Latino Methodology



53 West Baltimore Pike

Media, Pennsylvania 19063-5698

I N T E R N A T I O N A L C O M M U N I C A T I O N S R E S E A R C H

#7667

**KAISER/PEW 2002 LATINO SURVEY**

#### METHODOLOGY REPORT

In order to fully represent the opinions of Latino people living in the United States, ICR conducted interviews with a statistically representative sample of Latinos so that they could be examined nationally, in target regions of high Latino concentration, and by country of origin. Additionally, in order to make comparisons between Latino and non-Latino populations, the study also included a non-Latino quota segment on a national basis.

This study was conducted by telephone April 4 – June 11, 2002 among a nationally representative sample of 4,213 adults 18 years and older, including 2,929 Latinos and 1,284 non-Latinos. Amongst the non-Latinos were 1,008 non-Latino white adults and 171 non-Latino black adults. The error margins for each group are +/- 2.73 points (Total respondents), +/- 2.41 points (Latinos), +/- 3.04 points (non Latinos), +/- 3.32 points (non-Latino whites), and +/- 9.9 points (non-Latino blacks).

The study included interviews with a representative number of Latinos in 6 ethnic groups, based on country of origin: Mexican, Puerto Rican, Cuban, Salvadoran, Dominican, and Colombian.

Interviews were conducted with 1047 Mexicans, 317 Puerto Ricans, 343 Cubans, 204 Salvadorans, 235 Dominicans, and 214 Colombians. In total, 341 Central Americans and 394 South Americans were interviewed. The error margins for each group are +/- 3.31 points (Mexicans), +/- 6.65 points (Puerto Ricans), +/- 6.38 points (Cubans), +/- 10.11 points (Salvadorans), +/- 7.30 points (Dominicans) and +/- 10.45 points (Colombians). The error margin for Central Americans is +/- 8.08 points and for South Americans is +/- 7.95 points.

For this survey, ICR maintained a staff of Spanish-speaking interviewers whom, when contacting a household, were able to offer respondents the option of completing the survey in Spanish or in English. Fifty-seven percent of the Latino interviews were conducted predominantly in Spanish.

**Eligible Respondent**

For the Latino portion of the study, the survey was administered to any male or female age 18 and older that are of Latino origin or descent. The non-Latino portion of the study was conducted with any male or female age 18 and older that are of any non-Latino origin or descent.

##### Field Period

The field period for this study was April 4– June 11, 2002. The interviewing was conducted by ICR/International Communications Research in Media, PA. All interviews were conducted using the Computer Assisted Telephone Interviewing (CATI) system. The CATI system ensured that questions followed logical skip patterns and that the listed attributes automatically rotated, eliminating “question position” bias.

##### Sampling Methodology

The sample design employed in this project can best be described as a highly stratified, disproportionate RDD sample of the 48 contiguous states. Design considerations incorporated a number of objectives: a sample of the Hispanic adult population; oversamples of particular Latino subgroups defined by country of origin/background (COO); and, complimentary samples of African American (AA) and White/Other adult populations. A highly stratified telephone RDD sample was utilized due to the relative ease by which populations can be segregated by geography and Hispanic population density estimated reliably. This approach also insured strict probability sampling using an epsem RDD sample within sampling strata.

The sample utilized four levels of stratification. Primary stratification was at the regional level, using the four basic Census-defined geographies. Within region, geographic substrata were defined for the seven most populous Hispanic states: Arizona, California, Florida, Illinois, N ew Jersey, New York, and Texas. These seven states all contain Latino populations in excess of one million and in total over 80% of all Hispanics in the continental US. This basic stratification thus resulted in a total of eleven geographic sampling strata comprised of the seven high prevalence Latino States plus the residual population within each of the four regions.

The third level of stratification was accomplished at the telephone exchange-level (NPA-NXX). NXXs, within each of the eleven geographic strata were arrayed by Latino population incidence and segregated into three independent NXX-defined strata. Nominally, these NXX-strata correspond to High, Medium and Low incidence of Hispanic households (HHs). The NXX-level estimates of Latino HHs were those maintained in the GENESYS NXX-database; estimates of race/ethnicity are derived from the known geographic correspondence of each NXX in combination with estimates of the demographic characteristics of the relevant geographic components. Geographic correspondence being derived from the distribution of listed households, demographic characteristics from annual estimates of Claritas, Inc.

Table 1 details the basic geographic stratification along with the final NXX-level prevalence strata and corresponding Latino densities within each of the thirty-three sampling strata. For operational purposes the high incidence strata within each region were further stratified to identify NXXs in the range of >75%+ expected incidence and 50-75%. These groups were contacted with bilingual interviewers.

This basic design was also employed to secure the representative samples of White/Other and African-American households. Operationally this was accomplished by using a series of distinct replicates within each of the thirty-three geographic strata. As mentioned previously, the GENESYS data system was used to establish estimates of the Latino population for stratification purposes. At the same time, parallel estimates of the African American and White/Other populations were developed and utilized to establish sample allocations and replicate sizes for each stratum so as to establish a proportionate sample of each subgroup. So for example, in the first set of replicates all households were interviewed regardless of racial/ethnicity make-up; in the second set, White/Other households were terminated with only African American and Latino households interviewed, and in all remaining replicates, only Latino households were interviewed. This procedure enables us to secure approximately proportionate samples of the non-Latino population while maintaining one coherent stratification plan.

Allocation of Latino interviews across the NXX-based strata within each of the eleven Regional-State strata was determined using an optimal allocation scheme based on the incidences in Table 1.

Another objective of this project was to provide minimal sample sizes by country of origin. Specifically, we were interested in obtaining minimum sample sizes among Cubans, Columbians, Dominicans, and Salvadorans. In order to accomplish this we need to establish four independent RDD samples, each targeting high prevalence census tracts for one of the four groups. However, difficulties with the Census 2000 Hispanic country of origin question required us to first recast estimates for some of the subgroups.

It was initially presumed that the latest decennial census would provide the level of detail for the country of origin subgroups. However, a problem with the instructions for completing the census instrument has resulted in a substantial portion of the Latino population to be unclassifiable as to background and/or country of origin. The specific problem was a result of the instructions on the printed form. If Hispanic, the form queried specifically as to whether the person was (1) Mexican/Mexican American/Chicano, or (2) Puerto Rican, or (3) Cuban or OTHER SPANISH/HISPANIC/LATINO with a follow-up instruction to record the “group.” If one examines the 17.3% All Other Hispanics or Latinos by the “group” that was recorded, 10.2 %, or almost 60% wrote in “Spanish”, ”Hispanic” or “Latino” – a clear literal translation of what the instruction said to do. If these are removed, “unknown country of origin” becomes about 7%, which is pretty consistent with CPS experience that has averaged about 6.3% over the past two years.

Our previous experience on this type of survey is that we would expect Unknown Country/Regional Background to be on the order of 2%-3% as opposed to the CENSUS NEC category total of over 17%. This is problematic if we were to use the CENSUS subgroup totals alone as estimates of Latino population subgroup size since each subgroup would, on average, are understated by 17% of the size of that group (e.g., Dominicans). Looking at it another way and without any other changes we would need to weight up our NECs from say 3% to 17% - not a reasonable approach and we would knowingly and substantially understating the population in each subgroup. Consequently, we decided to account for the Census understatement of non-Listed COOs by allocating the excess “Others” over 6% across all country of origin groups except for Mexican, Puerto Rican and Cuban. [Table 2]

The targeted-subgroup samples were then developed. First, using 2000 Census data we computed country of origin estimates by State for the three target groups and from this we determined that coverage among the seven (7) original States was too low for both Columbians and Salvadorans. Consequently we added MA, MD, and VA, bringing the State-level coverage to 85.5% for Columbians, 88.1% for Salvadorans and 98.8 for Dominicans.

A file containing tract-level 2000 Census Latino country of origin breakdown was then compiled and “High Incidence” tracts identified for each country of origin. Once these tracts were identified, the NXXs serving each set were identified and assigned to one and only one sample group [Note: since these country of origin -target NXXs are all duplicated in the main sample frame they are oversampled relative to all other NXXs – the procedures for adjusting this are discussed in the next section.] In order to increase the efficiency of these RDD samples all telephone numbers corresponding to non-Hispanic listed households were eliminated prior to interviewing.

**Weighting and Estimation.**

The weighting and estimation phase was accomplished in two distinct phases. The weighting for the White/Other and African-American segments were accomplished first, as this was a relatively trivial process. The samples for these two groups had been designed to be approximately proportionate across the 33 sampling strata. The first step here was to adjust the actual final sample size to the expected proportionate sample within each State/Regional sample. These initial weights were then used as initial input weights to a sample balancing routine that controlled for age within sex for each of the basic eleven State/Regional strata.

The estimation phase of the Latino sample was more complex because of the planned disproportionate sampling as well as the country of origin oversamples and included a number of steps.

First, within each of the thirty-three geographic strata the sample was post-stratified based on the country of origin oversample definitions and country of origins of the completed interviews. In the main study, all Latinos were interviewed, regardless of country of origin, but in the oversamples, only Latinos with one of the four country of origins were sampled. Consequently, for each subset of NXXs we had two samples of the target country of origins: one from the Main Study one from the oversample. For each country of origin and its associated NXX subsets we also developed estimates of country of origin totals by overlaying 2000 Census tract data on each NXX – basically for every NXX we developed a country of origin distribution of the estimated Latino households using the previously non-response adjusted 2000 Census data. This process produced weights that brought the oversampled country of origin Latino interviews back into proportion with the non-target country of origins within the oversampled NXXs.

Second, for each of the thirty-three NXX-defined completed Latino interviews were then balanced in a single step using a sample balancing routine controlling for age within sex, specific country of origin subgroups and Total Latino households. As part of the balancing, the process was controlled to produce weights scaled to compensate for the disproportionate sampling.

##### Response Rate

The overall response rate for this study was calculated to be 56.6%.

The following formula was used to calculate the response rate:

*(Completes + Short interviews)*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*(Completes + Short interviews + Refusals + Projected number of refusals)*

where Projected number of refusals are calculated as:

## *Non-completion percentage x Non-contacted group*

Following is a full disposition of the sample selected for this survey:

TOTAL NUMBERS DIALED 103,030

Non-working, disconnected,

Business, or government 36,871

Ineligible 3,521

Data/modem/fax line 4,382

Language Barrier 1,262

Non-contacts

No answer 13,808

Busy 488

Callback 2,381

Refusals 14,180

Completed interviews 4,213

Completed short interviews with non-Latino households 21,924