reaching Hispanics

best practices in conducting high quality research for this critical audience

SSRS
research.refined.
Hispanics, one of the largest growing population segments in the United States, present special challenges to survey researchers. In addition to posing the same difficulties inherent to polling any small group, Hispanics also comprise subgroups representing different countries of origin. Some Hispanics have lived in the United States for several generations while others have only just arrived, and the ability to speak English and Spanish varies widely. Along with other cultural and demographic attributes unique to this group, such considerations require researchers to be particularly meticulous when attaining representative samples of the Hispanic population.

Regardless of their increasing numbers, Hispanics still currently comprise only 15 percent of the U.S. adult population, according to the U.S. Census Bureau’s Current Population Survey March 2013 supplement. This means that initial screening questions will encompass the lion’s share of cost in most Hispanic telephone surveys. In fact, a survey that costs $50,000 to interview any adult age 18 and older could cost three or four times as much when screening only for Hispanics, overall sample sizes being equal. For most researchers, this constitutes a significant cost concern that often spurs them to look for alternative sampling options.

Inexpensive alternative method of surveying Hispanics would be through the use of a convenience web panel. Like with any research, the key question one has to ask with regard to this data source is whether the non-probabilistic nature of recruiting panel members will lead to significant bias in the data gathered. Web panels have been successfully utilized for a range of research, and in many applications they will provide reliable approximations of the population in question. But in Hispanic research in particular, there are reasons to believe that data from non-probability web panels would be less trusted than for other populations. For the vast majority of online panels, two types of people will NOT be represented in the panel: 1) People who do not have access to the Internet and 2) People who cannot read English. This excludes 45% of the Hispanic population in the United States.

45% of Hispanics Would be Missed On Typical Web Panels
In addition, the 55% of Hispanics who theoretically could be included in an online panel are dramatically different than those who are missing. Namely, these potential Hispanic web panel members are highly acculturated. They are three times as likely to have parents who were born in the U.S.; more than twice as likely to be a citizen of the U.S.; nearly twice as likely to have a driver’s license; and six times as likely to have a four-year college degree, among other stark differences as shown below.

And furthermore, Hispanics who, by definition, cannot be part of a web panel are tremendously different in their consumption of products and media, and hold substantial differences of opinion. 46 percent of Web panel Hispanics regularly send text messages, compared to only 10 percent of non-web panel Hispanics. A third of web panel Hispanics use social networks, over twice the rate of non-web panel Hispanics. If one were to consider the consumption of Spanish language music via web panel, one would find that only about 10 percent of Hispanics are frequent purchasers of Spanish music. But because about 75% of non-web panel Hispanics consume Spanish language music, the true estimate for Hispanics overall is not 10 percent but 43 percent!

In short, the risk in using web panel data centers on whether a non-probabilistic approach will get the “true” answer. With Hispanics, it is a near mathematical impossibility, given that nearly half of Hispanics have almost no ability to be part of a web panel, and that these Hispanics in many respects are polar opposites of Hispanics who could be web panel participants.

For these reasons, high quality research of Hispanics continues to be principally telephone-based. The simplest way to survey Hispanics via telephone is to use dual-frame (landline and cell phones) random-digit-dialing (RDD). Cell phone interviews are critical given that now over HALF of all Hispanic households in the U.S. do not own a landline. But the problem with this method is its exorbitant cost, since interviewers must be paid to call households, only to screen out more than eight in ten of those households as not being Hispanic. For a survey of 1,000 respondents, this constitutes calling nearly 7,000 households, thereby incurring significant costs.
A second method is to employ a disproportionate stratified RDD design. This design first acquires the incidence for reaching a Hispanic household for every valid telephone exchange (or in the case of cell phones, rate centers) in the United States. Then it orders these from the highest incidence (some exchanges, such as exchanges in and around Little Havana in Miami, can reach incidences of over 80 percent or more Hispanic) to the lowest. These are then clustered into groups, called strata, then typically labeled as the high incidence group, medium, and low (and sometimes very high and very low as well). Then, simply put, the lion’s share of the completed interviews are collected from the high incidence groups, while only a few interviews are collected from the low incidence groups. Now, that same 1,000 completed interviews that required 7,000 screens in the RDD design may require only 3,500 screens in the disproportionate stratified design. At the end of the day, sophisticated weighting procedures correct the disproportionality entered in the sampling design to balance the study back to true national representation. Overall, the disproportionate stratified design meets all the criteria of a high-quality, methodologically rigorous design, while incurring significant cost savings.

However, there are instances where such cost savings are not enough. After all, the disproportionate stratified design still only increases the survey incidence of reaching Hispanics from the 15 percent found in the general adult population to somewhere in the range of 20 – 32 percent. So while such a design can double the incidence, thereby essentially cutting the cost of interviewing in half, such discounting still results in a study whose cost is roughly double that of a similar study of the general population. In such cases where cost is the ultimate driving factor, researchers have to turn to one final sampling option: the use of listed samples consisting of people with distinctive Latino surnames.

Surname sampling is an attractive option because it can cut the cost of a survey in half compared to a disproportionate stratified sampling design. However, there is a price to pay for the use of surname sampling in terms of coverage. First, again, only half of Hispanic households have a landline telephone. And only half of those have a listed telephone number. Thus, only 25 percent of Hispanics have any chance to be selected in this design. But furthermore, only about half of Hispanics have a recognizable Hispanic surname. Therefore the true coverage of this design is about 12 percent of all Hispanics. And therefore, one must ask whether respondents with distinctive surnames, with a listed landline telephone, are significantly different from Other Hispanics.

Surname sampling is an attractive option because it can cut the cost of a survey in half

In 2004, SSRS set out to document the potential bias this design may have. To investigate the differences between these and non-Latino surname respondents, twelve demographic variables were selected from a recent Hispanic survey conducted by SSRS, as well as five media-consumption variables, five corresponding language-of-media-consumption variables, and two Latino identity variables.
Figure 1 summarizes the mean differences between surname and non-surname respondents within demographic variables. It shows some stark differences between Latinos with and without Latino surnames. Most prominently, Latinos with surnames were just over half as likely to have attained a college degree. Furthermore, there was a 16 percentage point gap in voter registration and a $12,000 income gap, both favoring Latinos without Latino surnames. Latinos with surnames were far more likely to be Mexican and less likely to be Puerto Rican, and less likely to be living in the northeastern region of the United States. Finally, while Latinos with surnames were less likely to be born in the United States, they also were, on average, older. Indeed, within the twelve demographic measurements analyzed, totaling twenty-six variables, nearly half showed substantial differences by surname status.

And that was in 2004. Today, the differences are surely much larger given that virtually no Hispanics were “cell phone only” back then and over half are today.

As noted, even a highly disproportionate stratified design will lead to costs significantly higher than a general population survey, and much higher than a survey from a web panel. The challenge is, therefore, to find a method of covering all Hispanics at costs approaching those of traditional panel providers.
SSRS offers a range of options that attain high quality, randomly representative cross-sections of the Hispanic population at a cost only fractionally higher than web panels. Specifically, SSRS utilizes its suite of omnibus surveys, in which the screening is a covered cost and shared across clientele, to screen for Hispanics. And given the number of screens attained on a weekly and monthly basis, a cross-section of 1,000 Hispanics can be attained fairly quickly. And most importantly, because these omnibus surveys are conducted in English and Spanish and have a particular concern toward attaining representative cross-sections of Hispanics, the quality of the data attained from these sources is high.

The SSRS Omnibus survey is a national, weekly, dual-frame bilingual RDD telephone survey designed to meet standards of quality associated with publically released standalone custom research studies. Each weekly wave of SSRS Omnibus consists of 1,000 interviews, of which 500 are obtained with respondents on their cell phones and about 100 with Hispanics, of which at least 35 interviews are completed in Spanish. SSRS Omnibus includes weighting specifically designed to attain representative Hispanic cross-sections by adjusting for not just the typical parameters such as age and gender, but also by U.S. nativity, and (when weighting Hispanic-only samples) longevity of U.S. residency.

Additionally, SSRS’ sister company Centris conducts a monthly survey of 2,000 respondents, utilizing a randomized address-based sample. In short, random samples from the U.S. Postal Service mail delivery file (the CDSF) are generated and telephone numbers are appended where available. Those with telephone numbers are asked to participate in a telephone survey while those without a number are mailed an invitation to conduct the survey on the Internet. The monthly Centris survey is conducted in both English and Spanish.

Combined, these two sources collect hundreds of Hispanic interviews on a monthly basis.

Interested parties can insert any number of questions into our omnibus platforms, up to 12 minutes in length. As well, the demographic and screening questions do not “count” as part of that length; such questions are already asked in the omnibus surveys. The following demographics are captured for all Hispanic respondents:

- Age*
- Gender*
- Education*
- Employment Rate*
- Race*
- Born/Years in the U.S.
- Marital Status
- Party Identification
- Political Ideology
- Voter Registration
- Parental Status
- Religion
- Health Insurance Status
- Language Preference
- Income*
- Homeownership
- Household Size/Composition
- Age of Children
- Region*
- Metro Status*
- Hispanic Neighborhood Density
We stress that the samples from these omnibus surveys are fully random samples of the Hispanic population, providing representative cross-sections by age, acculturation, language use, generation, and other measures. As such we have a unique ability to conduct surveys of specific sub-populations of Hispanics if so desired.

The tables below provide the approximate sample sizes that can be attained with the combined approaches of our omnibus surveys. As well, SSRS has relationships with the largest web panel providers in the country. We are able to collect surveys of Hispanics from these providers, and we then execute a “propensity weighting” procedure that calibrates the responses of web panel Hispanics to the higher quality and more representative telephone Hispanics. We recommend only attaining a maximum of 50% of the interviews with the web panel with this approach, to maintain the overall quality of the study.
A final area of concern regarding high quality interviewing of Hispanics concerns the weighting of Hispanic data. As detailed in Dutwin (2014), there are a number of different weighting procedures out there, and they all vary considerably at the degree to which they make Hispanic samples from any data course representative. It is critical to not only weight Hispanics to specific “typical” targets such as age, gender, and education, but also to weight Hispanics to “Hispanic specific” targets such as heritage (Mexian, Puerto Rican, Cuban, Central American, South American, Other), foreign born/U.S. born status, and if born outside the U.S., number of years the person has lived in the U.S. (typically by decades, 0-10 years, 11-20 years, and 21 or more years). Only by utilizing such corrections or similar corrections will Hispanic segment samples be accurate by generation, by language use, and by acculturation.

In sum, conducting high quality research of Hispanics requires specialized knowledge and expertise. But it can be done, and fairly inexpensively. Regardless of the data source, it is important to understand the ways in which data collection techniques and weighting techniques can make the data as representative as possible.

For more information on this topic, please contact:

David Dutwin, Ph.D.
Executive Vice President/Chief Methodologist
SSRS
53 West Baltimore Pike
Media, PA 19063-5698
484-840-4406
ddutwin@ssrs.com