific questions on care received for health conditions (high blood pressure, asthma, diabetes, and other health conditions)
F	Cancer screening	71	1. Cancer screening services received (Pap smear, human papilloma virus, mammogram, colonoscopy/sigmoidoscopy exam, blood stool test)
G	Health center services	69	1. Usual source of care 2. Referrals 3. Language assistance received 4. Help received to access social programs 5. Health center services experience and satisfaction with a wide range of health center characteristics

(continued)

Table 3-1. Description of Health Center Patient Survey Instrument (continued)

Module	Topic	Number of Questions	Types of Questions
H	Health insurance	42	1. Current insurance coverage 2. Reasons for lack of insurance 3. Coverage provided by insurance
I	Prescription medication	15	1. Prescription services experience 2. Ease of getting prescription 3. Satisfaction level
J	Dental	46	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	51	1. Questions about feelings 2. Reason for inability to get or delay in getting mental health care 3. Where mental health treatment or counseling was received (prescription medication, group or individual counseling, inpatient treatment, etc.)
L	Substance use	116	1. Use of substances (cigarettes, alcohol, illicit drugs), substance abuse treatment
M	Prenatal care/ family planning (females aged 15–49)	22	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	16	1. Whether HIV test was received; if not, reason why
O	Living arrangements	16	1. Type of place where living currently 2. Crowding 3. Extent of past homelessness experience
Q	Income and assets	13	1. Family income 2. Receipt of income support and public assistance
R	Demographics	37	1. Birth place 2. Education 3. Sexual orientation 4. Marital status 5. Veteran status 6. 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 Grantee Recruitment

Recruitment of the sampled grantees and sites began in March 2014. The recruiting activity included making contact with each sampled grantee 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.

Advance packages were sent to the selected grantees, which marked the start of the recruiting process. The first contact with the grantee office was a brief telephone call in which the recruiters confirmed the name, title, and mailing address for the grantee chief executive officer or decision maker who would receive the introductory packet. The recruiters also captured the telephone number and e-mail address of the grantee contact.

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

Recruiters obtained permission or approval from all applicable grantee administrators and applicable review boards, such as local IRBs, prior to any grantee's participation. Each recruiter assisted, as appropriate, in preparing for the local review process. The grantees were provided with all the documentation required for study review and approval.

During the grantee recruitment process, a few grantees were found to be ineligible or refused to participate. Ineligibility occurred when the grantee 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. The most common reasons cited for refusal were privacy concerns and key decision makers' perception of an excessive burden on grantee and site staff and resources. The recruiters made final contact with the grantees after the sites were selected and indicated which sites had been selected and the patient interview quota at those sites.

4.2 Site Recruitment

The grantee organization and each associated site were usually recruited independently; that is, some sites required permissions and approvals beyond those obtained at the grantee level. Site-level contacts were sent an advance package that contained the same materials as the one sent to the grantees.

During the initial and follow-up telephone calls, the recruiter identified requirements and approvals for participation at the site level, potential barriers to participation, and critical information required by the data collection team.

After *all approvals* were obtained at the site level, recruiters initiated an letter of approval (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.

Once site cooperation was secured, RTI recruiters arranged and conducted site staff phone training, which was held prior to the start of data collection at the site. The training agenda included an

overview of the study and of RTI, the number of interviews needed, the roles of the primary and secondary contacts, and the role of the site receptionist.

4.3 Respondent Recruitment

Data collection for the HCPS was initiated in September 2014 and concluded in April 2015. The HCPS involved a face-to-face interview that was administered by trained FIs. A parent or guardian completed a proxy interview if the sampled patient was 12 years of age or younger. Sampled patients aged 13 years or older were interviewed directly.

Respondent selection was conducted through on-site sampling. Although the FIs were told which 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 at the site, 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 Patient Survey (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 he or she was 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 his or her parent or guardian, for selected children) and give him or her a copy of the HCPS brochure.

If the selected patients were interested in participating in the 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 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 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 upon a location and time to meet and complete the interview.

The Patient Survey instrument was programmed in five languages: English, Spanish, Chinese, Korean, 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, Field Supervisors (FS) 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 68% of the interviews were conducted in English, 28% in Spanish, 3% in Chinese, and less than 1% in Vietnamese. None of the interviews were conducted in Korean.

The process for interviewing non-English-speaking patients was identical to that for English-speaking patients. Consent and assent forms, brochures, and scripts were translated into Spanish, Chinese,

Korean, and Vietnamese. All auxiliary materials, such as showcards, were also translated into these languages.

4.4 Conducting the Interview

Once patients were sampled and recruited into the Patient Survey, the next step was to administer informed consent. There were three different types of interviews: (1) a proxy interview with a parent or guardian for a child 0–12 years of age, (2) a self-interview with an adolescent 13–17 years of age, and (3) a self-interview with an adult at least 18 years of age.

The informed consent form was read aloud to each participant. Spanish, Chinese, Vietnamese, and Korean versions of the consent forms were available for use by RTI certified bilingual FIs only for sample members who preferred to conduct the interview in these languages. 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. 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. This was especially true if the respondent's medical appointment at the site was urgent.

The Patient Survey 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 were provided with a showcard from which to select their answers.

Once the interview was completed, all respondents received remuneration for participating: $25 cash or a cash equivalent (Visa, Walmart, or Target gift card or gift certificate for grocery or discount store [food voucher]). 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.

4.5 Data Collection Results

The target interview goal for the Patient Survey was 6,600 completed interviews. The targets by funding type were 3,630 for CHC, 1,210 for MHC, 1,210 for HCH, and 550 for PHPC. **Table 4-1** provides a breakdown of each funding type's interview targets and final completion figures.

Table 4-1. Completed Interviews, by Funding Type

Funding Type	Target Interview Goal	Interviews Completed	Percentage of Interview Goal Completed
CHC	3,630	3,965	109.2%
MHC	1,210	1,217	100.6%
HCH	1,210	1,230	101.7%
PHPC	550	590	107.3%
Total	**6,600**	**7,002**	**106.1%**

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

The target interview goals by oversampled subgroups were 647 for Asians, 409 for AIANs, 251 for NHPIs, and 1,200 for patients aged 65 or older. **Table 4-2** provides a breakdown of each oversampled group's interview targets and final completion numbers. Interview goals were met for AIANs (163.8%) but not for Asians (69.7%), NHPIs (52.6%), and patients aged 65 and older (46%).

Table 4-2. Completed Interviews, by Oversampled Subgroup

Type	Target Interview Goal	Interviews Completed	Percentage of Interview Goal Completed
Asian	647	386	59.7%
American Indian/Alaska Native	409	670	163.8%
Native Hawaiian/Pacific Islander	251	132	52.6%
Aged 65 or older	1,200	552	46.0%
Total	2,507	1,740	69.4%

Table 4-3 shows an interviewing response rate of 91.4%. **Table 4-4** 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 87% for MHC patients to a high of 96% for HCH patients.

Table 4-3. Final Response Rate for Patient Survey

Sample Category	Number	% of Confirmed Eligibles
Interviewing		
Ineligible cases	2,396	—
Eligible cases	7,659	—
Refusals, breakoffs, and other nonresponses	674	8.8%
Total completed interviews	7,002	91.4%

Table 4-4. Final Cooperation Rate for Patient Survey

Sample Category		Number	% of Eligibles
CHC	Confirmed eligible	4,470	—
	Refusals, breakoffs, and other nonresponses	505	11.3%
	Total completed interviews	3,965	88.7%
PHPC	Confirmed eligible	629	—
	Refusals, breakoffs, and other nonresponses	39	6.2%
	Total completed interviews	590	93.8%
HCH	Confirmed eligible	1,283	—
	Refusals, breakoffs, and other nonresponses	53	4.1%
	Total completed interviews	1,230	95.9%
MHC	Confirmed eligible	1,399	—
	Refusals, breakoffs, and other nonresponses	182	13.0%
	Total completed interviews	1,217	87.0%

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

4.6 Quality Control Procedures

RTI employed various techniques to ensure high-quality survey data collection. First, interview data quality was monitored closely through data examination. 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

- 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;
- timing data 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 computer-assisted recorded interview (CARI) recordings.

Second, CARIs were used to verify and monitor the quality of the work of the field 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 nine sections were recorded, depending on the instrument skip patterns.

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.

Third, FSs conducted telephone verifications with each site in their region using a standardized project-approved script. This verification required the FSs to call each site contact within a few days of the start of data collection at the site. The purpose of this call was to gather some quick feedback on the FI's performance and the overall pre-visit planning process of the Patient Survey. In particular, the FSs obtained the site staff's perceptions of interviewer demeanor, behavior, and attendance in addition to obtaining any other relevant feedback. During this call, the FSs also obtained further details on any issues that might have occurred at the site.

No incidents of falsification were discovered through the CARI or call verification processes.

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. Seven 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 post-processing, edit programs were written to duplicate 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. In the case of ambiguities (e.g., nested items that should be blank but contained a value), the gate question was treated as the accurate response and the nested items were recoded as legitimate skips.

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 one pound.
- Years of residency in the United States were 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.
- Gender 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 for youth income at $100,000.

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." During the data disclosure process some response categories with low counts were combined into "other" or dropped from the PUF.

5.3 Constructed Variables

Very few constructed or created variables are available in the HCPS PUF to avoid confusion around their construction, meaning, or use. The exceptions are *final_race*, *insured*, *uninsured, Agecat, Edit_gen, Education, Urban, Numper, and FPL*. These variables are created from questionnaire items and are described below.

Final_race is a combination of race and ethnicity. This uses question INT1a (Are you of Hispanic, Latino, or Spanish origin?) in combination with INT2 (What race or races do you consider yourself to be. You may select one or more.) If the response to INT1a was yes for being Hispanic, that respondent's value for *final_race* was Hispanic. If the response to INT1a was no, the race question (INT2) was considered. While respondents were allowed to select one or more race, respondents who selected multiple races were coded into Non-Hispanic Other. This group also contains race groups (e.g. Native Hawaiian) with too few respondents small to stand alone for confidentiality. This is the same process used for other prominent national surveys such as the National Survey on Drug Use and Health (NSDUH).

Insured and *uninsured* were coded based on the questions INS2 through INS12. If a respondent was uninsured for 6 or more months in the past 12-month reference period, that person was considered uninsured. The reverse is true: If a respondent was insured for 6 or more months in the 12-month reference period, they were considered insured. These variables are legitimately skipped for some respondents. Youth between the ages of 13-17 were not asked the questions in the Insurance Module.

Agecat is a 5-level variable grouping respondents into age categories. This is based on the recoded continuous age variable (INTAGE_R).

Edit_gen is a two-level variable indicating the respondent's biological sex a birth. This variable was edited for transgender respondents using INT3 and INT3_OTH or INT3_SPEC, if needed.

Education is a 5-level variable, including missing values. The questionnaire item *DMO4_R* is the recoded version of "What is the highest grade or year of school you completed?" *Education* uses that to collapse into fewer categories.

Urban is a two-level variable indicating if the interview took place in an urban or in a rural location. This was based the HRSA Uniform Data System (UDS) location information for the grantee. This may not be the same as the actual health center location where the patient interview occurred. All other location data is removed from the PUF.

Numper indicates the number of people living in the respondent's household. This is a 5-level variable based on the INC1c or INC1d variables.

FPL represents the federal poverty status of the respondent. This is a seven-level variable with cut-points relating to the health center patient population. The variable is constructed using income information, which is not provided on the PUF, the number of people supported by that income, and the location (state) of the household.

6. Weighting

6.1 Weighting

The goal of the 2014 HCPS is to produce estimates of the characteristics of all members of the health center patient population, not just the individuals who completed surveys. Sample weights allow the results to be extended from just the survey respondents to the entire target population. Therefore, throughout the analysis tables, when unweighted Ns or percentages are provided, 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 Ns 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 2014 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 list 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—Grantee Selection	
#1	Inverse Probability of Grantee Selection
#2	Adjustment for Percentage of Grantees Released
#3	Grantee 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

6.1.1 Weight Component #1: Inverse Probability of Grantee Selection

Weight component #1 reflected the grantees' probability of selection at the first stage of the sample design. The selection probabilities for grantees in sampling Strata 2, 4, 7, 8, and 9 were 1 because all grantees in those strata were selected (see Table 2-9). In other sampling strata, the selection probability for the i^{th} grantee 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, …, 10, corresponding to 10 grantee sampling strata); i is the grantee index (sequential number that is applied after each stratum is sorted) on the frame within a sampling stratum; n_h is the number of grantees selected in the h^{th} sampling stratum; and S_{hi} is the size measure, which is the number of patients served by each grantee from the 2012 UDS data. The weight component weight #1 was calculated as

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

6.1.2 Weight Component #2: Adjustment for Percentage of Grantees Released

As shown in Table 2-9, 251 grantees were selected in two batches. We selected 215 grantees in the first batch and anticipated that we could recruit 165 grantees assuming an 80% of recruitment rate. During the grantee recruitment, we found the recruitment rate was lower than 80%, and we selected extra 36 grantees in the second batch. Among 36 grantees we released 31 to the field. Weight component #2 accounted for the percentage of grantees released adjustment, and it was calculated as

$$wt2 = \frac{M_h}{m_h}$$

where M_h is the number of grantees selected and N_h is the number of grantees released in sampling stratum h.

6.1.3 Weight Component #3: Grantee Nonresponse Adjustment

Weigh component #3, the grantee nonresponse adjustments accounted for failure to recruit a grantee, and was calculated as

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

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

6.1.4 Weight Component #4: Inverse Probability of Site Selection

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

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

where s_{fij} is the number of patients in site j within grantee i for funding program f. When three sites with the largest patient volume were selected, the selection probability was 3 divided by total number of sites for a specific funding program.

The weight component #4 was calculated as

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

6.1.5 Weight Component #5: Site Nonresponse Adjustment

Weight component #5 accounted for failure to recruit a site within a grantee for a specific funding program and 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} grantee for funding program f.

6.1.6 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} grantee for funding program f in the survey year. s_{fij} was estimated in the formula below:

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

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 grantee recruitment, it was 1 if the site served only one patient type.

The weight component #6 was calculated as

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

6.1.7 Weight Component #7: Patient Nonresponse Adjustment

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

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

where s is for all selected patients and r is for respondents.

6.1.8 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 the final 2013 UDS for 1,155 grantees included in the grantee sampling frame. While the sample was selected based on the 2012 UDS, the poststratification used the most updated information, which was the 2013 UDS. The final 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**.

The poststratification adjustment factor was calculated using general exponential model (GEM; Folsom & Singh, 2000). Because of the oversampling for PHPC, MHC grantees, and grantees 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, they need to be adjusted. GEM has the

feature to control extreme weight while performing poststratfication 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 + 3*Interquartile Range (IQR), where IQR is the difference between the 75 percentile and 25 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 2014 UDS for each funding program. **Table 6-3** summarizes the variables that were controlled in the GEM.

Table 6-2. Proposed Variables in Poststratification

Variable	Number of Levels	Category
Census Region	4	Northeast; Midwest; South; West
Location Type	2	Urban; Rural
Age Group	9	0–4; 5–12; 13–19; 20–24; 25–34; 35–44; 45–54; 55–64; 65+
Gender	2	Male; Female
Race	5	White; Black; Native American/Alaskan; 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

NOTE: FPL = federal poverty level.

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

Variables	CHC	MHC	HCH	PHPC
Census Region	All[a]	All	All	All
Location Type	All	All	All	All
Age Group	All	0-19; 20-64; 65+[c]	≤44 vs. 45+[e]	0-19; 20-64; 65+
Gender	All	All	All	All
Race	All	White; Black; Others[d]	All	White; Black; Others
Hispanic	All	All	All	All
Insurance Status	All	None	None	Insured vs. Uninsured[f]
Poverty Level	None[b]	None	None	None

NOTE: CHC = Community Health Center Program; HCH = Healthcare for 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] None means no level was kept in the GEM model.
[c] Nine age groups were collapsed to three age groups.
[d] Five race categories were collapsed to three race categories..
[e] Nine age groups were collapsed to two age groups.
[f] Five levels of insurance status were collapsed to two levels, with insurance vs. without insurance.

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 two levels (≤44 and 45+) for HCH, and five levels of insurance status were collapsed to two levels (insured and uninsured) in the poststratification adjustment for PHPC. There were many patients having "unknown" poverty level in the data; thus, the poverty variable could not be kept in any GEM model.

6.1.9 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 21.2 million reported by all 1,155 eligible grantees in their final 2013 UDS reports.

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 the weights, and would have a value of one if the weights were equal. In the HCPS, the different sampling rates for grantees 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 for CHC, HCH, and PHPC as CHC, HCH, and PHPC grantees were selected from various sampling strata, while MHC grantees were concentrated in fewer sampling strata as shown in Table 2-9. After poststratification adjustment we were able to bring UWEs down for all funding types as we applied extreme weight control features in GEM.

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	3,965	3,965	1,217	1,217	1,230	1,230	590	590	7,002	7,002
Sum	19,905,655	20,923,779	610,360	753,081	534,837	522,869	538,061	211,925	21,588,913	22,411,654
Mean	5,020	5,277	501	619	435	425	912	339	3,083	3,201
Minimum	33	7	4	4	4	1	10	3	4	1
Median	1,621	1,625	189	258	130	153	270	236	669	609
Maximum	267,763	54,573	5,426	9,326	33,422	4,321	29,634	3,637	267,763	54,573
UWE	8.08	3.61	3.56	3.16	15.67	3.46	10.52	2.52	12.28	5.60

NOTE: ANALWT = final analysis weights; CHC = Community Health Center Program; HCH = Healthcare for Homeless Program; MHC = Migrant Health Center Program; NR = nonresponse; PHPC = Public Housing Primary Care Program.

[a] NR adjusted weight is the weights before poststratfication, 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.

As shown in **Table 6-4**, the weights within each funding program had relatively smaller variation; the UWE varied from 2.52 to 3.61. On average, each CHC patient represented about 5,277 patients in the CHC patient population, each MHC patient represented 619 patients in the MHC patient population, each HCH represented 425 patients in the HCH patient population, while each PHPC patient represented 339 patients in PHPC population. Thus, when data were combined for all four funding programs, the weight variation was anticipated to be greater. The UWE for combined Patient Survey data was 5.60.

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
$G_{hi} = n_h \frac{S_{hi}}{\sum_i S_{hi}}$	G_{hi}	Selection probability for the i^{th} grantee within h^{th} stratum	Output from PROC SURVEYSELECT in SAS
	n_h	Prespecified number of grantees selected for the study in h^{th} stratum	RTI calculates the sampling rates and allocates grantee 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} grantee within h^{th} stratum	BPHC's 2012 UDS
	$\sum_i S_{hi}$	Total number of patients the grantees served in the year prior to the survey year in h^{th} stratum	BPHC's 2012 UDS
$C_{fij} = \begin{cases} 1, & or \\ \frac{3s_{fij}}{\sum_j s_{sij}} \end{cases}$	C_{fij}	Selection probability for j^{th} site within i^{th} grantee 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} grantee for funding program f	RTI recruiters collect this information from the grantee 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} grantee for funding program f	Sum of s_{fij} within the grantee for a specific funding program
$P_{fijk} = \frac{m_{fij}}{S_{fij}}$	P_{fijk}	Selection probability of patient k from grantee i, site j for funding program f	Calculate from the formula
	m_{fij}	Number of selected patients to yield n_{fij} complete interview from grantee 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} grantee for funding program f	RTI recruiters collect this information from the grantee or site in recruiting process
$wt1 = 1/G_{hi}$	$wt1$	Inverse of probability of grantee selection	Inverse of G_{hi}
$wt2 = \frac{M_h}{N_h}$	$wt2$	Percentage of grantee released adjustment, where M_h is the number of grantees selected and N_h is the number of grantees released in sampling stratum h	Calculate from the formula
$wt3 = \frac{N_h}{n_h}$	$wt3$	Grantee nonresponse adjustment, where N_h is the number of grantees released and n_h is the number of grantees recruited in sampling stratum h	Calculate from the formula

(continued)

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

Formula	Terms	Description	Data Source
$wt4 = 1 / C_{fij}$	$wt4$	Inverse of probability of site selection	Inverse of C_{fij}
$wt5 = \frac{N_{fi}}{n_{fi}}$	$wt5$	Site nonresponse adjustment, where N_{fi} is the number of sites selected and n_{fi} is the number of sites recruited in i^{th} grantee for funding program f	
$wt6 = 1/P_{fijk}$	$wt6$	Inverse of probability of patient selection	Inverse of P_{fijk}
$w_{fijk} = wt_1 * wt_2 * wt_3 * wt_4 * wt_5 * wt_6$	w_{fijk}	Design weights for each selected patient	Product of six design based weight components corresponding to three selection stages
$wt_7 = \sum_s w_{fijk} / \sum_r w_{fijk}$	wt_7	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
	$\sum_r W_{fijk}$	Sum of the design weights of completed interview within a site	Sum of w_{fijk} of completed interviews within a site
wt_8	wt_8	Poststratification adjustment done by each funding program; adjusts weights to BPHC's 2013 UDS total number of patients for various demographic domains	GEM developed at RTI; control totals are from BPHC's 2013 UDS
$ANALWT_{fijk} = wt_1 * wt_2 * wt_3 * wt_4 * wt_5 * wt_6 * wt_7 * wt_8$	$ANALWT_{fijk}$	Final analysis weight	Product of design weight, nonresponse, and poststratification adjustments

7. Electronic Codebook for the Public Use File

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 the question matches the researcher's question of interest. There is no special software for using the PUF codebook. The codebook is Section 508–compliant.

7.1 Using the Data Files

The data is provided in three data file formats: SAS, SPSS, and Stata. Users should download the file corresponding with their statistical software preference. Users of other software, such as R, 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 catalogue (file with extension .sas7bcat) and include the following text at the top of their program to ensure 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 2014 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 2014";
    proc surveyfreq data=PS_Data;
        tables INT4a;
        strata VESTR;
        cluster VEREP;
        weight analwt;
    run;
```

SUDAAN:

```
proc crosstab data=indata filetype=sas design=wr deff;
nest VESTR VEREP;
weight analwt;
```

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 Patient Survey dashboard located on the Bureau of Primary Health Care's website: http://www.bhpc.hrsa.gov. Use this website to navigate to or search for the dashboard where users can select Patient Survey outcomes, demographic groups, and filter down to populations of interest. Estimates and bar charts are presented and users may also choose to see confidence intervals. All estimates and confidence intervals are weighted and properly account for the complex survey design. This type of point-and-click analysis may be especially helpful to health center managers, policy makers, students, and the general public.

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.

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 November 12, 2002, from http://www.amstat.org/sections/srms/Proceedings, 2000 edition.

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

Passel, J. S. & Cohn, D. (2011). *Unauthorized Immigrant Population: National and State Trends, 2010*. Washington, DC: Pew Hispanic Center.