

## A City-Based Design That Attempts to Improve National Representativeness of Asians

*Steven Pedlow*<sup>1</sup>

This article describes a case study on the potential of using smaller geographical units in an area probability design, and reports the challenges of collecting a nationally representative sample for this hard-to-reach population. The Census Integrated Communications Program Evaluation (CICPE) was designed to evaluate the promotional campaign's effect on Decennial Census participation for six race/ethnicity groups of interest. A nationally representative Core sample was designed to collect interviews for Hispanics, non-Hispanic African-Americans, and non-Hispanic Whites. However, it was impractical to include the rarer Asian, American Indian and Alaska Native (AIAN), and Native Hawaiian and Other Pacific Islander (NHOPI) populations in the Core design. For the Asian sample, we designed a separate area probability sample.

Traditional area probability sampling designs use counties or metropolitan areas as first-stage units, but smaller geographical units can better target hard-to-reach populations. The CICPE Asian sample used cities as the first-stage units.

*Key words:* Area probability sampling; CICPE; Decennial Census; Hard-to-Reach.

### 1. Introduction

Every ten years, the U.S. Census Bureau attempts to count every American through the Decennial Census. For the 2000 Decennial Census, the Census Bureau responded to declining mail participation in the 1990 Decennial Census (which had increased the costs of in-person enumeration visits) with a greatly expanded outreach and promotion campaign called the “Partnership and Marketing Program” (PMP). NORC at the University of Chicago was contracted to conduct the 2000 Partnership and Marketing Program Evaluation (PMPE), an independent evaluation of the PMP which included a series of three face-to-face in-person surveys: before the PMP (autumn 1999); during the PMP, which was also the time period that included the mailing of Census forms to housing units and the beginning of data collection by mail (winter/spring 2000); and after the PMP during the nonresponse follow-up operation of the 2000 Decennial Census (summer 2000), which involved in-person visits to housing units that did not mail back their Census questionnaire. The Census Bureau was sensitive to differential impact of the PMP by race/ethnicity, and so the sample was equally divided among six different race/ethnicity groups, including Asians.

In 2010, the Census Bureau took the lessons learned from 2000 and designed an Integrated Communications Program (ICP) to encourage mail participation in the 2010

<sup>1</sup> NORC/University of Chicago – Statistics/Methodology, 55 E Monroe St., Suite 2000, Chicago, IL 60603, U.S.A. Email: [pedlow-steven@norc.uchicago.edu](mailto:pedlow-steven@norc.uchicago.edu)

Decennial Census. NORC at the University of Chicago again conducted an independent evaluation of the ICP called the “Census Integrated Communications Program Evaluation” (CICPE), which again utilized three waves of in-person face-to-face interviewing that matched up to before, during, and after the Census ICP. The same six race/ethnicity groups from the 2000 evaluation were again of interest. In 2000 and 2010, NORC’s sample designs included a Core sample (See Section 2 for details) that was a nationally representative area probability sample to collect interviews from Hispanics, non-Hispanic African-Americans, and non-Hispanic Whites. However, supplemental samples were necessary for the remaining three race/ethnicity groups: Asians (the focus of this article), American Indian and Alaska Natives (AIAN), and Native Hawaiian and Other Pacific Islanders (NHOPI).

For the 2000 PMPE, NORC’s Asian Supplemental Sample collected all interviews from the five U.S. cities with the largest Asian populations. For the 2010 CICPE, we wanted a more nationally representative sample of Asians. This article describes how we used a city-based area probability sample to greatly increase the coverage of our Asian sample. The tradeoff between noncoverage error and screening cost has already been widely recognized and discussed in many books and papers, but it seems that the screening costs for Asians (4.2 percent of the national population) and other hard-to-reach groups as well as the difficulty of targeting Asians (and other hard-to-reach groups) at larger geographies has prevented such groups from being studied with nationally representative studies.

Prior to the CICPE, no independent surveys had attempted a nationally representative sample of Asians. The only two surveys that have attempted to be nationally representative for Asians previous to the CICPE were embedded within much larger studies that allow a minority of a large number of second-stage units to oversample Asians. The National Health and Nutrition Examination Survey (NHANES) started to oversample Asians in 2011 to create nationally representative estimates, but it does this by oversampling second-stage units with high Asian eligibility rates within a survey that completes 5,000 interviews per year. The second-stage units with high Asian eligibility rates are in areas where many other second-stage units with low Asian eligibility rates have been selected. The latest NHANES sample documentation (Curtin et al. 2013) describes that 1,440 second-stage units were used from 2007-2010 (sample documentation for 2011-2014 is not yet available). Only a small percentage of the second-stage units in 2011-2014 will produce Asian respondents, but the large size of the overall study still allows a sizable and representative sample of Asians. The second nationally representative survey of Asians prior to CICPE is the National Latino and Asian American Study (NLAAS), whose sample design was completely integrated with the National Comorbidity Survey Replication (NCS-R) national sample design by selecting 474 pairs of segments in the same areas: one to be nationally representative and one to oversample Asians or Latinos (Heeringa et al. 2004). The Collaborative Psychiatric Epidemiology Surveys website ([http://www.icpsr.umich.edu/icpsrweb/CPES/about\\_cpes/sample\\_design.jsp#nlaas](http://www.icpsr.umich.edu/icpsrweb/CPES/about_cpes/sample_design.jsp#nlaas)) concedes that without this pairing, survey costs would have been prohibitively high for the NLAAS. The NLAAS used only 317 second-stage units because they expected near-zero interviews in 157 of their second-stage units. Both of these surveys use counties as the smallest first-stage unit for selection. The CICPE shows that nationally representative surveys of Asians and other hard-to-reach groups can be attempted by using first-stage units smaller than counties

without need for a larger survey to support them because these groups can be better targeted with smaller first-stage units.

Looking ahead to the rest of this article, Section 2 will discuss the 2000 PMPE and 2010 CICPE sample designs, including a review of Area Probability Sampling. Section 3 will describe the details of the 2010 CICPE Asian sample design. Section 4 will present some results from fielding the 2010 CICPE Asian sample. Section 5 shows some demographic comparisons between the 2010 CICPE and the national Asian population. Section 6 discusses the limitations to our study, especially with regard to its national representativeness. Finally, Section 7 summarizes this article.

## **2. The 2000 PMPE and 2010 CICPE Sample Designs**

Both the 2000 PMPE and the 2010 CICPE had sample designs that included three waves of data collection. The first wave of data collection took place before the main campaign elements, the second wave took place while the campaign peaked, and the third wave took place after the mail participation deadline to avoid in-person follow-up. Both designs had a sample size that was an idealized 3,000 interviews per wave divided equally among six race/ethnicity groups: Hispanics, non-Hispanic African-Americans, non-Hispanic Whites, American Indian and Alaska Natives (AIAN), Native Hawaiian and Other Pacific Islanders (NHOPI), and Asians. The second wave of the 2010 CICPE design was compressed into a shorter time period, so the sample size was dropped to 2,100.

A nationally representative area probability sample called the “Core” sample was designed to collect interviews for the three largest race/ethnicity groups: Hispanics, non-Hispanic African-Americans, and non-Hispanic Whites. National coverage as part of this Core sample was impractical for the three smaller race/ethnicity groups, so three supplemental samples were necessary. This article focuses on the Asian Supplemental Samples for the 2000 PMPE and 2010 CICPE.

At the time of the 2010 CICPE sample design, the latest source of information on local Asian populations was the 2000 Decennial Census, as the American Community Survey had not yet released small area data. According to the 2000 Census, there were 11,898,828 U.S. Non-Hispanic Asians (Barnes and Bennett 2002), alone or in combination, comprising 4.2 percent of the U.S. population at the time. This figure includes those who marked Asian, regardless of whether other race boxes were marked on the census form; the 2000 Census was the first Decennial Census where race was asked using a “mark all that apply” format.

Since there are many more Hispanics, non-Hispanic African-Americans, and non-Hispanic Whites in the U.S. population than there are Asians, collecting enough Asian interviews through the Core sample would require impractically large screening samples with heavy subsampling of the eligibles for the higher population race/ethnicity groups. In fact, national coverage itself was considered impractical during the planning of the 2000 PMPE, as shown by Wolter et al. (2002), which collected all of its supplemental Asian interviews from the five U.S. cities with the largest Asian populations: New York, Los Angeles, San Francisco, Chicago, and Seattle. At the time of the 2000 Census, 18.8 percent of the U.S. Asian population lived in these five cities. This meant that the 2000 PMPE did not attempt to achieve a nationally representative Asian sample, since the coverage of that sampling frame was only 18.8 percent of the U.S. Asian population. Within these five cities,

6.5 percent of the population was Asian. If the sample was equal probability within these cities, the eligibility rate for the Asian sample would be 6.5 percent. We can also refer to this rate as the screening “hit rate.”

For the 2010 CICPE design, our intention was to improve coverage through a national design. Most national face-to-face surveys in the United States use a multi-stage area probability (AP) sampling design that selects clusters of housing units to interview in order to reduce data-collection costs (Kish 1965). In a multi-stage AP sampling design, a set of large clusters are first selected (first-stage units). Within the selected large clusters, sets of small clusters are selected (second-stage units). Finally, within these selected small clusters, individual housing units are selected for interviewing. The basic objective for a multi-stage AP sampling design is a nationally representative equal-probability sample permitting optimal statistical efficiency. To achieve this, AP samples use probability proportional to size (PPS) sampling in which “larger” areas have a greater selection probability. The measure of size often used for the probabilities is the number of housing units, usually derived from Census data.

Most national area probability samples have first-stage units that are county based, often even using larger metropolitan statistical areas (MSAs) where present (Lohr 2009). However, the key idea in this article is that hard-to-reach populations are better targeted at small geographies. Within the large first-stage unit areas for typical area probability designs, the smaller second-stage areas are often block based, either in terms of blocks, block groups, or entire census tracts. Of course, national samples cannot use first-stage geographies as small as individual blocks; this would require too many clusters that are too spread out to be cost effective. However, NORC has a history of using smaller geographies as first-stage units to better oversample race/ethnicity groups.

The National Longitudinal Survey of Youth 1979 cohort (NLS79) obtained a nationally representative set of interviews with youths who were 14-21 years old while oversampling Hispanic and African-American youths. To do this, NORC split the task into two parts. First, a nationally representative area probability sample was used to get a nationally representative mix of Hispanic, non-Hispanic African-American, and non-Hispanic non-African-American youths. A second area probability sample was used to obtain only Hispanic and African-American youths. This “Supplemental” sample did not differ in its design from the nationally representative “Cross-Sectional” sample, but different areas were selected to better target Hispanic and African-American youths.

The National Longitudinal Survey of Youth 1997 cohort (NLSY97) took this design one step further (Moore et al. 2000). In the “Supplemental Sample” design, all first-stage units were counties rather than entire MSAs in urban areas. Remembering that our goal was to oversample Hispanic and African-American youths, MSAs often have central city counties with a high rate of minority youths surrounded by outlying, more rural areas with lower concentrations of Hispanic and African-American youths. Our strategy allowed us to separate counties with many minority youths from surrounding counties in the same MSA with fewer of them. Counties were still considered too large to target Asians, so we used places defined by the Census as the first-stage clusters in the 2010 CICPE sample design. Places defined by the Census include cities, towns, villages, as well as other “census-designated places,” but we will simplify our language in this article and refer to our first-stage clusters as “cities.”

### 3. The 2010 CICPE Asian Sample

Our first task in selecting a city-based Asian sample was to construct a sampling frame of cities. The most recent data available at the time was still the 2000 Decennial Census. Our first step was to set a threshold of 1,000 Asians for a city to be included, which led to a set of 1,261 U.S. cities that included 75.6 percent of all U.S. Asians. While our universe did not represent 100 percent coverage, it did represent a substantial increase over the 18.8 percent coverage for the 2000 PMPE Asian sample design. Within these 1,261 cities, the population is 7.8 percent non-Hispanic Asians. This population percentage of non-Hispanic Asians (7.8 percent) is our estimated eligibility rate, even though this may differ from the actual eligibility rate based on the project's protocol to interview the person most likely to handle the incoming mail. [Table 1](#) gives the coverage and estimated eligibility rates for the many different non-Hispanic Asian population thresholds that we could have used for our sampling frame of cities.

[Table 1](#) shows that as the threshold decreases, the coverage increases while the eligibility rate decreases. [Table 1](#) also shows that 12.71 percent of Asians live outside of cities (in unincorporated places, including rural areas). So it is not possible to achieve 100 percent coverage without sampling unincorporated places. As the eligibility rate decreases, more screening becomes necessary to find the same number of eligible households. Though coverage could be higher with county-level sampling, the eligibility rates would be so much lower that such a sample would be cost prohibitive.

Keeping in mind our goal of 500 Asian interviews in each wave, we needed to balance the number of cities where we would have to hire staff against the cluster size determined by the average number of interviews we would need to collect in each city. Increasing the number of cities would increase the cost, while decreasing the number of cities would increase the clustering and therefore the design effect. Balancing these two factors, we decided to select a representative sample of 25 cities (requiring an average of 20 interviews per city) with probability proportional to the city population of non-Hispanic Asians. Our design gave every Asian in our frame of 1,261 cities an equal chance of being in

*Table 1. Threshold options for the Asian frame of cities*

Minimum Number of Non-Hispanic Asians	Eligible Cities	Asian Population Coverage (%)	Eligibility Rate (%)
100,000	8	20.46	12.28
50,000	18	25.92	12.69
25,000	46	34.07	12.30
10,000	153	48.12	10.58
5,000	321	57.87	9.67
2,500	653	67.59	8.72
1,000	1,261	75.57	7.83
500	2,059	80.32	7.04
250	3,015	83.17	6.50
100	4,569	85.29	5.97
50	6,039	86.16	5.68
25	7,823	86.69	5.46
1	18,608	87.29	5.05

one of our selected cities. New York and Los Angeles, two of the cities used for the 2000 PMPE design, were selected with certainty because they each contain more than 1 in 25 of the Asians within our sampling frame (359,684). Every other city in our frame had a chance of selection equal to their Asian population divided by 359,684. Of the three other cities used in the 2000 PMPE, two (San Francisco and Chicago) were selected while Seattle was not.

We also wanted representativeness among different Asian subgroups. Census 2010 population counts are available at the place level for the following subgroups: Indian Asians, Chinese Asians (excludes Taiwanese), Filipino Asians, and Other Asians. To make sure that all four subgroups were represented, we divided the non-certainty cities into four equal groups: 1) those where at least 26.83 percent of Asians were Indian Asians, 2) those where at least 24.27 percent of Asians were Chinese Asians, 3) those where at least 24.84 percent of Asians were Filipino Asians, and 4) those where at least 54.77 percent of Asians were Other Asians. Within these four groups, the sampling frame was serpentine sorted by Census Region, State, and Asian Population Percentage. [Table 2](#) shows the distribution of respondents by Asian subgroup, which includes the three subgroups used during sampling, as well as three other major subgroups and an “Other” category. [Table 2](#) also includes the Census 2010 distribution among Asian Americans ([Hoeffel et al. 2012](#)).

Due to the Census data limitations, only the top three Asian subgroups in [Table 2](#) could be controlled in our selection process. [Table 2](#) shows that we had fewer Asian Indian respondents and more Filipino respondents than Census 2010 would lead us to expect. The surplus of Chinese respondents can be partly explained by the Census exclusion of Taiwanese Chinese (our question did not separate them). Among the “Other” subgroups, our total sum is quite close to the Census sum, though our sample did have more Japanese respondents and fewer of the rarer Other groups.

Eleven of our twenty-five selected cities were in California. Other states with more than one city selected were Hawaii, New York, and Texas. [Table 3](#) gives details on the 25 selected cities.

The Asian population sizes in our 25 cities range from the minimum of 1,000 to around 850,000, with a median of approximately 23,000. Asian population percentages range from under two percent to over 70 percent, with a median of 12.86 percent.

We then selected entire census tracts as our second-stage clusters within our selected cities. We decided to select five tracts from each noncertainty city so that we would have

*Table 2. Distribution of 2010 CICPE Asian sample by Asian subgroup*

Asian Subgroup	Wave 1 (%)	Wave 2 (%)	Wave 3 (%)	Census 2010 (%)
Asian Indian	11.9	13.1	13.7	18.4
Chinese (excl. Taiwanese)	21.3	26.5	24.2	21.8
Filipino	28.7	26.2	22.8	19.7
Japanese	20.5	20.3	20.3	7.5
Korean	7.0	4.3	7.5	9.9
Vietnamese	7.4	5.6	6.8	10.0
Other	3.3	4.0	4.6	12.7

Table 3. The 25 cities in the 2010 CICPE Asian sample

City	State	Asian Population	Asian Population Percentage (%)
New York city	New York	857,094	10.70
Los Angeles city	California	396,352	10.73
San Jose city	California	252,818	28.25
San Francisco city	California	250,364	32.23
Honolulu CDP*	Hawaii	244,698	65.84
Chicago city	Illinois	137,039	4.73
Houston city	Texas	111,511	5.71
Sacramento city	California	74,634	18.34
Philadelphia city	Pennsylvania	73,403	4.84
Stockton city	California	52,631	21.59
Anaheim city	California	42,171	12.86
Arcadia city	California	24,886	46.91
Hilo CDP*	Hawaii	23,206	56.94
Bellevue city	Washington	20,741	18.93
Riverside city	California	16,311	6.39
Bakersfield city	California	12,036	4.87
Lakewood city	California	11,870	14.96
Canton CDP*	Michigan	7,252	9.50
Pacifica city	California	6,805	17.73
Aiea CDP*	Hawaii	6,423	71.22
Syracuse city	New York	5,566	3.78
Port Arthur city	Texas	3,546	6.14
Stafford city	Texas	3,272	20.87
Marlborough city	Massachusetts	1,514	4.18
Rio Rancho city	New Mexico	1,006	1.94

\*CDP = Census Designated Place

approximately 125 tracts for 500 interviews (an average clustering of four interviews per tract). This means that every tract selected represents 359,684/5, or roughly 72,000 Asians. Since New York and Los Angeles both have more than 359,684 Asians, we selected extra tracts for them. We selected ( $5 * 857,094/359,684 = 11.9$ ) twelve tracts in New York and ( $5 * 396,352/359,684 = 5.5$ ) six tracts in Los Angeles. Only one selected city contains less than five Census tracts to select from; the Aiea CDP only contains two tracts, so both were selected (and each had  $5/2 = 2.5$  times as many housing units selected). Thus we selected a total number of 130 Census tracts, which resulted in an average of 3.8 interviews per selected Census tract. The selected tracts had an even larger range of Asian population percentages, ranging from 87 percent to less than one percent. An equal probability sample using these 130 Census tracts would have resulted in a sample eligibility rate of 7.8 percent (almost twice the national eligibility rate), but this can be increased by oversampling in tracts with a higher proportion of Asian residents. We actually designed a sample with an expected eligibility rate of 26 percent, but this required some tracts to be oversampled by a factor of 50. Differential sampling weights that result from such a skewed oversample would have created a large design effect, which would have greatly reduced our effective sample size.

Table 4. Planned and actual unweighted eligibility rates

Statistic	2000 PMPE (%)	2010 CICPE (%)
Eligibility Rate		
First-Stage	6.5	7.8
Planned	unknown	12.5
Wave 1	22.2	10.3
Wave 2	13.3	12.9
Wave 3	18.9	8.5

As in most statistical design decisions, the amount of oversampling involved a balance between lowering screening costs versus keeping variance due to differential weighting low. With help from the Census Bureau, we agreed to limit the design effect so that the loss in effective sample size due to differential sampling would be no greater than 20 percent (a design effect due to differential sampling no greater than 1.25). Our approach decreased the differential oversampling from a factor of 50 to a factor of 3. In so doing, we incurred higher screening costs, but maintained the effective sample size closer to the number of interviews. Our specific strategy was to oversample tracts with eligibility rates of at least 20 percent by a factor of 3, and to oversample tracts with eligibility rates between 10 and 20 percent by a factor of 2. With this strategy, our estimated eligibility rate was 12.5 percent, three times as large as the national eligibility rate.

#### 4. 2010 CICPE Asian Sample Field Results

We were able to meet our sample targets for the Asian sample, but our actual unweighted eligibility rates were lower than our estimate for two out of the three waves. This is not surprising when the observed eligibility rate is lower than the planned eligibility rate. Households that are eligible are the most difficult households at which to achieve cooperation (even at the screener level), so eligible households often have a lower screener response rate than ineligible households. This results in a lower eligibility rate instead of a lower response rate, so this is often referred to as “hidden interview nonresponse” within the screener nonresponse. Table 4 shows that the 2000 PMPE achieved higher eligibility rates, which were due to a higher level of oversampling. The 2000 PMPE Asian design oversampled areas with eligibility rates of at least 20 percent by a factor of 5 (Wolter et al. 2002).

Table 5 compares the weighted response rates from 2010 CICPE against the unweighted 2000 PMPE response rates.

Weighted rates are not available from the 2000 PMPE, but unweighted rates are not appropriate for the 2010 CICPE because of the mixed-mode data collection procedures

Table 5. Response rates for 2000 PMPE and 2010 CICPE

Wave	2000 PMPE (Unweighted) (%)	2010 CICPE (Weighted) (%)
Wave 1	57.2	50.7
Wave 2	71.0	64.2
Wave 3	60.8	73.8

that included subsampling of nonrespondents for in-person follow-up. Nevertheless, the average response rates for both studies are around 63 percent, and both studies have their lowest response rate in the first wave. The response rates are higher for the 2000 PMPE in the first two waves, but the response rate is much higher in the third wave for the 2010 CICPE.

Nonresponse bias is usually immeasurable, but the 2010 CICPE study is an exception. With Census Bureau cooperation, we were able to match the entire set of our selected households to 2010 Decennial Census response data. Nonresponse bias was an important issue for the 2010 CICPE since it was designed around probable/actual response to the 2010 Decennial Census, and it is logical to think that nonrespondents to our survey would be more likely to be nonrespondents to the 2010 Decennial Census. Table 6 gives the actual mail response rates to the 2010 Decennial Census by April 18 for three different types of 2010 CICPE respondents, as well as for interview nonrespondents and those households for which we could not determine eligibility (screener nonrespondents). We used the April 18 date cutoff because this marks the start of the in-person follow-up effort. Decennial forms are mailed near the end of March, Census Day is April 1, and any households returning their mail forms after April 18 may still have been visited in-person. Minimum cost is achieved for households whose mail questionnaire is received prior to April 18.

The three types of CICPE respondents are: 1) Refusers – those respondents who were (soft) refusals at one time, 2) Difficult Respondents – those respondents who had more than the median number of visits before responding, and 3) Easy Respondents – those respondents who responded after less than the median number of visits. All of the mail response rates in Table 6 are weighted. Table 6 shows that our respondents did have higher mail return rates. As expected, the easy Asian respondents had the highest mail return rate by April 18 (64.3 percent) while the nonrespondents had the lowest mail return rate by April 18 (53.0 percent). We estimated the eligible proportion among those with unknown eligibility status and counted them as nonrespondents. Combining the two nonrespondent categories together, the mail return rate was 53.6 percent. Combining the three respondent categories together, the mail return rate was 62.7 percent. Combining all five categories together, the mail return rate for our entire Asian sample was 59.4 percent.

The nonresponse bias is the difference between the estimate for only the respondents (62.7 percent) and the entire population of interest, represented by the entire sample (59.4 percent). Therefore, our Asian sample’s nonresponse bias is  $62.7 - 59.4 = 3.3$  percent.

*Table 6. Mail response rates for the Asian CICPE sample (weighted)*

Outcome	Return Rate (%)	Response Status (%)	ALL (%)
Unknown Eligibles	54.5	Nonrespondents:	ALL:
Nonrespondents	53.0	53.6	59.4
Respondents – Refusers	61.9	Respondents:	
Respondents – Difficult Respondents – Easy Respondents	62.2 64.3	62.7	

Since this is positive, our respondents are more likely to be mail responders by April 18 than our nonrespondents, which is the expected direction of our nonresponse bias. Interestingly, of our six race/ethnicity groups, four had almost no nonresponse bias (our estimates of nonresponse bias were less than one percent), while our American Indian and Alaska Native sample had a negative nonresponse bias, meaning that respondents were less likely to be mail responders by April 18 than our entire sample of American Indian and Alaska Native housing units (Datta et al. 2012 has more details for all six race/ethnicity groups).

## 5. Demographic Comparisons to the National Asian Population

We collected demographic data from our respondents, which allows us to compare our set of respondents with national control totals to examine how representative our sample is. Table 2 (above) already showed that our sample has more Filipino and Japanese respondents and fewer Asian Indian respondents than the national totals. Table 7 compares the age and gender distribution of our respondents against Current Population Survey Annual Social and Economic Supplement data from March 2010.

We had expected age and gender to be skewed by the project's protocol that interviewed the person most likely to handle the incoming mail. However, Table 7 shows that the CICPE Asian Sample gender distribution is very close to the national average of 47.1 percent male, 52.9 percent female. Our respondents are more likely to be 65 years of age or older and less likely to be 18–29 years of age, but our distribution is consistent with most surveys that have lower response rates among this young age group and higher response rates for senior citizens.

As an additional check on the representativeness of our Asian sample, Table 8 shows the highest degree earned by our 2010 CICPE Asian Sample respondents and the

Table 7. Distribution of 2010 CICPE Asian sample by age and gender

Gender/Age	Wave 1 (%)	Wave 2 (%)	Wave 3 (%)	CPS, 2010 (%)
Male 18-29	19.7	20.1	18.0	24.6
Male 30-44	31.6	34.3	35.2	33.7
Male 45-64	32.1	29.6	29.7	30.9
Male 65 +	16.7	16.0	17.2	10.8
<b>Total Male</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
Female 18-29	16.8	11.8	13.5	22.4
Female 30-44	33.7	35.0	34.1	32.8
Female 45-64	29.4	33.6	33.0	31.4
Female 65 +	20.1	19.5	19.5	13.4
<b>Total Female</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
ALL 18-29	18.3	15.1	15.5	23.4
ALL 30-44	32.4	34.2	34.1	33.2
ALL 45-64	31.3	33.0	31.9	31.1
ALL 65 +	18.1	17.6	18.5	12.2
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
<b>Percentage Male</b>	<b>46.2</b>	<b>43.4</b>	<b>49.3</b>	<b>47.1</b>

Table 8. Distribution of 2010 CICPE Asian sample by highest degree

Highest Degree	Wave 1 (%)	Wave 2 (%)	Wave 3 (%)	ACS 2009 (%)
No High School Diploma	6.6	6.2	7.7	15.0
High School Diploma	36.8	34.3	33.3	35.0
College Degree	35.8	35.8	35.3	30.0
Graduate/Professional Degree	20.8	23.7	23.7	20.0
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

corresponding national percentages from the 2009 American Community Survey (U.S. Census Bureau 2011).

Table 8 shows that we have fewer respondents with no high school diploma than the American Community Survey, but our distributions are similar otherwise.

## 6. Limitations

Some compromises were made in order to limit the cost of the Asian sample, the largest of which is that the frame did not attain 100 percent coverage due to the minimum size requirement. Since we needed an average of 20 interviews per city per wave, we did need to set a minimum size, but we set the minimum size at 1,000 Asians in order to keep the eligibility rate to almost 8 percent. Setting a lower minimum size would have resulted in additional screening costs. In fact, Table 1 shows that using cities as the first-stage sampling units limits the coverage to 87 percent. To achieve 100 percent coverage, a sample of unincorporated areas would be necessary.

One minor limitation is that we did have larger clusters in the Aiea, Hawaii CDP since the interviews there were collected in only two Census tracts rather than five because there are only two tracts in the Aiea CDP. Clustering could also have been reduced if we had selected more than 25 cities, but adding field staff in more locations would have created significant additional costs.

Finally, our nonresponse bias does show that Asians are the only race/ethnicity group with a bias larger than one percent. One possible explanation is that we only collected interviews in English and Spanish and so the Asians were likely to have the highest language barrier among the six race/ethnicity groups. If English-speaking Asians were more likely to respond to the Census, this would result in a positive nonresponse bias as observed.

## 7. Summary

Even for a hard-to-reach population, it may be possible to attempt a nationally representative sample if the population can be targeted by local areas. Asians make up only 4.2 percent of the U.S. population, but are more clustered by city than larger first-stage sampling units such as counties or metropolitan areas. We have achieved a 75.6 percent coverage rate for a national sample of Asians for the Census Integrated

Communications Program Evaluation (CICPE) study by selecting 25 cities from a frame of 1,261 U.S. cities with a population of at least 1,000 Asians. While we could not come closer to 100 percent coverage in a cost-effective way and expensive screening was still necessary, we believe that our data is more representative of U.S. Asians than the 2000 PMPE study taking place in only five U.S. cities as well as any list-assisted telephone survey using Asian surnames or other methods with unknown biases (Davern et al. 2007). Higher coverage is possible with county-level sampling, but with eligibility rates that are lower enough to make it cost prohibitive. This is why nationally representative samples of Asians have not been previously attempted without a larger survey structure for support. While the central idea of this article may not be a theoretical breakthrough, it has not been acted on previously, greatly reducing the ability to collect nationally representative data on Asians and other hard-to-reach groups. If samplers could break away from the “rule” that first-stage units must be counties, area probability samples could more flexibly collect nationally representative data for a wider range of applications.

## 8. References

- Barnes, J. and Bennett, C. (2002). The Asian Population: 2000, a Census 2000 Brief. Washington, D.C.: Bureau of the Census. Available at: <http://www.census.gov/prod/2002pubs/c2kbr01-16.pdf> (accessed December 2013).
- Curtin, L., Mohadjer, L., Dohrmann, S., Kruszan-Moran, D., Mirel, L., Carroll, M., Hirsch, R., Burt, V., and Johnson, C. (2013). National Health and Nutrition Examination Survey: Sample Design, 2007-2010, National Center for Health Statistics. Vital and Health Statistics, Series 2, Number 160. Available at: [http://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_160.pdf](http://www.cdc.gov/nchs/data/series/sr_02/sr02_160.pdf) (accessed December 2013).
- Datta, A., Yan, T., Evans, D., Pedlow, S., Spencer, B., and Bautista, R. (2012). The 2010 Census Integrated Communications Program Evaluation (CICPE) Final Report. Washington, D.C.: Bureau of the Census. Available at: [http://www.census.gov/2010census/pdf/2010\\_Census\\_ICP\\_Evaluation.pdf](http://www.census.gov/2010census/pdf/2010_Census_ICP_Evaluation.pdf) (accessed December 2013).
- Davern, M., McAlpine, D., Ziegenfuss, J., and Beebe, T. (2007). Are Surname Telephone Oversamples an Efficient Way to Better Understand the Health and Healthcare of Minority Group Members? *Medical Care*, 45, 1098–1104.
- Heeringa, S., Wagner, J., Torres, M., Duan, N., Adams, T., and Berglund, P. (2004). Sample Designs and Sampling Methods for the Collaborative Psychiatric Epidemiology Studies (CPES). *International Journal of Methods in Psychiatric Research*, 13, 221–240, DOI: <http://www.dx.doi.org/10.1002/mpr.179>.
- Hoeffel, E., Rastogi, S., Kim, M.O., and Shahid, H. (2012). The Asian Population: 2010, a Census 2010 Brief. Washington, D.C.: Bureau of the Census. Available at: <http://www.census.gov/prod/cen2010/briefs/c2010br-11.pdf> (accessed December 2013).
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Lohr, S. (2009). *Sampling: Design and Analysis (Second Edition)*. Pacific Grove, CA: Duxbury Press.
- Moore, W., Pedlow, S., Krishnamurty, P., and Wolter, K. (2000). The National Longitudinal Survey of Youth 1997 (NLSY97) Technical Sampling Report. Chicago:

- NORC at the University of Chicago. Available at: <http://www.bls.gov/nls/nlsy97techsamp.pdf> (accessed December 2013).
- U.S. Census Bureau (2011). Profile America Facts for Features, Asian/Pacific American Heritage Month: May 2011. Washington, D.C.: Bureau of the Census. Available at: [http://www.census.gov/newsroom/releases/archives/facts\\_for\\_features\\_special\\_editions/cb11-ff06.html](http://www.census.gov/newsroom/releases/archives/facts_for_features_special_editions/cb11-ff06.html) (accessed December 2013).
- Wolter, K., Calder, B., Malthouse, E., Murphy, S., Pedlow, S., and Porras, J. (2002). Census 2000 Evaluation: Partnership and Marketing Program Evaluation. Washington, D.C.: Bureau of the Census. Available at: <http://www.census.gov/pred/www/rpts/D.1.PDF> (accessed December 2013).

Received February 2013

Revised November 2013

Accepted January 2014