# Primary Health Care Patient Surveys

Statistical Design Plan

# Table of Contents

# List of Exhibits

# 1. Introduction

The Primary Health Care Patient Surveys (PHCPS), sponsored by the Health Resources and Services Administration (HRSA), aim to collect data on patients who use health centers funded under Section 330 of the Public Health Service Act. Results from the Patient Surveys 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 service. The Patient Surveys will collect data from the clients of health centers funded through four BPHC grant programs: the Community Health Center Program (CHC), the Migrant Health Center Program (MHC), the Health Care for the Homeless Program (HCH), and the Public Housing Primary Care Program (PHPC).

To this end, the BPHC funded two contracts:

- The Community Health Center Patient Survey (CHC) contract will collect and analyze data from clients of the CHC program. This program serves low-income individuals.

- The Health Center Special Populations Patient Survey (Special Populations) contract will collect and analyze data from clients of the MHC, HCH, and PHPC. Respectively, these three programs serve migrant workers, homeless individuals, and residents of public housing.

Because some of the Section 330-funded health center grantees (grantees) receive grants through more than one of the aforementioned grant programs, extensive coordination between the two contracts will create efficiencies that will allow for larger sample sizes and ensure consistency between the two studies. Therefore, the sample design for the CHC and Special Populations studies reflects the decision by BPHC to coordinate these two studies using a harmonized sampling and data collection approach.

In the PHCPS, the primary analytic units are patients who receive services from the funded grantees. The primary analytic units are clustered within the health center sites (sites) within a grantee. Because most of the grantees operate more than one site, the sites are clustered within the grantees. RTI International[1] will use a three-stage sample design in which the grantees are selected as the primary sampling units (PSUs), sites are selected within selected grantees, and patients are selected within selected sites. Because of the high costs involved with recruiting a grantee and hiring a field interviewer (FI) to perform the data collection, we will select an independent patient sample from each funding program for grantees receiving multiple funding programs. The sample design allows us to obtain more patient interviews with fewer data collection costs due to the high costs of recruiting grantees.

In this report, we summarize the three-staged sample design that will be used for the Patient Surveys. The sample design will allow for controlled sampling of important characteristics to ensure that certain comparisons can be made both within the CHC and Special Populations studies, and to other national studies.

---

[1] RTI International is a trade name of Research Triangle Institute.

# 2. Target Population

The target population for the PHCPS is defined as persons receiving face-to-face services from a CHC, MHC, HCH, or PHPC grantee, and as persons receiving these services from a clinical staff member who exercises independent judgment in the provision of services.[2] Clients of grantees located within the 50 United States and the District of Columbia are included; clients of grantees within U.S. territories and possessions are excluded.

Only persons who received services through one of these grantees at least once in the year prior to the current visit were considered eligible for the survey. This eligibility criterion was used because many of the questions in the survey ask about services received in the past year; individuals without previous visits would not have been able to answer these questions and, therefore, were not considered eligible. This eligibility criterion was also implemented in the BPHC's previous Community Health Center Survey (2002) and Healthcare for Homeless Survey (2003).

---

[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.

# 3. Grantee Sample Selection

This section discusses the first stage of sample selection, which is the selection of grantees. The process of selecting grantees includes sample frame construction, stratification, and selection of stratified probability proportional to size (PPS) grantee samples.

## 3.1    Sampling Frame Construction

Updated BPHC Uniform Data System (UDS) grantee-level data will be used to construct a sampling frame for the first stage of selection. The UDS is compiled each year from annual data submissions by each Section 330-funded grantee. The UDS contains data on key characteristics of the grantees, such as the type of grant funding received, geographic region, urban/rural location, number of sites within a grantee, number of patients served, and other information. These same grantee characteristic profiles will be used in stratification and sample selection. In this statistical design plan, we used data from the 2006 UDS (2007 UDS data have not yet been collected) to illustrate the design plan. Once OMB approval has been received, the final sample will be drawn using the most current UDS data.

The 2006 UDS data were collected from 1,002 grantees. Some grantees will be excluded from the sampling frame, including

- twenty-eight grantees located in U.S. territories or possessions (i.e., those in Puerto Rico, the Virgin Islands, and the Pacific Basin),
- six grantees funded through the CHC program that only operated school-based sites (see **Section 4.1** for more detail on this decision),
- seven grantees that received MHC funding only and that served clients through a voucher program, and
- any grantee that is no longer a Section 330-funded grantee.

A total of 961 eligible grantees reporting in 2006 will be included in the grantee sampling frame. We show the distribution of key grantee characteristics in **Exhibits 1** and **2**. **Exhibit 1** breaks the grantees down by funding program, region, urban/rural location, and number of sites. In the grantee sampling frame, 732 grantees had a single funding program, while 229 grantees received funding from multiple programs. The majority of grantees, roughly 90% and including grantees participating in a single and/or multiple funding programs, received some CHC funding.

The number of sites within a grantee ranged from 1 to 91. There were 654 grantees that had at least three sites, with an average of about six sites per grantee. A little over one-third of the grantees had four to nine sites. The South had 340 grantees, while the West had 265. The Northeast and Midwest had roughly the same number of grantees with 176 and 180, respectively. Slightly more grantees were in rural areas than were in urban areas.

Another important grantee characteristic is the number of patients in 2006 (**Exhibit 2**). Among the 961 eligible grantees in the grantee sampling frame, the number of patients receiving at least one face-to-face encounter for services during 2006 varied among the grantees, ranging from 139 to 203,556 and averaging 15,168. The total number of patients was approximately 14.6 million.

**Exhibit 1.  Grantee Characteristics in the Sampling Frame**

| Domain Category | Number of Grantees | Percent Distribution |
|---|---|---|
| **Funding Program Received** | | |
| C | 651 | 67.74% |
| H | 68 | 7.08% |
| M | 9 | 0.94% |
| P | 4 | 0.42% |
| CH | 82 | 8.53% |
| CM | 98 | 10.20% |
| CP | 15 | 1.56% |
| PH | 5 | 0.52% |
| CMH | 16 | 1.66% |
| CMP | 2 | 0.21% |
| CPH | 9 | 0.94% |
| CMPH | 2 | 0.21% |
| *Total* | *961* | *100%* |
| **Region**[a] | | |
| Northeast | 176 | 18.31% |
| Midwest | 180 | 18.73% |
| South | 340 | 35.38% |
| West | 265 | 27.58% |
| *Total* | *961* | *100%* |
| **Urban/Rural Location** | | |
| Urban | 467 | 48.60% |
| Rural | 494 | 51.40% |
| *Total* | *961* | *100%* |
| **Number of Sites** | | |
| 1 | 160 | 16.65% |
| 2 | 147 | 15.30% |
| 3 | 125 | 13.01% |
| 4–9 | 353 | 36.73% |
| 10–14 | 93 | 9.68% |
| 15–19 | 42 | 4.37% |
| ≥20 | 41 | 4.27% |
| *Total* | *961* | *100%* |

NOTE: C = Community Health Center Program; H = Healthcare for Homeless Program; M = Migrant Health Center Program; P = Public Housing Primary Care Program.

[a] "Region" refers to Census region here.

**Exhibit 2.   Distribution of Patients Served in 2006**

| Patient Distribution | Number of Patients |
|---|---|
| Range of Number of Patients | |
|     Minimum | 139 |
|     25th Percentile (Q1) | 5,1272 |
|     Median | 10,321 |
|     75th Percentile (Q3) | 19,539 |
|     Maximum | 203,556 |
| | |
| Mean Number of Patients per Grantee | 15,168 |
| | |
| Total Number of Patients Across All Grantees | 14,561,166 |

## 3.2    Stratification Variables

Comparing the CHC survey to the National Health Interview Survey and comparing survey outcomes between funding programs are the primary analytic goals for BPHC. Therefore, our sample design goals are as follows:

- Select a sufficient number of patients to complete 4,522 interviews (2,210 for the CHC contract and 2,312 for the Special Populations contract, as per the modified contract).

- Within Special Populations, maintain roughly the same number of patient interviews for both HCH and MCH with slightly fewer from the PHPC.

As shown in **Section 2**, the majority of grantees receive CHC funding, while relatively few grantees receive PHPC and/or MHC funding. A random selection of grantees without any stratification would result in very small grantee sample sizes, and consequently small patient sample sizes, for the MHC and PHPC funding programs (**Exhibit 3**).

**Exhibit 3.   Grantee and Patient Yields from Unstratified Random Sampling**

| Grantee Funding Type | Number of Grantees Selected | Expected Number of Complete Interviews |
|---|---|---|
| C | 105 | 3,255 |
| H | 22 | 682 |
| M | 15 | 465 |
| P | 4 | 124 |
| **Total** | **146** | **4,526** |

NOTE: C = Community Health Center Program; H = Healthcare for Homeless Program; M = Migrant Health Center Program; P = Public Housing Primary Care Program.

In this scenario, the number of selected grantees is determined using proportional allocation by the number of grantees for each funding type, as shown in **Exhibit 1**. A simple random sample of grantees is selected. The result is displayed in **Exhibit 3**. In this selection scenario, we select 115 unique grantees. If a selected grantee participates in multiple funding programs, we would take an independent sample of each funding program. For example, if a grantee receiving both CHC and MHC funding is recruited, this grantee would be counted as a CHC grantee and also as an MHC grantee. Therefore, there are 105 CHC grantees, which count as more than 70% of the total yielded grantees. The PHPC program has only 4 grantees. To calculate the expected completed interviews, we further assume the same number of completed patient interviews is obtained from each funding program in a grantee and 4,522 total

completed interviews for both studies. The patient sample size for MHC and PHPC is very small. Ultimately, we would have very limited statistical power to perform comparisons.

To facilitate the comparison of survey outcomes between funding programs, a stratified sampling method with different sampling rates for selecting grantees within each stratum is necessary. To this end, we will create four mutually exclusive strata by grouping grantees according to the types of funding they receive. These four groups will serve as the first-stage strata and are defined as follows:

- Stratum 1: Grantees with CHC Funding Only.
- Stratum 2: All Grantees with PHPC Funding.
- Stratum 3: Remaining Grantees with MHC Funding.
- Stratum 4: All Remaining Grantees Not Included in Strata 1–3.

The number of grantees within each stratum is displayed in **Exhibit 4**.

### Exhibit 4.    Definition of First-Stage Stratification

| First-Stage Strata | Grantee Funding Type | Number of Grantees |
|---|:---:|:---:|
| Stratum 1: Grantees with CHC Funding Only | C | 651 |
| Stratum 2: All Grantees with PHPC Funding | P; CP; PH; CMP; CPH; CMPH | 37 |
| Stratum 3: Remaining Grantees with MHC Funding | M; CM; CMH | 123 |
| Stratum 4: All Remaining Grantees Not Included in Strata 1–3 | H; CH | 150 |
| **Total** | | **961** |

NOTE: C = Community Health Center Program; H = Healthcare for Homeless Program; M = Migrant Health Center Program; P = Public Housing Primary Care Program.

The above first-stage strata are used to ensure that the selected grantees are representative to the four funding programs. To ensure the grantees with single funding type of MHC or HCH are represented in the grantee sample, we split Stratum 3 and Stratum 4 into two second-stage strata as shown in **Exhibit 5**: Stratum 3.1, Stratum 3.2, Stratum 4.1, and Stratum 4.2.

### Exhibit 5.    Definition of First-Stage and Second-Stage Stratification

| First-Stage and Second-Stage Strata | Grantee Funding Type | Number of Grantees |
|---|:---:|:---:|
| Stratum 1: Grantees with CHC Funding Only | C | 651 |
| Stratum 2: All Grantees with PHPC Funding | P; CP; PH; CMP; CPH; CMPH | 37 |
| Stratum 3: Remaining Grantees with MHC Funding | M; CM; CMH | |
| Stratum 3.1: CM and CMH Grantees | CM; CMH | 114 |
| Stratum 3.2: M Grantees | M | 9 |
| Stratum 4: All Remaining Grantees Not Included in Strata 1–3 | H; CH | |
| Stratum 4.1: CH Grantees | CH | 82 |
| Stratum 4.2: H Grantees | H | 68 |
| **Total** | | **961** |

NOTE: C = Community Health Center Program; H = Healthcare for Homeless Program; M = Migrant Health Center Program; P = Public Housing Primary Care Program.

Furthermore, to ensure the selected grantee sample within six first-stage and second-stage strata are representative of grantees with different patient sizes, we further split six strata into several third-stage strata according to the patient size of a grantee. In each of six strata, we calculate the 33rd and 66th percentile of patient size. Grantees with patient sizes over the 66th percentile are defined as "Large" grantees, grantees with patient sizes below the 33rd percentile are defined as "Small" grantees, and grantees with patient sizes between the 33rd and 66th percentiles are defined as "Medium" grantees. In order to have the minimum sample size be larger than 10 in each final stratum, some first-, second-, and third-stage strata are collapsed due to small sample size. Thus, there are a total of 12 final strata in the grantee sample stratification, as shown in **Exhibit 6**.

## Exhibit 6.    Definition of Final Stratification

| Three-Stage Strata | Grantee Funding Type | Final Strata | Number of Grantees |
|---|---|---|---|
| Stratum 1: Grantees with CHC Funding Only | C | | |
| Stratum 1.1.1: Large | | 1 | 179 |
| Stratum 1.1.2: Medium | | 2 | 234 |
| Stratum 1.1.3: Small | | 3 | 238 |
| Stratum 2: All Grantees with PHPC Funding | P; CP; PH; CMP; CPH; CMPH | | |
| Stratum 2.1.1: Original Stratum 2 | | 4 | 37 |
| Stratum 3: Remaining Grantees with MHC Funding | M; CM; CMH | | |
| Stratum 3.1: CM and CMH Grantees | CM; CMH | | |
| Stratum 3.1.1: Large | | 5 | 72 |
| Stratum 3.1.2: Medium | | 6 | 31 |
| Stratum 3.1.3: Small | | 7 | 11 |
| Stratum 3.2: M Grantees | M | | |
| Stratum 3.2.1: Original Stratum 3.2 | | 8 | 9 |
| Stratum 4: All Remaining Grantees Not Included in Strata 1-3 | H; CH | | |
| Stratum 4.1: CH Grantees | CH | | |
| Stratum 4.1.1: Large + Medium | | 9 | 51 |
| Stratum 4.1.2: Small | | 10 | 31 |
| Stratum 4.2: H Grantees | H | | |
| Stratum 4.2.1: Large | | 11 | 19 |
| Stratum 4.2.2: Medium + Small | | 12 | 49 |
| **Total** | | | **961** |

NOTE: C = Community Health Center Program; H = Healthcare for Homeless Program; M = Migrant Health Center Program; P = Public Housing Primary Care Program.

In addition to the 12 strata for grantee sample selection discussed above, we will sort the sampling frame by region (Northeast, Midwest, South, and West), urban/rural location, and number of sites per grantee within each final stratum 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 will provide implicit stratification according to the order of the units in a stratum. The selected grantee samples will be distributed among various regions, urban/rural locations, and number of sites.

## 3.3    Select Stratified PPS Sample of Grantees

In the multi-stage sample design, when the cluster size varies greatly, the unequal probability sampling of cluster within each stratum will result in estimates of population characteristics, especially population totals that have lower variance than those obtained from clusters with equal probability (Levy

and Lemeshow, 1999). As mentioned in **Section 3.1**, the grantees differ widely with respect to the number of patients served. PPS sampling is a commonly used method of unequal probability sampling in which the probability of a cluster being sampled is proportional to the level in that cluster of some size measure. We will use PPS sampling to select the grantee sample from each final stratum. The size measurement will be the number of patients who visited the grantee for services.

Before selecting a grantee sample from each final stratum, we will determine the grantee sample allocation for each final stratum. We plan to recruit 115 unique grantees to participate in the studies to achieve our targeted 4,522 completed patient interviews, 2,210 interviews for CHC and 2,312 interviews for Special Populations. The grantees with PHPC- and MHC-funded programs will be over-sampled, while grantees with CHC-funded programs will be under-sampled. We start by determining the grantee sample allocation to the first-stage strata (4 strata), and then allocate the grantee sample to the sub-stratum at second- and third-stage stratification. The grantee sample allocation determination steps are ordered as follows:

*Step 1:*    Select 31 (85% of 37) grantees receiving PHPC funding in Stratum 2.

*Step 2:*    Subtract the 31 PHPC grantees in Stratum 2 from the 961 total grantees (n = 930).

*Step 3:*    Calculate the proportion of remaining grantees in Stratum 3 and Stratum 4. The proportion of grantees in Stratum 3 is 13.23% (123 / 930), and the proportion of grantees in Stratum 4 is 16.13% (150 / 930).

*Step 4:*    Subtract the 31 PHPC grantees in Stratum 2 from the total selected grantee sample of 115 (n = 84).

*Step 5:*    Determine the number of grantees to be selected for Stratum 3 and Stratum 4 using the percentages calculated in Step 3; the number of grantees in Stratum 3 is 11, and the number of grantees in Stratum 4 is 14.

*Step 6:*    Over-sample Stratum 3 by applying an over-sampling ratio of 2.5 (note that this ratio is determined to have roughly the same number of interviews for MHC and HCH grantees) to the proportionate sample of 11, resulting in 28 grantees in Stratum 3.

*Step 7:*    Retain a proportionate sample for Stratum 4 from Step 5, which is 14.

*Step 8:*    Allocate to Stratum 1 the remaining grantees that have not been allocated to Stratum 2, Stratum 3, or Stratum 4 (42 grantees).

*Step 9:*    Determine the grantee sample size for each sub-stratum proportionally to the number of grantees in each sub-stratum.

As a result, 42 unique grantees are selected from Stratum 1, 31 unique grantees are selected from Stratum 2, 28 unique grantees are selected from Stratum 3, and 14 unique grantees are selected from Stratum 4. **Exhibit 7** displays the sample allocation of grantees and the sampling rate for each stratum. The overall sampling rate is 11.97%. The grantee sample allocation to the final 12 strata is shown in **Exhibit 8**.

## Exhibit 7.    Grantee Sample Allocation of a Stratified Disproportionate Sampling for the First-Stage Strata

| Strata | Total Number of Grantees | Selected Grantees | Sampling Rate |
|---|---|---|---|
| Stratum 1: Grantees with CHC Funding Only | 651 | 42 | 6.45% |
| Stratum 2: All Grantees with PHPC Funding | 37 | 31 | 83.78% |
| Stratum 3: Remaining Grantees with MHC Funding | 123 | 28 | 22.76% |
| Stratum 4: All Remaining Grantees Not Included in Strata 1–3 | 150 | 14 | 9.33% |
| **Total** | **961** | **115** | **11.97%** |

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

**Exhibit 8.    Grantee Sample Allocation to Final Strata**

| Three-Stage Strata | Grantee Funding Type | Final Strata | Number of Grantees Selected in the Sample |
|---|---|---|---|
| Stratum 1: Grantees with CHC Funding Only | C | | |
| Stratum 1.1.1: Large | | 1 | 12 |
| Stratum 1.1.2: Medium | | 2 | 15 |
| Stratum 1.1.3: Small | | 3 | 15 |
| Stratum 2: All Grantees with PHPC Funding | P; CP; PH; CMP; CPH; CMPH | | |
| Stratum 2.1.1: Original Stratum 2 | | 4 | 31 |
| Stratum 3: Remaining Grantees with MHC Funding | M; CM; CMH | | |
| Stratum 3.1: CM and CMH Grantees | CM; CMH | | |
| Stratum 3.1.1: Large | | 5 | 16 |
| Stratum 3.1.2: Medium | | 6 | 7 |
| Stratum 3.1.3: Small | | 7 | 3 |
| Stratum 3.2: M Grantees | M | | |
| Stratum 3.2.1: Original Stratum 3.2 | | 8 | 2 |
| Stratum 4: All Remaining Grantees Not Included in Strata 1–3 | H; CH | | |
| Stratum 4.1: CH Grantees | CH | | |
| Stratum 4.1.1: Large + Medium | | 9 | 5 |
| Stratum 4.1.2: Small | | 10 | 3 |
| Stratum 4.2: H Grantees | H | | |
| Stratum 4.2.1: Large | | 11 | 2 |
| Stratum 4.2.2: Medium + Small | | 12 | 5 |
| **Total** | | | **116** |

NOTE: C = Community Health Center Program; H = Healthcare for Homeless Program; M = Migrant Health Center Program; P = Public Housing Primary Care Program.

To account for selected grantees' refusals to participate, we will select 15% more grantees for each stratum. We assume an 85% grantee recruitment rate. The remaining 15% of grantees selected will be held in reserve to replace grantees refusing to participate in the study.

After a PPS selection of grantees in each of the 12 strata is completed, 116 grantees will be in the sample. As stated in **Section 3.2**, an independent sample will be selected for each funding program if a selected grantee has multiple funding programs, which will yield 192 funding programs out of 116 grantees. To achieve interview targets of 2,210 CHC patients and 2,312 special population patients, the number of complete interviews for each funding type is calculated and displayed in **Exhibit 9**.[3] For this grantee sample selection scenario, there are 99 CHC grantees, 31 HCH grantees, 31 MHC grantees, and 31 PHPC grantees from which we will select the next stage sample and sites. We discuss this selection further in **Section 4**.

---

[3] Note that during the sampling plan implementation, the sample realization may yield a slightly different distribution of grantees for each funding type.

**Exhibit 9.    Yield of the Grantee Funding Type and Patients of a Stratified Disproportionate Sampling**

| Funding Program | Number of Grantees for Each Funding Program | Number of Patients per Grantee for Each Funding Program | Number of Completed Interviews for Each Funding Program |
|---|---|---|---|
| C | 99 | 23 | 2,277 |
| H | 31 | 27 | 837 |
| M | 31 | 27 | 837 |
| P | 31 | 22 | 682 |
| **Total** | **192** | | **4,633** |

NOTE: C = Community Health Center Program; H = Healthcare for Homeless Program; M = Migrant Health Center Program; P = Public Housing Primary Care Program.

We expect 2,277 interviews for CHC and 2,356 interviews for Special Populations, which is roughly on target with the goal of 2,210 interviews for CHC and 2,312 for Special Populations (**Exhibit 9**). For the Special Populations survey, the number of interviews for PHPC-funded programs is smaller than the number of interviews for HCH- and MCH-funded programs. The reason for this difference is that the PHPC patient population is much smaller than the HCH and MCH patient populations. We discuss this difference in detail in **Section 5**, where we discuss the patient sample selection.

In **Exhibit 10**, we display the grantee and patient sample distribution by region and urban/rural area from the sampling realization discussed above. The distributions of grantee and patient sample by region and the patient sample distribution by urban/rural area are very similar to the distributions of grantee frame and patient population. The grantee sample has a slightly higher proportion of urban grantees than the proportion in the grantee frame. The reason for this difference is that we selected 31 PHPC grantees out of 37, and those PHPC grantees are mainly in urban areas (35 in urban areas, 2 in rural areas). Therefore, our proposed grantee sample selection and patient sample selection methods produced grantee and patient samples that represented the target population in different regions and urban/rural areas very well.

**Exhibit 10.   Grantee and Patient Sample Distribution by Region and Urban/Rural Area**

| Domains | Grantee Frame | | Grantee Sample | | Patient Population | | Patient Sample | |
|---|---|---|---|---|---|---|---|---|
| | N | % | n | % | N | % | n | % |
| **Region** | **961** | **100%** | **116** | **100%** | **14,561,166** | **100%** | **4,633** | **100%** |
| Northeast | 176 | 18.3% | 23 | 19.8% | 2,942,832 | 20.2% | 931 | 20.1% |
| Midwest | 180 | 18.7% | 22 | 19.0% | 2,672,756 | 18.4% | 839 | 18.1% |
| South | 340 | 35.4% | 37 | 31.9% | 4,516,264 | 31.0% | 1,521 | 32.8% |
| West | 265 | 27.6% | 34 | 29.3% | 4,429,314 | 30.4% | 1,342 | 29.0% |
| **Urban/rural** | **961** | **100%** | **116** | **100%** | **14,561,166** | **100%** | **4,633** | **100%** |
| Urban | 467 | 48.6% | 65 | 56.0% | 8,236,600 | 56.6% | 2,643 | 57.0% |
| Rural | 494 | 51.4% | 51 | 44.0% | 6,324,566 | 43.4% | 1,990 | 43.0% |

When we select the real sample for the study, to meet the targeted number of complete interviews of 4,522 (2,210 for CHC and 2,312 for Special Populations), we may need to adjust the sampling rates at grantee selection for each stratum described previously and the number of interviews per grantee for a specific funding program.

## 3.4    Grantee Selection Probability

The selection probability for the $i^{th}$ grantee within the $h^{th}$ stratum is given by

$$G_{hi} = n_h \frac{S_{hi}}{\sum_i S_{hi}}, \qquad (1)$$

where $h$ is the index for the strata (Stratum 1, Stratum 2, Stratum 3, and Stratum 4); $i$ is the index for grantees on the frame within each stratum; $n_h$ is the number of grantees to select in the $h^{th}$ stratum; and $S_{hi}$ is the size measure, which is the number of patients in each grantee.

We are aware that applying different sampling rates for each stratum and taking grantee samples through PPS sampling causes an increase of variability of the selection probability. Consequently, this application of different sampling rates increases variability of sampling weights, namely increases the unequal weighting effect (UWE) on the variances of sample estimates and reduces the statistical power of the analysis. To minimize the impact of UWE, we will select sites within grantees using PPS sampling in the second stage of selection, and we will select the same number of patients per grantee in the third stage of sample selection. We address these issues in more detail in **Section 4** and **Section 5**.

# 4. Site Sample Selection

As discussed previously, more than half of the grantees have three or more sites, and, in general, those grantees with more sites tend to have more patients. Furthermore, the grantees are selected with the PPS method at the first stage of selection, which means that grantees with large numbers of patients have a higher probability of being selected in the sample. As a result, we expect a fair number of the grantees recruited to have more than three sites. We will allow at most three sites for each funding program within a grantee to be in the PHCPS; therefore, for those grantees with more than three sites, we will select three from their larger total. This section discusses the second stage of selection, which is the selection of sites from participating grantees that have multiple sites.

## 4.1    Determine Eligible Sites within Participating Grantees

Once a grantee is recruited and agrees to conduct the study in its sites, our recruiters will work with the grantee's administration to identify eligible sites. The following eligibility criteria will be used:

- The site should participate 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 temporary clinic.
- The site is not a school-based health center.
- The site is not a specialized clinic, excepting clinics providing OB/GYN services.

Due to the complexity of recruiting school-based sites, including the extensive efforts associated with getting permission from schools and parents/guardians to interview the adolescent patients, recruiting stand-alone, school-based sites is not feasible within the current survey schedule and budget. Therefore, such sites will be excluded from the Patient Surveys. Although these sites are excluded, we will not necessarily be excluding all patients who receive school-based health services. Some children who receive medical care at a school-based health center site may also receive some of their care at a non-school-based Community Health Center site near their residence.

After the eligible sites are identified, we will collect or verify the following information from/with each participating grantee:

- number of eligible sites serving each client type (i.e., migrants, homeless, public-housing, and low-income);
- address and contact information for each eligible site; and
- number of patients served in each eligible site, overall and by type of client.

## 4.2    Evaluate Distances between Eligible Sites

In most cases, only one field interviewer (FI) will be hired to collect data within each participating grantee. Therefore, it is desired that sites are within manageable distances for the FI(s). The grantees tend to operate sites in relatively localized areas. Experience from the 2002 Community Health Center Survey showed that, out of the 70 grantees selected for the study, 82% of the sites selected were within 30 miles of the grantee's headquarters, with an additional 9.2% of the sites being within 45 miles of the central health center location. The 2002 Community Health Center data suggest that the majority of the selected sites are within an area that can be covered by one FI without incurring extensive travel costs. We expect that the same will be true for the current surveys. However, our sampling staff will evaluate distances between the administrative office/central site and the associated sites as soon as possible to determine if any modifications are needed to the selection of sites within the grantee, or if special data collection arrangements should be made. We will pay special attention to sites that are located more than 100 miles from the administrative office/central site. The Project Officer will be consulted if issues of distance arise.

## 4.3    Site Selection and Selection Probability

If there are three or fewer sites for a population type (i.e., migrant, homeless, public-housing, and low-income) and all of the sites are within a manageable distance by one FI, all of the sites will be included in the study. If one site is far from the other sites and the other sites are close to each other, the two sites that are close to each other will be selected. However, if all three sites are far from each other, we will select the site that has the largest patient volume. Similarly, when two sites for a specific funded program are far from each other, the one with the largest number of patients will be selected. Again, these special cases will be reviewed with the Project Officer.

For grantees with more than three sites for a population type, we will use a PPS sampling method similar to the one discussed in **Section 3.3** to select three sites from the sites within a manageable distance for one FI to cover. The number of patients for the sites under a specific funding program will serve as the size measure in the PPS sampling. For the grantees who participate in multiple funding programs, an independent PPS selection of sites will be conducted for each funding program, if needed.

The selection probability for the $j^{th}$ site within the $i^{th}$ grantee for $f$ funding program is given by

$$C_{fij} = \begin{cases} 1 & , \text{ if } 3 \text{ or fewer sites are all selected, or} \\[2em] \dfrac{3s_{fij}}{\sum\limits_{j} s_{fij}} & , \text{ if } 3 \text{ sites are selected through PPS sampling,} \end{cases} \tag{2}$$

where $s_{fij}$ is the number of patients in site $j$ within grantee $i$ for funding program $f$.

# 5. Patient Sample Selection

Because of the mobile nature of some of the target populations of this study, a random sample of patients will be chosen for interviews as they enter the site and register with the receptionist for services. A field interviewer will visit a selected site for a predetermined number of days in the sampling period to conduct interviews. The receptionist will be instructed to select the first eligible patient registered after the FI informs the receptionist that he/she is ready for the next interview. The receptionist will read to the selected patient a brief script about the study and direct the patient to the FI for questions or participation. The receptionist will be asked to keep track of the number of patients who enter the site and the number of patients selected while the FI is at the site to conduct data collection. The receptionist can either use tally marks to count patients as they enter or complete a table based on the sign-in sheet or appointment list before the FI leaves the site. The patient count sheets for each FI data collection visit will be sent to RTI for data entry, and counts will be used to calculate the analysis weights for the study.

To minimize the UWEs of selecting a PPS sample, the same number of patients will be selected from the grantees in each of the funding programs. As shown in **Exhibit 9** in **Section 3.3**, 99 CHC grantees, 31 MHC grantees, 31 HCH grantees, and 31 PHPC grantees are recruited for CHC and Special Populations. To achieve 2,210 completed interviews, we expect 23 patient interviews will be completed from each participating CHC grantee. For Special Populations, we will achieve 2,312 completed interviews. However, the goal of roughly the same interviews for MCH, HCH, and PHPC is difficult to achieve. In the 2006 UDS, there were 701,623 patients from the HCH program and only 129,280 patients from the PHPC program. The PHPC population is much smaller than the HCH population. We are concerned about the amount of time an interviewer may need to spend in the sites in order to contact eligible PHPC patients. Therefore, we have reduced the production goal for the PHPC population. We expect that 27 patient interviews will be completed from each participating HCH and MHC grantee, and 22 patient interviews will be completed for each participating PHPC grantee. The reduced goal of 22 completed interviews per PHPC grantee may still be too aggressive. We will check the patient volume for each selected site for served PHPC patients and consult with the BPHC Project Officer if this goal becomes problematic.

Within each grantee, if more than one site is selected into the study for a specific funding program, the number of patient interviews within that grantee will be divided equally among those sites. For example, if three sites are selected within an HCH grantee, 16 patients will be surveyed from each site.

If a grantee participates in more than one funding program, independent patient samples will be selected for each funding program. If a site is chosen for multiple funding programs, the receptionist at the site will be asked to track and to select patients on the FI visiting dates for all funding programs. The FI will screen participating patients to determine patient population types (i.e., homeless, migrant, public-housing, or low-income) and will use the appropriate questionnaire to conduct the patient interviews.

The selection probability of patient $k$ from within grantee $i$, site $j$ for funding program $f$ is given by

$$P_{fijk} = \frac{n_{fij}}{e_{fij} s_{fij}}, \tag{3}$$

where $n_{fij}$ is the number of completed interviews from grantee $i$, site $j$ for funding program $f$; $e_{fij}$ is observed patient eligibility rate in grantee $i$, site $j$ for funding program $f$; and $s_{fij}$ is the number of patients in grantee $i$, site $j$ for funding program $f$.

The probability of a patient selected in the study is the product of $G_{hi}$, $C_{fij}$, and $P_{fjik}$ in Formulas (1), (2), and (3), respectively. That is

$$\pi_{hfijk} = \frac{n_h s_{hi}}{\sum_i s_{hi}} \frac{3 s_{fij}}{\sum_j s_{fij}} \frac{n_{fij}}{e_{fij} s_{fij}}. \tag{4}$$

For a specific funding program in a grantee that has three sites selected through PPS sampling, the patient selection probability is maintained in roughly the same manner (self-weighting) within each stratum, as shown in Formula (4). Because $n_{fij}$ is the same for each site, $s_{fij}$ is cancelled out, and, within a grantee, the proportion of $\sum_j s_{fij}$ and $s_{hi}$ is a constant. However, for a specific funding program in a grantee that has three or fewer sites that are all selected, the $C_{fij}$ is 1. To maintain the same equal selection probability property, the number of interviews per site ($n_{fij}$) should be allocated proportionally to the number of patients of the site rather than equally allocating the interviews to the sites. That is

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

where $n_{fi}$ is the number of interviews from a grantee for funding program $f$.

# 6. Sample Sizes and Statistical Power

Statistical tests attempt to use data from samples to determine whether a difference exists in a population or between two populations. An example of a statistical test would be to test the null hypothesis that the number of uninsured children aged 12 or younger does not differ between the CHC and MCH populations. The power of the test is the probability that the test will find a statistically significant difference between two populations as a function of the size of the true difference between those two populations. There is always a chance that the samples will appear to support or to refute a tested hypothesis when the reality is the opposite. That risk is quantified as the statistical significance level. We use significance level of 0.05 to calculate statistical power in this document.

We are using a three-stage sample design in which the grantees are selected as the PSUs, sites are selected within grantees, and patients are selected within sites. For a specific funding program, up to three sites within a grantee will be selected. Sample sizes for grantees, sites, and patients are based on an integrated sample design across the four funding programs. The sample design considers the overlap of funding programs between sites and the clustering effect of the sites on the demographics of the patients. The clustering effect and the previously mentioned UWE together contribute to the design effect (*deff*), which is a measure of the precision gained or lost by the use of the more complex design instead of a simple random sample. The design effect is a function of the clustering effect and the UWEs. A design with a large *deff* will reduce the statistical power of the analysis.

Results from the 2002 Healthcare for Homeless and Community Health Center User Survey analyses provide valuable insights into the amount of clustering and *deff* that might be incurred in the upcoming CHC and Special Populations studies and into our ability to make comparative analyses to the other study populations. Selecting patients within sites within grantees produces a clustering effect. We approximated the intra-cluster correlation (ICC) and the UWE to more accurately gauge the effective sample size and statistical power between funding programs for the key outcome measures by using the 2002 CHC Survey data. We used the sample size from the stratified disproportionate sampling scenario from **Exhibit 9** and estimated the sample size for each funding type using demographic information and some key measures from the 2006 UDS. **Exhibits 11–16** display the statistical power of detecting a 10% difference between surveys of interest for five outcomes. **Exhibit 11** displays the statistical power of detecting a 10% difference between the CHC population and the population in National Health Interview Survey (NHIS). **Exhibit 12** displays the statistical power of detecting a 10% difference between the interested domains within the CHC population. **Exhibit 13** displays the statistical power of detecting a 10% difference between the CHC population and the previous 2002 CHC population. **Exhibit 14** displays the statistical power of detecting a 10% difference between the CHC population and the PHPC population. **Exhibit 15** displays the statistical power of detecting a 10% difference between the MHC population and the HCH population. **Exhibit 16** displays the statistical power of detecting a 10% difference between the HCH population and the previous 2002 HCH population.

For the comparative analyses between the CHC and other national surveys (such as NHIS), there should be sufficient power to provide meaningful comparison for all five outcome measures due to the increase in the sample size of the CHC population and the large sample size from the NHIS as shown in **Exhibit 11**. The statistical powers of detecting a 10% difference between the domains within the CHC population, and comparisons between the CHC population and the previous 2002 CHC population, are reasonably high for most of the comparisons (**Exhibit 12** and **Exhibit 13**). The statistical power of detecting a 10% difference between the CHC population and the PHPC population for five outcome measures is pretty good. When comparing patients diagnosed with hypertension, the chance of detecting 10% of the difference is 83.7%. This percentage indicates sufficient power to detect differences between members of these populations. When looking at females with hypertension in the Community and Public Housing populations, the power is reduced to 64.9% due to the reduction in sample size as shown in **Exhibit 14**. The ability to detect meaningful differences between the MHC population and the HCH population and between the HCH population and the previous 2002 HCH population will be somewhat limited to domains with higher sample sizes as shown in **Exhibit 15** and **Exhibit 16**.

**Exhibit 11. Statistical Power to Detect a 10% Difference between the Community Health Population and the National Health Interview Survey**

| Domain | CHC versus NHIS | | | | |
|---|---|---|---|---|---|
| | Smoking | Drinking | Asthma | Hypertension | Diabetes |
| **Total** | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| **Age Group** | | | | | |
| under 18 | 95.3% | 92.7% | 100.0% | 96.0% | 100.0% |
| 18–34 | 89.3% | 85.2% | 99.6% | 90.5% | 100.0% |
| 35–49 | 80.4% | 75.2% | 98.3% | 81.9% | 99.9% |
| 50+ | 82.5% | 77.5% | 98.7% | 84.0% | 99.9% |
| **Race/Ethnicity** | | | | | |
| NH-White | 96.7% | 94.6% | 100.0% | 97.2% | 100.0% |
| NH-Black | 85.5% | 80.7% | 99.2% | 86.8% | 100.0% |
| HISP | 96.6% | 94.5% | 100.0% | 97.2% | 100.0% |
| **Gender** | | | | | |
| Male | 98.0% | 96.5% | 100.0% | 98.4% | 100.0% |
| Female | 99.8% | 99.5% | 100.0% | 99.9% | 100.0% |
| **Insurance Status** | | | | | |
| Insured | 99.8% | 99.6% | 100.0% | 99.9% | 100.0% |
| Uninsured | 97.8% | 96.2% | 100.0% | 98.2% | 100.0% |
| **Language** | | | | | |
| English | 100.0% | 99.9% | 100.0% | 100.0% | 100.0% |
| Non-English | 92.4% | 88.8% | 99.8% | 93.3% | 100.0% |

NOTE: Power calculations are based on two-sample $t$-tests comparing prevalence rates of five outcomes with a 0.05 level of significance.

**Exhibit 12.  Statistical Power to Detect a 10% Difference within the Community Health Population**

| Domain | Comparisons within CHC | | | | |
|---|---|---|---|---|---|
| | Smoking | Drinking | Asthma | Hypertension | Diabetes |
| **Age Group** | | | | | |
| Under 18 vs. 18–34 | 66.2% | 62.8% | 86.7% | 67.4% | 93.2% |
| Under 18 vs. 35–49 | 59.2% | 56.1% | 80.0% | 60.4% | 87.8% |
| Under 18 vs. 50+ | 60.8% | 57.6% | 81.5% | 61.9% | 89.1% |
| 18–34 vs. 35–49 | 55.3% | 52.0% | 77.0% | 56.4% | 85.8% |
| 18–34 vs. 50+ | 56.6% | 53.3% | 78.6% | 57.8% | 87.2% |
| 35–49 vs. 50+ | 51.9% | 48.5% | 74.8% | 53.0% | 84.5% |
| **Race/Ethnicity** | | | | | |
| NH-White vs. NH-Black | 64.4% | 61.2% | 84.6% | 65.6% | 91.5% |
| NH-White vs. HISP | 76.0% | 72.4% | 93.6% | 77.2% | 97.6% |
| NH-Black vs. HISP | 65.6% | 61.5% | 88.4% | 67.0% | 95.1% |
| **Gender** | | | | | |
| Male vs. Female | 87.2% | 84.0% | 98.5% | 88.2% | 99.7% |
| **Insurance Status** | | | | | |
| Insured vs. Uninsured | 86.2% | 83.5% | 97.4% | 87.0% | 99.2% |
| **Language** | | | | | |
| English vs. Non-English | 80.0% | 77.3% | 94.1% | 80.9% | 97.4% |

NOTE: Power calculations are based on two-sample $t$-tests comparing prevalence rates of five outcomes with a 0.05 level of significance.

**Exhibit 13. Statistical Power to Detect a 10% Difference between the 2002 Community Population and the 2009 Community Population**

| Domain | CHC09 versus CHC02 | | | | |
|---|---|---|---|---|---|
| | **Smoking** | **Drinking** | **Asthma** | **Hypertension** | **Diabetes** |
| **Total** | 99.2% | 98.8% | 100.0% | 99.4% | 100.0% |
| **Age Group** | | | | | |
| Under 18 | 71.6% | 68.0% | 91.0% | 72.9% | 96.2% |
| 18–34 | 60.6% | 57.0% | 82.8% | 61.9% | 90.7% |
| 35–49 | 49.9% | 46.6% | 72.2% | 51.0% | 82.1% |
| 50+ | 52.1% | 48.8% | 74.6% | 53.2% | 84.2% |
| **Race/Ethnicity** | | | | | |
| NH-White | 75.3% | 71.7% | 93.1% | 76.5% | 97.3% |
| NH-Black | 55.5% | 52.1% | 78.1% | 56.7% | 87.1% |
| HISP | 75.0% | 71.4% | 93.0% | 76.2% | 97.3% |
| **Gender** | | | | | |
| Male | 79.9% | 76.5% | 95.4% | 81.0% | 98.5% |
| Female | 92.2% | 90.0% | 99.3% | 92.9% | 99.9% |
| **Insurance Status** | | | | | |
| Insured | 92.6% | 90.4% | 99.3% | 93.2% | 99.9% |
| Uninsured | 79.1% | 75.6% | 95.0% | 80.2% | 98.3% |
| **Language** | | | | | |
| English | 95.9% | 94.4% | 99.8% | 96.4% | 100.0% |
| Non-English | 65.5% | 61.8% | 86.7% | 66.7% | 93.5% |

NOTE: Power calculations are based on two-sample *t*-tests comparing prevalence rates of five outcomes with a 0.05 level of significance.

**Exhibit 14. Statistical Power to Detect a 10% Difference between the Community Health Population and the Public Housing Population**

| Domain | CHC versus PHPC | | | | |
|---|---|---|---|---|---|
| | Smoking | Drinking | Asthma | Hypertension | Diabetes |
| **Total** | 82.8% | 80.6% | 95.1% | 83.7% | 97.9% |
| **Age Group** | | | | | |
| Under 18 | 43.3% | 41.2% | 61.0% | 44.2% | 69.7% |
| 18–34 | 33.2% | 31.5% | 47.7% | 33.8% | 55.5% |
| 35–49 | 23.6% | 22.5% | 33.6% | 24.0% | 39.3% |
| 50+ | 21.2% | 20.4% | 29.8% | 21.6% | 34.8% |
| **Race/Ethnicity** | | | | | |
| NH-White | 16.2% | 15.7% | 21.8% | 16.5% | 25.0% |
| NH-Black | 42.6% | 40.1% | 61.6% | 43.5% | 71.0% |
| HISP | 45.4% | 43.1% | 63.4% | 46.2% | 72.0% |
| **Gender** | | | | | |
| Male | 41.8% | 39.9% | 58.5% | 42.6% | 66.7% |
| Female | 63.9% | 61.3% | 82.4% | 64.9% | 89.0% |
| **Insurance Status** | | | | | |
| Insured | 61.0% | 58.4% | 79.5% | 62.0% | 86.6% |
| Uninsured | 45.8% | 43.7% | 63.7% | 46.7% | 72.1% |
| **Language** | | | | | |
| English | 60.9% | 58.4% | 79.0% | 61.8% | 86.0% |
| Non-English | 44.6% | 42.2% | 63.3% | 45.5% | 72.2% |

NOTE: Power calculations are based on two-sample $t$-tests comparing prevalence rates of five outcomes with a 0.05 level of significance.

**Exhibit 15. Statistical Power to Detect a 10% Difference between the Migrant Population and the Homeless Population**

| Domain | MHC versus HCH | | | | |
|---|---|---|---|---|---|
| | Smoking | Drinking | Asthma | Hypertension | Diabetes |
| **Total** | 68.3% | 64.5% | 89.1% | 69.5% | 95.1% |
| **Age Group** | | | | | |
| Under 18 | 20.4% | 18.8% | 34.1% | 21.0% | 43.8% |
| 18–34 | 26.0% | 24.2% | 41.0% | 26.6% | 50.4% |
| 35–49 | 22.6% | 21.3% | 33.5% | 23.0% | 40.1% |
| 50+ | 16.2% | 15.3% | 23.4% | 16.5% | 28.0% |
| **Race/Ethnicity** | | | | | |
| NH-White | 9.9% | 9.6% | 12.3% | 10.0% | 13.8% |
| NH-Black | 7.3% | 7.2% | 8.4% | 7.4% | 9.0% |
| HISP | 32.0% | 29.1% | 54.9% | 33.0% | 69.2% |
| **Gender** | | | | | |
| Male | 41.1% | 38.4% | 61.3% | 42.0% | 71.6% |
| Female | 39.2% | 36.3% | 60.8% | 40.2% | 72.2% |
| **Insurance Status** | | | | | |
| Insured | 31.0% | 28.6% | 49.7% | 31.8% | 61.0% |
| Uninsured | 48.0% | 45.0% | 69.7% | 49.1% | 79.6% |
| **Language** | | | | | |
| English | 27.2% | 25.9% | 38.7% | 27.7% | 45.2% |
| Non-English | 23.3% | 21.3% | 41.1% | 24.1% | 54.1% |

NOTE: Power calculations are based on two-sample $t$-tests comparing prevalence rates of five outcomes with a 0.05 level of significance.

**Exhibit 16.  Statistical Power to Detect a 10% Difference between the 2002
Homeless Population and the 2009 Homeless Population**

| Domain | HCH09 versus HCH02 | | | | |
|---|---|---|---|---|---|
| | Smoking | Drinking | Asthma | Hypertension | Diabetes |
| **Total** | 71.4% | 67.5% | 91.5% | 72.7% | 96.6% |
| **Age Group** | | | | | |
| Under 18 | 16.2% | 15.2% | 24.6% | 16.5% | 30.4% |
| 18–34 | 26.4% | 24.5% | 41.7% | 27.0% | 51.3% |
| 35–49 | 32.6% | 30.3% | 51.3% | 33.5% | 62.1% |
| 50+ | 22.0% | 20.5% | 34.6% | 22.5% | 42.8% |
| **Race/Ethnicity** | | | | | |
| NH-White | 32.6% | 30.3% | 51.3% | 33.5% | 62.1% |
| NH-Black | 34.1% | 31.6% | 53.4% | 35.0% | 64.3% |
| HISP | 22.6% | 21.0% | 35.5% | 23.1% | 43.9% |
| **Gender** | | | | | |
| Male | 48.1% | 44.7% | 71.3% | 49.2% | 81.8% |
| Female | 37.7% | 35.0% | 58.4% | 38.7% | 69.6% |
| **Insurance Status** | | | | | |
| Insured | 28.1% | 26.1% | 44.5% | 28.8% | 54.4% |
| Uninsured | 56.1% | 52.5% | 79.7% | 57.4% | 88.8% |
| **Language** | | | | | |
| English | 64.4% | 60.5% | 86.7% | 65.7% | 93.8% |
| Non-English | 16.3% | 15.3% | 24.9% | 16.6% | 30.8% |

NOTE: Power calculations are based on two-sample *t*-tests comparing prevalence rates of five outcomes
with a 0.05 level of significance.

# 7. Data Collection

## 7.1     Schedule

PHCPS survey data will be collected over a period of 4 months. Although data collection was originally scheduled for March through May 2009, we will revise our timeline and request a no-cost extension, per BHPC's request to allow more time to review the draft questionnaire and receive OMB clearance. A revised schedule will be provided in the final version of Deliverable 4: Implementation Plan.

To reduce site burden, we will minimize the data collection period at each site. Because the estimated time to complete each interview with Special Population respondents is 3.5 hours and the estimated time to complete each interview with CHC respondents is 4.92 (see **Section 7.2** for how this estimate was created), and because the typical interviewer is only available for part-time employment, an average of 1 week of data collection will be required for every six interviews. Production goals, including all four patient populations, per site, will range from 8 to 99, which means that the field period in any given site could be as short as 2 weeks or as long as 11 weeks. However, in cases in which the production goal exceeds 50 patients, and as deemed appropriate by the site, special accommodations can be made, such as staffing an interviewer who can be at the site full time or bringing in an additional interviewer to help complete interviews in a shorter period of time.

## 7.2     Costs

The three primary field costs associated with all completed cases are interviewer labor, mileage incurred by interviewers, and incentives paid to respondents. Our statistical design and data collection plans assume interviews will be completed at a rate of 3.5 hours each for Special Population respondents and 4.92 hours each for CHC respondents. These figures include time for driving to and from a facility, waiting to be approached by eligible patients, screening potential participants, administering informed consent, administering an interview, updating field status codes and completing other administrative paper work, shipping material back to RTI, and participating in regular conference calls with his/her field supervisor. We assume that interviewers will require reimbursement for an average of 36 miles per completed interview. Finally, we have budgeted for $25 in incentives for each survey participant.

# 8. Strengths and Limitations of Study Design

Sample designs that maximize the ability to make inferences about a target population will also have limitations due to budget and schedule constraints. This section addresses the strengths and limitations in the sample design for the Patient Surveys.

## 8.1 Strengths

The three-stage PPS sample design will produce a sample of grantees, health centers, and patients that will spread the samples across the United States and across urban/rural locations and various grantee sizes according to the numbers of patients and sites. The resulting sample of patients will provide BPHC with data that will allow them to make references to the patient population receiving services through CHC, MHC, PHPC, and HCH.

The sample design has stratified the grantees into groups by funding program in order to provide samples of close to 2,210 patients in CHC and 2,312 patients in Special Populations, as discussed in **Section 5**. The sample sizes are based on the selection of 115 unique grantees, but, because we are allowing patient samples to be selected from each funding program in which the grantee participates, the patient sample sizes will be equivalent to selecting samples from 192 grantees. This sample design, which takes advantage of the multiple funding programs received by some of the grantees, results in a 67% increase of efficiency in recruiting the grantees for the PHCPS. The patient interviews from more grantees will result in better statistical power than if the patient samples had only been selected from a total of 115 grantees, with each grantee only representing one funding program.

The combined sample of patients from the four funding programs will be sufficient for comparative analyses with national estimates of U.S. residents from the NHIS on a number of outcomes and subpopulations of patients and U.S. residents. Comparative analyses between the funding programs may be adequate for larger subgroups of the patient populations.

## 8.2 Limitations

Although the sample design takes advantage of the multiple funding programs received by some of the grantees, due to budget restrictions, the patient sample sizes for each funding program will not be large enough for comparative analyses of patient characteristics between the funding programs for certain less-prevalent subgroups of the patient populations. The low statistical power estimates from such subgroups, such as comparisons of smoking, drinking, asthma, hypertension, and diabetes prevalence between MHC and HCH, are illustrated in the statistical power exhibits (**Exhibits 11–16**).

An additional limitation pertains to capturing seasonal variation in health care needs and service utilization. The time constraints for completing the study within the contract time period will limit the data collection period to 3 months. Because the data collection period will occur over a period of time that is less than 1 year, the study will not be able to address any seasonal fluctuations in the types of services provided to the health center patients during different seasons of the year. The spring data collection will underestimate patients who enter the health centers for flu shots, typically during late fall, and for cold- and flu-related illnesses that typically occur during the winter months. The short time period for data collection may also miss groups of migrant workers who are migrating to certain areas of the United States to work in fields that produce crops that need to be harvested in the spring, while some of the health centers in the study may be in areas where the crops are harvested in the fall.

Finally, for those funding programs in a grantee that has three or fewer sites that are all selected, to reduce the UWE, the number of patients selected from each site should be proportional to its number of patients (as discussed in **Section 5**). For the ease of field operation, we will allocate the patient sample equally to the sites in a similar way to how the grantees with three sites are selected through PPS sampling. In doing so, however, we could inflate the UWEs and consequently lose some statistical analysis power.

# 9. References

Chromy, J. R. (1981). "Variance Estimations for a Sequential Sample Selection Procedure" in D. Krewski, R. Platek, and J.N.K. Rao, eds. *Current Topics in Survey Sampling*. New York: Academic Press, Inc.

Levy, P. S., and Lemeshow, S. (1999). *Sampling of Population*: *Methods and Applications*. Third edition. New York: John Wiley and Sons.