Oversampling Small Low Income Communities in California using Telephone Surveys

Ismael Flores Cervantes[1] and J. Michael Brick[1]
[1]Westat, 1600 Research Blvd., Rockville, MD 20850

Abstract

This paper describes the challenges in oversampling small low-income communities in a random-digit-dial telephone survey. The California Health Interview Survey (CHIS) is a telephone survey conducted since 2001 that explores issues in public health and health care and monitors changes over time in California. As part of CHIS 2009, additional supplemental samples were included to increase the representation of young families in 14 underserved California communities. As with any survey of rare populations, the size of the communities and other eligibility requirements presented challenges for sample design, data collection, and weighting. A combination of approaches that included stratified sampling, two-phase sampling, and screening were used to sample these communities. The impact of the inclusion of the supplemental samples on weighting is also described.

1. Introduction

Due to the characteristics of telephone survey methodology, targeting small geographic represents a challenge because of the large screening effort needed to identify respondents in these areas. The difficulty of the task increases when the population of interest includes families in underserved communities which traditionally have been difficult to sample using telephone surveys. Among the reasons that make it difficult to reach this population are language barriers, predisposition not to participate in surveys, and lack or intermittent telephone service. These factors in combination with budget constraints make the implementation of a telephone survey difficult for this type of population. The successful implementation of this type of survey requires a coordinated use of different sampling procedures and data collection protocols. The objective of the sampling procedures is to increase the efficiency of the screening effort by defining sampling strata with a high prevalence of the population of interest. On the other hand, data collection protocols can increase the likelihood of response when this population is approached by using multilingual interviewers and incentives. We describe the sampling procedures used to sample 14 underserved communities in California as supplemental samples of a larger or main telephone survey. Section 2 describes the characteristics of the population of interest and the sample design of the supplemental samples. Section 3 describes sampling procedures used for rare populations and how they were applied in this case. Section 4 describes the impact of the weighting process of the main sample. The discussion is presented in Section 5.

2. The Building Healthier Communities Supplemental Samples

Fourteen small underserved communities in California were sampled as part of the 2009 California Health Interview Survey (CHIS). These communities were located in small but well-defined geographically areas in California[1] as shown in Figure 2-1.The communities were participants of the program Building Healthy Communities (BHC) which was a ten-year community initiative sponsored by The California Endowment. The focus of the BHC programs varied by community and ranged from improving access to health care, health insurance, food access, economic opportunity, and youth development[2]. The California Endowment (TCE) is a private health foundation that provides grants to community-based organizations to help improve social and economic environments, and promote the health of children and families in these communities[3].

Figure 2.1: Communities participating in the program Building Healthy Communities

Communities	
1	Boyle Heights
2	Central/Southeast/Southwest Fresno
3	Central Long Beach
4	Central Santa Ana
5	City Heights (San Diego)
6	Coachella
7	Del Norte County/Adjacent Tribal Lands
8	East Oakland
9	East Salinas
10	Richmond
11	South Los Angeles
12	South Kern County
13	South Sacramento
14	Southwest Merced/East Merced County

Source: The California Endowment at http://www.calendow.org/home.aspx

Table 2-1 shows the demographic and socioeconomic characteristics of the population in these communities. In general, these communities have large percentages of Hispanics, children under 18 years old, unemployed persons 16 years old or older, families below poverty, and adults with less than a high school.

Due to budget constraints, the BHC sample was designed to be a supplement of a larger survey. The reduction in cost was achieved by reutilizing resources already developed for the main survey. These resources included questionnaire design, sampling and data collection protocols. Since it was a supplemental sample, the number of cases from the main and supplemental samples could be combined to increase the sample size for analysis of these communities. At the same time, this approach had limitations from the framework already in place. For example, data collected in the main survey had to meet needs of users of the supplemental sample. Any additional questions asked only in the

[1]The area of the communities was defined in terms of Census 2000 block groups.
[2]www.calendow.org/healthycommunities
[3]A listed household is defined as a household whose telephone number and address appears in a white pages and it geographic location has been determined

supplemental sample had limited usability because they were only available to respondents in the supplement. The BHC sample also inherited issues affecting the main survey. In this case, since the main survey was a telephone survey, the supplemental sample also suffered from declining response rates, increasing undercoverage due to cell phone use, and issues related to telephone surveys for studying low income populations.

Table 2.1: Socio-demographic characteristics of communities participating in the program Building Healthy Communities

Community		Population	Socio demographic characteristic (Percentages)				
			Latino	Age 0-17 years old	Average Labor Force (16+ years old) Unemployed	Families Below Poverty	Less than a High School Diploma
1	Boyle Heights	73,297	92	35	13	33	69
2	Central/Southeast/ Southwest Fresno	78,284	69	39	22	43	63
3	Central Long Beach	86,046	66	39	18	42	59
4	Central Santa Ana	91,156	94	38	9	24	77
5	City Heights (San Diego)	77,937	63	37	12	35	52
6	Coachella	40,589	98	35	11	26	60
7	Del Norte County/ Adjacent Tribal Lands	29,949	15	22	12	17	28
8	East Oakland	90,770	53	35	14	23	44
9	East Salinas	32,413	95	39	22	25	76
10	Richmond	28,238	64	35	13	25	46
11	South Los Angeles	93,623	78	37	15	35	54
12	South Kern County	65,154	73	36	22	21	54
13	South Sacramento	68,277	41	34	13	25	40
14	Southwest Merced/ East Merced County	51,188	63	36	19	30	48
	Average	64,780	69	36	16	29	55

Source: Table Revised by Healthy City, September 2008. Data compiled from Claritas, 2008.

The goal of the BHC sample design was to complete 400 interviews of eligible persons in each community. The definition of eligibility had two components. The first component was based on geography that was determined after geocoding the home address of the respondents collected at the end of the interview. Only households located inside the community were eligible for the study. The second component was demographic, where children under 18 years old or adults 40 years old or younger were eligible for the study.

Due to budget constraints, a cell phone sample component was not part of BHC sample and the BHC sample had to rely on the cell phone component of the main sample. The cell phone samples were drawn by area codes that covered large geographic areas in the main sample; a costly and intensive screening operation would have been required to target the supplemental sample to the very small geographic areas to conduct cell phone interviews. Despite this limitation, there was no undercoverage from the exclusion of cell phone users in the BHC communities because they were sampled as part of the cell phone component of the main survey.

The main survey was the 2009 California Health Interview Survey (CHIS) (California Health Interview Survey, 2011a). CHIS was a random digit dialing (RDD) telephone survey of California's population first administered in 2001 which collects data on public health and access to health care in California. The main component of CHIS 2009 was a large landline sample with a sample size of 53,602 completed interviews[4] allocated to produce estimates by county. There were two supplemental samples: a cell phone supplemental sample with a size of 3,728 interviews drawn by California regions used to address the undercoverage due to the increasing use of cell phones; and a surname supplemental sample of a size of 2,608 cases used mainly to increase the representation of smaller ethnic groups within the state residents in the sample. The small size of the cell phone sample compared to the landline component had an impact on the estimates for small areas, in specific for the BHC communities as described in Section 4.

3. Sampling

A combination of the same sampling methods used in CHIS 2009 to increase the sample of smaller ethnic groups were also used to improve the efficiency of the screening operation in the BHC sample. These methods were disproportionate sampling, two-phase sampling, and screening. These approaches are described in the following sections.

3.1 Disproportionate stratified sampling

Disproportionate stratified sampling is a method that reduces the amount of screening for a small subgroup when the population can be divided into strata that differ in subgroup prevalence. The reduction of screening is achieved by oversampling the strata with higher prevalence of the subgroup at higher rates (Flores Cervantes and Kalton, 2008). In this approach in telephone surveys, sampling strata were created by classifying telephone exchanges[5] in the landline frame into geographic substrata with high and low prevalence of the population of interest. The classification of exchanges used geographic information of the listed households in the exchanges. This geographic information is available from coverage reports produced by specialized commercial firms that compile and update information such as the number of listed households and their geography in the exchanges.

The BHC subsampling strata were created within the CHIS 2009 strata after evaluating the coverage and prevalence rate of eligible households as follows. First, a list of telephone exchanges with at least one listed household in the block groups that defined a community was compiled using all exchanges in the CHIS stratum. Then the list of exchanges was sorted in descending order by the proportion of listed households within the exchange that fell inside the community. In the next step, the cumulative prevalence rate and cumulative coverage rate were computed for each exchange on the list. The prevalence rate was computed as the ratio of cumulative total households in the community to the cumulative total households in the exchanges up to the current exchange on the list. The coverage rate was computed as ratio of the cumulative total households in the community up to the current exchange to the total households in the community. In the last step, the BHC stratum was created by a choosing a cut-off point for a predefined coverage rate after evaluating the coverage and prevalence of different

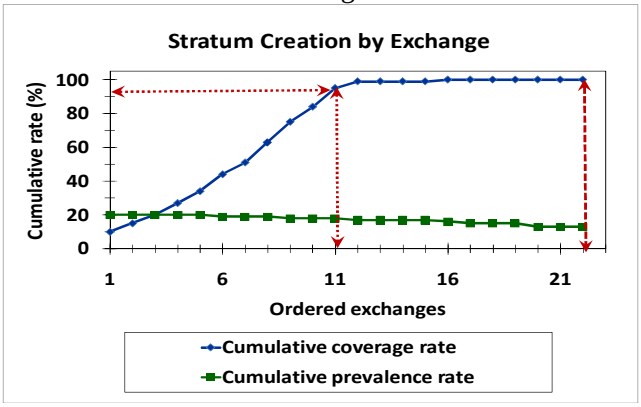
[4] This total includes adult, child and adolescent interviews

[5] A telephone exchange consist of the area codes (the first 3 digits) and the prefixes (the next 3 digits) of the 10-digit telephone numbers in the United States

sets of exchanges. The BHC substratum was defined by the group of exchanges where the cumulative coverage rate is less or equal to the cut-off point.

A graphical representation of the relationship between coverage and prevalence is shown in Figure 3-1. The plot was created using the cumulative coverage and prevalence rates for the BHC community in Kern County. The horizontal axis represents the ordered exchanges and the vertical axis represents the cumulative rates. The green and blue lines represent the cumulative prevalence and coverage rates respectively. For example, a cut-off point of 100 percent (indicated by the red dashed line) defines a substratum with 21 exchanges which contains all listed households of the community (i.e., 100 percent coverage). The corresponding prevalence rate is 13.0 percent indicated by the intersection of the vertical line and the green line. In this case, we expected to find 13 BHC households for every 100 contacted household in this stratum.

Figure 3.1: Graphical representation of the creation of the BHC stratum using telephone exchanges



A common practice for increasing the efficiency of screening is to trade lower coverage (i.e., introduce bias) for a higher prevalence by selecting a different cut-off point. For example, Figure 3-1 shows that the 90 percent cut-off stratum (i.e., the substratum contains only 90 percent of the households in the community) with 11 exchanges has a prevalence rate of 18 percent. This approach may produce biased estimates because part of eligible population in the excluded exchanges was not sampled. However, in the BHC design, the excluded exchanges (exchanges with higher cut-off point and exchanges with no households in the community within the CHIS stratum) were sampled as part of the CHIS main sample. Although the bias was not a problem, the differential sampling rates between the CHIS and BHC strata reduced the precision of BHC estimates when households in the communities were sampled in the non-BHC stratum.

The situation shown in Figure 3-1 is common in telephone surveys of small geographic areas where there are not important changes in the prevalence rate at different cut-off points. Therefore a different approach was evaluated and implemented in the BHC samples. In this approach, 100 banks rather than exchanges were used to create BHC substrata. Since the 100-bank was a smaller unit than the exchange, this method achieved a better mapping between the geographic area and the telephone numbers. The plot in Figure 3-2 shows the graphical representation of this approach. Although this plot is constructed in a similar way to the plot based on exchanges, the cumulative prevalence and coverage rates are computed by cumulating banks rather than exchanges. In the case for South Kern County, the 100 percent cut off includes 2,060 100-banks that include 100

percent of the listed households in the community with a prevalence rate of 34.7 percent, much higher than the prevalence rate with a cut-off point of 90 percent for a stratum crated at the exchange level. The prevalence was further increased to 46.4 percent by selecting a cut-off point of 90 percent. This BHC stratum contained 1,041 banks.

Figure 3.2: Graphical representation of the creation of the BHC stratum using banks

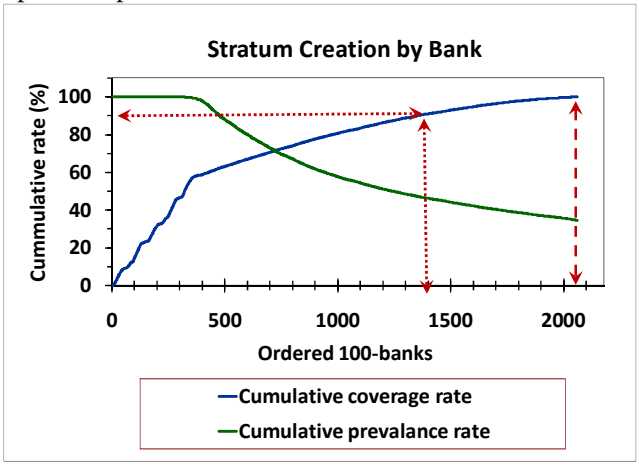


Table 3-1 compares the prevalence rates of strata created using exchanges and banks with 90 and 100 percent cut-off points for the BHC communities. The last column shows the reduction in screening comparing the 90 percent cut-off points of the two approaches. The reduction was computed as the ratio of the number of households needed to screen in order to contact the same number of eligible households based on strata created using bank or exchanges. The reduction in screening varied by community with an average reduction of 15.6 percentage points.

Table 3.1: Expected BHC prevalence rate (percentage)

		Exchange		Bank		Reduction in screening effort
		Cut-off point		Cut-off point		
Community		100	90	100	90	
1	Boyle Heights	5.0	43.0	34.7	44.7	3.8
2	Central Santa Ana	19.0	29.0	28.1	34.2	15.3
3	Central/Southeast/Southwest Fresno	12.0	36.0	30.3	39.2	8.2
4	City Heights	12.0	28.0	27.2	35.0	19.9
5	Coachella Valley	19.0	68.0	51.1	78.2	13.0
6	East Oakland	11.4	43.9	35.6	49.3	10.9
7	East Salinas (Alisal)	24.0	43.0	37.2	45.9	6.3
8	Long Beach	15.0	54.0	43.0	55.6	2.9
9	Richmond	17.0	27.0	27.2	31.2	13.5
10	Sacramento	8.0	26.0	23.9	32.0	18.6
11	South Los Angeles	8.0	26.0	23.5	30.7	15.4
12	South Kern	13.0	18.0	34.7	46.4	61.2
13	Southwest Merced/East Merced County	26.0	48.6	51.7	56.4	13.8
14	Del Norte County Adjacent Tribal Lands	100.0	100.0	100.0	100.0	N/A
Average*		14.6	37.7	34.5	44.5	15.6

*Del Norte is excluded because it matched a CHIS sampling strata no BHC substratum was created

3.2 Two-phase sampling

Two-phase sampling or double sampling was also used to increase the efficiency of sampling the communities. In this approach, basic and inexpensive data are collected for the first-phase sample. This information is used to create subsampling substrata with varying prevalence of membership of the group of interest. In the second phase, a disproportionate stratified sample is selected, oversampling the higher-density strata (Flores Cervantes and Kalton, 2008).

As part of the CHIS 2009 protocols, the sample was preprocessed to remove nonproductive numbers (i.e., nonworking and business numbers) and to obtain a mailing address for the remaining numbers. These procedures were implemented to improve the efficiency of dialing and to increase response rates. In order to implement the two-phase approach, telephone numbers with a mailing address were geocoded to determine their geographic location. Using this information, four substrata for additional subsampling were created as shown in Table 3-2.

Table 3.2: Creation of second-phase subsampling substrata and subsampling rates

Second phase sampling substratum	Telephone number working /residency status	Did telephone number have a matched address?	Was the mailing address geocoded to a community?	Subsampling rate
1	Nonworking/ business	n/a	n/a	0
2	Residential	Yes	No	0
3	Residential	Yes	Yes	1
4	Residential	No	n/a	*R*

Stratum 1 did not include any households and no telephone number was retained for the second selection (i.e., second phase subsampling rate r=0). The likelihood of households in stratum 2 to be in the community was small so no telephone numbers in this stratum were retained. Finally, the prevalence rate in stratum 4 was likely to be high so telephone numbers in this stratum were retained with certainty (i.e., subsampling rate r=1). Since the prevalence rate in stratum 4 was unknown, this stratum was subsampled at a rate r computed so it yielded the minimum sample size of telephone numbers to draw in the first phase with the following constraints:

- The total number of completed BHC eligible interviews in a community was 400.
- The maximum design effect for the estimate of the total eligible population computed using the BHC sample was 1.5.

Mathematically the problem was equivalent to minimizing the sample n to draw in the first phase constrained to

$$DEFF = \left(\sum_{h \geq 3} A_h r_h \right) \left(\sum_{h \geq 3} \frac{A_h}{r_h} \right) \leq c \text{, and } \sum_h n\, r_h P_h P A_h = m_0 \text{;}$$

where $DEFF$ was the design effect, c was a constant for the maximum value allowed for the design effect (i.e., $c = 1.5$), A_h was the proportion of BHC population in stratum h, r_h was the second phase subsampling sampling rate in stratum h (i.e., $r_1 = r_2 = 0$, and $r_3 = 1$), m_0 was the total BHC target sample size (i.e., $m_0 = 400$), P_h was the

proportion of telephone numbers in stratum *h*, and PA_h was the proportion of the population in stratum *h* that is BHC eligible. Initial assumptions about the quantities P_h, A_h, and PA_h were made using the main CHIS sample. The initial assumed proportions were low and the optimization was recomputed several times during data collection after they were revised using the results of BHC sample.

The overall proportion of telephone numbers drawn at the first phase was 84 percent, but it varied by community. All telephone numbers in the BHC strata were drawn in the first phase in 4 communities and this percentage was higher than 90 in 4 communities. The lowest rate was 58 percent in two communities. In those communities where all telephone numbers were drawn at the first phase, all cases in stratum 4 (i.e., without a matched address) were retained in the second phase. In these communities, there was no increase in the design effect due to the subsampling of strata 4. The previous expression of the DEFF assumed that there was no misclassification of BHC eligible cases among strata. As a result, no sample was drawn from stratum 2. When this assumption does not hold, the estimates from these samples are typically biased because of the exclusion of misclassified cases from the sample. However, such undercoverage did not exist in the BHC design because the misclassified cases in stratum 2 were sampled as part of the main CHIS sample. Although there was no undercoverage, the estimates were expected to be less precise due to the differential sampling rate among the BHC and CHIS strata.

3.3 Screening

Screening was also used to prevent ineligible respondents from completing the interview during data collection. Screening questions to determine the demographic eligibility of the household were included as part of the questionnaire. However, the eligibility of the BHC sample also depended on the geographic location of the respondent and this information was known only after the respondent's home address was geocoded. As a result, interviews could be completed and found to be ineligible at a later time. Additional questions such as the ZIP code of residence were added. Although the ZIP codes generally included areas larger than the community, they screened out many ineligible respondents before administering the extended interview.

Table 3-3 shows the achieved sample size for the BHC sample and the proportion of the completed interviews that were BHC eligible. On average, 88 percent of the completed interviews met the BHC eligibility definition. In other words, 730 completed interviews were discarded at the end of the study. The table also shows that the overall target sample was almost met (i.e., 0.4 percentage shortage). However, the results varied among the communities with 12.4 percent shortage in City Heights and an excess of 22.5 percent in South Los Angeles.

Table 3.3: Completed interviews and target sample size by community in the BHC sample

	Community	Completed interviews	BHC eligible	Eligibility rate %	Target (excluding CHIS samples)	Excess or shortage %
1	Boyle Heights	369	344	83.3	364	-5.5
2	Central Santa Ana	486	405	81.9	381	6.3
3	Central/Southeast/ Southwest Fresno	443	363	82.3	371	-2.2
4	City Heights	344	283	91.4	323	-12.4
5	Coachella Valley	384	351	91.3	382	-8.1
6	East Oakland	436	398	92.4	386	3.1
7	East Salinas (Ailsa)	409	378	88.8	374	1.1
8	Long Beach	409	363	80.7	377	-3.7
9	Richmond	445	359	86.9	389	-7.7
10	Sacramento	427	371	81.3	371	0.0
11	South Los Angeles	477	388	87.5	379	2.4
12	South Kern	522	457	89.8	373	22.5
13	Southwest Merced/ East Merced County	343	308	99.0	333	-7.5
14	Del Norte County Adjacent Tribal Lands	401	397	87.6	381	4.2
	Total/	5,895	5,165	88.0	5,184	-0.4

4. Weighting

In this section, we describe the impact of the inclusion of the BHC sample in the creation of the sampling weights. The creation of the CHIS weights is a complex process described in the CHIS 2009 Methodology reports (California Health Interview Survey, 2011b). The process involved multiple steps carried out separately for the landline and cell phone samples. In general, the sampling weights were adjusted for nonresponse at the screener and extended interview levels. It also included adjustment factors that took into account the probability of selection of the sampled person and any subsampling of respondents made during data collection. Since the sampled design in CHIS followed a dual frame approach with a common population sampled in different frames, the landline and cell phone samples were combined using a composite factor applied to the weights. The combined weights were then were trimmed and raked to control totals in the last step of weighting. Since the BHC sample was drawn from strata created within the landline frame (see Section 3), its inclusion to the weighting process only affected the creation of the landline weights.

In general, cases not eligible for the BHC sample (those not retained for additional subsampling in the second phase or those screened out because they did not meet the BHC geographic or demographic eligibility) were eligible for interviewing as part of the main CHIS sample. In other words, although they were considered ineligible in the BHC sample, they were treated as nonrespondents in the main CHIS. Therefore ineligible BHC cases were adjusted using regular weighting classes nonresponse adjustments that included respondents from the main sample. This was the same approach used to weight the ineligible cases from the surname samples in the main CHIS weighting.

The impact of the BHC sample and the changes made to the weighting process to the landline sample are described below:

- *Creation of the base weights*: The landline base weights were created taking into account the multiple probabilities of selection since the number could be sampled from either the landline or surname frames. In CHIS, the base weights were created conditioned on the observed sample in strata formed by the intersection of the surname and landline frames. With the inclusion of the BHC sample, the same approach was followed and it was assumed that the sample was drawn from strata formed by similar groups that that included the BHC frame, remaining of CHIS frame not included in the BHC frame, and the surname frame. The frame counts needed to create the base weights were computed taking into account that the assumed strata were created using groups of banks and not exchanges as in the regular CHIS process.
- *Adjustment for BHC sample with an address geocoded outside the BHC area (stratum 2)*: These cases were adjusted as if they were nonrespondents in the main CHIS. This required identifying CHIS cases in the BHC strata with a geocoded address outside the BHC area. The general adjustment factor applied to the CHIS cases was

$$f_c = \frac{\sum_{i \in ineligible\ BHC} wgt_i + \sum_{i \in CHIS} wgt_i}{\sum_{i \in CHIS} wgt_i},$$

  which was computed separately by community.
- *Adjustment for BHC sample without a mailing address (stratum 4):* Cases in these groups were subsampled at the optimal rate described in Section 3. These cases were also adjusted for nonresponse taking into account the CHIS cases in this group and those that BHC sample that were selected in the second phase. The form of the adjustment was

$$f_c = \frac{\sum_{i \in BHC_s} wgt_i + \sum_{i \in BHC_{\bar{s}}} wgt_i + \sum_{i \in CHIS} wgt_i}{\sum_{i \in BHC_s} wgt_i + \sum_{i \in CHIS} wgt_i}$$

  where BHC_s and $BHC_{\bar{s}}$ are the group of BHC cases retained and excluded in the second phase selection respectively.

- *Adjustment for demographic or geographic ineligibility at the end of the screener interview.* These cases completed the screener interview but were ineligible for BHC sample. These were adjusted as nonrespondents taking into account the CHIS and surname samples. The form of the adjustment was

$$f_c = \frac{\sum_{i \in BHC_R, BHC_{NR}, BHC_{IN}} wgt_i + \sum_{i \in CHIS_R, CHIS_{NR}} wgt_i + \sum_{i \in SUR_R, SUR_{NR}, SUR_{IN}} wgt_i}{\sum_{i \in BHC_R} wgt_i + \sum_{i \in CHIS_R} wgt_i + \sum_{i \in SUR_R} wgt_i}$$

  where BHC_R BHC_{NR}, and BHC_{IN} were the respondent, nonrespondent, and ineligible cases in the BHC sample respectively, $CHIS_R$, and $CHIS_{NR}$ were the respondent and nonrespondent cases in the CHIS sample respectively, and SUR_R,

SUR_{NR}, and SUR_{IN} were the respondent, nonrespondent, and ineligible cases in the surname sample respectively.

Other adjustments such as adjustment for unknown residential status, unknown eligibility, and screener nonresponse were implemented in the same way as in the regular CHIS but with separate weighing classes for the BHC communities.

After the screener and extended interview weighting adjustments, the weights were poststratified to telephone usage before the cell and landline sample were combined using a composite factor. In the final step, the weights were raked to control totals. One additional raking dimension with the total population of each BHC community was added to the CHIS raking dimensions. Eligible and ineligible BHC cases with a completed extended were raked to the BHC totals.

Although the combined CHIS and BHC sample was weighted without any major changes to the weighting process, the final weights for the BHC communities had very large coefficients of variation. Further investigation showed that the cell phone sample was 0.6 percent of the total sample size in the community, the cell phone sample accounted an average 37 percent of total sum of weights (i.e., estimate of the population in the community). This was the result of the large relative sampling rate between the combined CHIS-BHC landline sample and the CHIS cell phone sample. Although the CHIS 2009 cell phone sample was small relative to the landline sample, it did not have such impact for estimates at the state or county levels. However, estimates for very small areas such as the communities where the landline respondents were further oversampled with the inclusion of the BHC landline samples, the relative sampling rates between the landline and cell phone sample became very large. The analysis showed that the landline respondents were on average oversample 46 times more than cell phone respondents. Since these large variations in weights were not acceptable for the production of estimates, the weights in the cell phone sample were trimmed separately by community in order to reduce their impact on the estimates. The cut-off value for trimming the cell phone sample weights were set to 5 times the largest landline weight within community. Cell phone weights larger than the cut-off value were trimmed before raking. The new final weights had considerable smaller coefficient of variation and the cell phone sample represented 1.9 percent of the total sum of weights. By reducing the impact of the cell phone sample in the estimates it was implicitly assumed that there are no differences between cell phone users and nonusers in the communities. Heath profiles and estimates produced using these weights can be found at The California Endowment at http://www.calendow.org/communities/building-healthy-communities/.

5. Discussion

Low income communities in small geographic areas present a challenge for telephone sampling methodology. The difficulties arise from a combination of factors such as the very large screening effort needed to contact respondents of this population, limited budgets, and characteristics of the population such as propensities for responding to the survey or prevalence of telephone use. A way to reduce costs for small surveys that target this type of population is to field the survey as a supplemental sample to a larger survey. In the case of the BHC, it was a supplemental sample within the CHIS 2009 main sample. Still, a combination of procedures such as disproportionate stratified sampling two-phase sampling was needed to increase the efficiency of sampling. However, as in most

telephone surveys where these procedures are used, these procedures have modest savings that are important in studies with limited budgets. In the case of the BHC and CHIS samples, these sampling procedures helped achieve the expected target goals although with variability by community. However, including a landline BHC supplement combined with a relatively small cell phone sample in the main CHIS survey produced a large variability in the weights. This was the result of the high relative differential rates between the BHC/CHIS landline sample and CHIS cell phone sample. The impact of the cell phone sample and the variability of the weights were reduced by trimming the cell phone sample before raking. Weights with less variability could be achieved by increasing the cell phone sample size as done in more recent cycles of CHIS. Another alternative is to use approaches such as address based sampling (ABS) to collect the telephone of the respondent for completing the interview though the telephone. This approach is currently tested in two communities in 2012.

6. References

California Health Interview Survey (2011a). *CHIS 2009 Methodology Series: Report 5 – Weighting and Variance Estimation*. Los Angeles, CA: UCLA Center for Health Policy Research.

California Health Interview Survey (2011b). CHIS 2009 Methodology Series: Report 1 - Sample Design. Los Angeles, CA: UCLA Center for Health Policy Research,

Flores Cervantes, I., and Kalton, G. (2008). Methods for sampling rare populations in telephone surveys. In *Advances in Telephone Survey Methodology*. (Eds., J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. de Leeuw, L. Japec, P. Lavrakas, M.W. Link and R.L. Sangster). Hoboken, NJ: Wiley, 113-132.