CONSIDERATIONS OF SURVEY ERROR IN SURVEYS OF HISPANICS

DAVID DUTWIN*
MARK HUGO LOPEZ

Abstract As the largest and fastest-growing minority population in the United States, Hispanics have become an increasing focus of survey research. The vast body of Hispanic research evidences myriad options regarding sampling, data collection, and weighting, each of which can affect the resulting data about this population. Typical survey designs feature simple random samples (sometimes obtained as part of larger omnibus or general-population surveys), stratified RDD, “top market,” and surname designs. In addition, some studies obtain interviews in English only, while others offer both English and Spanish but make choices regarding the use and allocation of bilingual interviewers. Finally, there are a range of considerations in the weighting of Hispanic survey data. Utilizing data from a national omnibus survey, the General Social Survey, and the Pew Hispanic Center National Survey of Latinos, this article explores these three foci: sampling, interviewing language, and weighting. We report on what we find to be best practices and the implications of failing to enact these practices, as measured by bias and variance in survey estimates of Hispanics.

As the largest and fastest-growing minority population in the United States, Hispanics have naturally become an increasing focus of survey research. There are myriad decisions to make when operationalizing surveys of Hispanics, including some that are unique or at least of special consideration when interviewing Hispanics in the United States. Significantly, many choices in designing and executing surveys of Hispanics will have a particular and robust impact on survey error. To date, however, several critical aspects of Hispanic research have garnered little or no attention.

DAVID DUTWIN is executive vice president and chief methodologist for SSRS/Social Science Research Solutions in Media, PA, USA. MARK HUGO LOPEZ is director of Hispanic research at the Pew Hispanic Center in Washington, DC, USA. *Address correspondence to David Dutwin, SSRS, 53 West Baltimore Pike, Media, PA 19063, USA; e-mail: ddutwin@ssrs.com.
We approach the issue through the lens of total survey error by asking which choices faced by researchers of Hispanics have the greatest impact on estimates of Hispanic public opinion and survey research and are therefore likely to be the most significant potential contributors to survey error. We point to three principal areas of particular concern. First is the design of the sample itself. Many popular designs do not afford full coverage (the percentage of the total population that has any probability for being selected to participate in the survey) of the Hispanic population, while others purport to provide near-full coverage but with disproportionate probabilities of selection that both save costs and introduce variance. Second, researchers make decisions as to the effort that will be made to interview Hispanics in Spanish, which again may impact coverage and therefore, potentially, bias. The third decision researchers need to make relates to the procedures for weighting Hispanic data. Many studies cull Hispanic interviews over time in a general-population framework and utilize the overarching weight for the general population. Alternatively, Hispanics can be weighted to Hispanic population estimates, specifically using the same target variables (e.g., gender, age, educational attainment) that one would ordinarily use for the general population. Notably, however, this procedure ignores the fact that there are measures particular to Hispanics, such as heritage (e.g., Mexican, Puerto Rican, Cuban, etc.) that may be important in further reducing nonresponse bias in the weighting of Hispanics. Overall, we find that the choices made when conducting Hispanic research lead to considerable differences in the total survey error introduced into surveys of Hispanics. We summarize those findings in this article and discuss the implications of these best practices with regard to Hispanic research.

Total Survey Error

Much has been written about total survey error (TSE) (Kish 1965; Groves and Lyberg 2010), including Public Opinion Quarterly’s special issue on this very topic (Biemer 2010). In that issue, contributors provided multiple frameworks for understanding TSE, all focusing on two overarching concerns: bias, measured as the difference between the survey estimate and a “gold standard” of that parameter, and variance, measured by one of many possible metrics that describe the range and spread of observations of an estimate. While the problem with bias is self-explanatory (namely, that one has arrived at the wrong estimate), variance error increases variation in an estimate without necessarily contributing to bias. Still, the greater the variance, the larger the standard error of the estimate, reducing the statistical power available for significance testing.

Regardless of the model, the goal of a TSE approach is to minimize TSE whatever the particular survey design and “price point” might be. Using insights gained from a TSE approach, researchers are better able to identify and make decisions regarding trade-offs between designs that minimize costs at the expense of increased variance and those that minimize costs at the expense of increasing
bias. This article is not meant to provide a full review of TSE, nor is it meant to review potential survey errors that are not particular to Hispanic research. For instance, we do not focus on data-processing errors equally likely to emerge in surveys of different populations. Rather, we focus on areas of particular concern for surveys of Hispanics, based on the choices one must make when conducting surveys of Hispanics—namely, the impact of the survey design itself, administration of the survey, particularly the availability of bilingual interviewing, and efforts at reducing errors of representation through weighting.

Surveying Hispanics: Sample Design

Interviews of Hispanics can be collected using one of two different overarching strategies. Hispanic interviews can be attained via screening studies where all non-Hispanics are terminated in an initial screening process and are not administered the full interview. Alternatively, Hispanics can be interviewed as part of larger population studies, typically those whose sampling frame includes all adults within the general population.

For screener studies, a major barrier is cost. Given that Hispanics constitute only 14.3 percent of the U.S. adult population (ACS 2010), a survey of Hispanics will cost far more than a comparable survey of the general population. Therefore, researchers often utilize nonsimple random sample (SRS) strategies to collect their data. Most typical are stratified designs in which geographies, defined within telephone exchanges, are clustered into strata, and strata of high Hispanic household incidence (that is, percentage of all households that are Hispanic) are oversampled compared to strata of low incidence.

The Pew Hispanic Center (PHC) National Survey of Latinos (NSL), conducted annually since 2002, has utilized a variety of designs in which landline, and more recently, cell-phone telephone exchanges are clustered by their expected incidence of reaching Hispanics. In 2010 and 2011, for example, the PHC executed a design in which all landline telephone numbers were divided into those that are listed in a telephone directory and associated with a Hispanic surname, and all other landline telephone numbers. Using data from Marketing System Group’s Genesys sampling software, the remaining telephone numbers were split, by exchange, into expected Hispanic household incidences, as shown in Table 1. Similarly, cell-phone area codes were stratified by expected Hispanic incidence. Overall, the landline surname strata and very high-incidence strata were oversampled, while the low- and medium-incidence strata were undersampled. Again stratification is a common strategy, utilized in many other studies (see, for example, LeFauvre, Srinath, and Arday 2006; Ponce et al. 2006; Kalton 2009).

1. Note that the low-incidence strata were not undersampled significantly, or at all, because the study utilized a prescreened sample from a national omnibus survey to attain Hispanic interviews in the low strata.
Table 1. Pew Hispanic Center Stratified Design

<table>
<thead>
<tr>
<th>Strata</th>
<th>Percent of Hispanic households (within phone type)</th>
<th>Expected survey incidence (%)</th>
<th>Interviews</th>
<th>Percent of interviews (within phone type)</th>
<th>Percent who cannot speak English well (modeled from ACS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landline surname</td>
<td>30.5</td>
<td>78</td>
<td>547</td>
<td>39.0</td>
<td>38</td>
</tr>
<tr>
<td>Landline very high</td>
<td>19.4</td>
<td>54</td>
<td>253</td>
<td>18.5</td>
<td>25</td>
</tr>
<tr>
<td>Landline high</td>
<td>18.8</td>
<td>34</td>
<td>194</td>
<td>14.5</td>
<td>21</td>
</tr>
<tr>
<td>Landline medium</td>
<td>17.2</td>
<td>18</td>
<td>136</td>
<td>10.4</td>
<td>18</td>
</tr>
<tr>
<td>Landline low</td>
<td>14.2</td>
<td>4</td>
<td>197</td>
<td>17.6</td>
<td>16</td>
</tr>
<tr>
<td>Cell-phone high</td>
<td>37.6</td>
<td>40</td>
<td>758</td>
<td>58.7</td>
<td>32</td>
</tr>
<tr>
<td>Cell-phone medium</td>
<td>33.1</td>
<td>24</td>
<td>331</td>
<td>27.7</td>
<td>25</td>
</tr>
<tr>
<td>Cell-phone low</td>
<td>29.3</td>
<td>7</td>
<td>179</td>
<td>14.3</td>
<td>18</td>
</tr>
</tbody>
</table>

Cost saving is the primary reason to enact a stratified, dual-frame design such as that of the NSL to interview Hispanic households: These designs can attain survey incidences of reaching Hispanic households of up to 30 percent. Therefore the survey requires only half the number of screener interviews compared to an SRS survey, and costs are approximately 25 percent less than a comparable SRS survey. Such a design, properly weighted, should not be any more biased than a dual-frame survey utilizing a simple random sample. However, the variance will be greater, since weighting must correct for the disproportionate probability of selection of strata sampling. This is well understood, as all stratified designs will attain a higher design effect than comparable simple random sample designs. That said, researchers can model an optimal allocation scheme for the stratification, finding in effect the “minmax” point for the lowest margin of error at the lowest price point (Tucker, Casady, and Lepkowski 1992).

Another screener-based design option is to conduct a “top market” survey of Hispanics, which interviews only households in the most populous Hispanic markets. This design can be advantageous to in-person research in that it limits the number of locations to which interviewers must travel. Table 2 shows a typical top-market design. A top-ten-market design achieves a 30 percent survey incidence, and covers 46 percent of all U.S. Hispanic households. The top ten markets above the line in table 2 represent a respectable “spread” of markets, without undue concentration in Texas, for example, and other parts of the Southwest. It is also possible to develop a top-market design where the top markets are defined by Hispanic incidence rather than raw population, or by strategically substituting some of the low-incidence markets for high-incidence markets. For example, one could substitute Chicago, Dallas, and San Francisco for McAllen, El Paso, Fresno, and Albuquerque. This design would increase incidence only from 30 to 34 percent, however, and would surely skew more significantly to Mexicans than to Hispanics of other backgrounds.

A final design that has been popular in market research is the surname design, a highly cost-effective option in which survey incidence typically approaches 80 percent. Specifically, we estimate that a surname-only design will cost only about 55 percent of a stratified dual-frame design and less than 40 percent of a SRS dual-frame design. Importantly, however, this landline-only design, which once covered just over 30 percent of all Hispanic households, now covers only about 21 percent of Hispanics thanks to the migration of Hispanics to CPO status. Also noteworthy is evidence that sampling by

---

2. Notably, another potential design choice would be to eschew the dialing of cell phones, independent of other decisions made regarding the sample design. At best, however, this strategy would only cover just under two-thirds of all Hispanic households, given that Hispanics lead all major racial/ethnic groups in owning only cell phones. Specifically, the most recently available NHIS estimate for cell-phone-only status (CPO) of Hispanics, January to July 2010, finds that 33 percent of Hispanics are CPO (Blumberg and Luke 2011). Thus, while landline-only studies may be conducted by those looking for low-cost options, researchers need to be cognizant of the substantial undercoverage. The bias inherent in such a design has been documented and is significant (Dutwin, Keeter, and Kennedy 2010).
Table 2. Metropolitan Statistical Areas with the Largest Populations of Hispanics

<table>
<thead>
<tr>
<th>MSA</th>
<th>Hispanics</th>
<th>All adults</th>
<th>Incidence (%)</th>
<th>Cumulative coverage (%)</th>
<th>Cumulative incidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles–Long Beach, CA</td>
<td>3,837,690</td>
<td>9,490,885</td>
<td>40</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>New York–Northeastern NJ</td>
<td>3,009,875</td>
<td>13,385,071</td>
<td>22</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>Houston–Brazoria, TX</td>
<td>1,303,928</td>
<td>3,986,511</td>
<td>33</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Miami–Hialeah, FL</td>
<td>1,270,626</td>
<td>1,866,491</td>
<td>68</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>Riverside–San Bernardino, CA</td>
<td>1,251,904</td>
<td>2,958,733</td>
<td>42</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>Chicago, IL</td>
<td>1,236,427</td>
<td>6,727,442</td>
<td>18</td>
<td>36</td>
<td>31</td>
</tr>
<tr>
<td>Dallas–Fort Worth, TX</td>
<td>1,075,913</td>
<td>4,416,464</td>
<td>24</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>San Francisco–Oakland–Vallejo, CA</td>
<td>734,127</td>
<td>3,762,393</td>
<td>20</td>
<td>42</td>
<td>29</td>
</tr>
<tr>
<td>San Antonio, TX</td>
<td>727,119</td>
<td>1,399,353</td>
<td>52</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td>Phoenix, AZ</td>
<td>688,538</td>
<td>2,774,941</td>
<td>25</td>
<td>46</td>
<td>30</td>
</tr>
<tr>
<td>San Diego, CA</td>
<td>644,981</td>
<td>2,277,704</td>
<td>28</td>
<td>48</td>
<td>30</td>
</tr>
<tr>
<td>Washington, DC/MD/VA</td>
<td>530,339</td>
<td>4,107,792</td>
<td>13</td>
<td>50</td>
<td>29</td>
</tr>
<tr>
<td>McAllen–Edinburg–Pharr–Mission, TX</td>
<td>446,735</td>
<td>505,433</td>
<td>88</td>
<td>51</td>
<td>29</td>
</tr>
<tr>
<td>El Paso, TX</td>
<td>443,494</td>
<td>551,624</td>
<td>80</td>
<td>53</td>
<td>30</td>
</tr>
<tr>
<td>Orlando, FL</td>
<td>378,724</td>
<td>1,613,901</td>
<td>23</td>
<td>54</td>
<td>29</td>
</tr>
<tr>
<td>Denver–Boulder, CO</td>
<td>371,613</td>
<td>1,901,231</td>
<td>20</td>
<td>55</td>
<td>29</td>
</tr>
<tr>
<td>Las Vegas, NV</td>
<td>361,103</td>
<td>1,445,258</td>
<td>25</td>
<td>56</td>
<td>29</td>
</tr>
<tr>
<td>Fresno, CA</td>
<td>342,041</td>
<td>743,504</td>
<td>46</td>
<td>57</td>
<td>29</td>
</tr>
<tr>
<td>Atlanta, GA</td>
<td>335,744</td>
<td>3,552,993</td>
<td>9</td>
<td>58</td>
<td>28</td>
</tr>
<tr>
<td>Fort Lauderdale–Hollywood, FL</td>
<td>328,873</td>
<td>1,351,017</td>
<td>24</td>
<td>59</td>
<td>28</td>
</tr>
<tr>
<td>Austin, TX</td>
<td>322,004</td>
<td>1,171,211</td>
<td>27</td>
<td>60</td>
<td>28</td>
</tr>
<tr>
<td>San Jose, CA</td>
<td>318,709</td>
<td>1,333,366</td>
<td>24</td>
<td>61</td>
<td>28</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>MSA</th>
<th>Hispanics</th>
<th>All adults</th>
<th>Incidence (%)</th>
<th>Cumulative coverage (%)</th>
<th>Cumulative incidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tampa–St. Petersburg–Clearwater, FL</td>
<td>311,923</td>
<td>2,158,969</td>
<td>14</td>
<td>62</td>
<td>28</td>
</tr>
<tr>
<td>Albuquerque, NM</td>
<td>278,573</td>
<td>651,026</td>
<td>43</td>
<td>63</td>
<td>28</td>
</tr>
<tr>
<td>Philadelphia, PA/NJ</td>
<td>265,086</td>
<td>3,958,892</td>
<td>7</td>
<td>64</td>
<td>27</td>
</tr>
<tr>
<td>Other MSAs</td>
<td>8,570,004</td>
<td>96,778,449</td>
<td>9</td>
<td>90</td>
<td>17</td>
</tr>
<tr>
<td>Not in an MSA</td>
<td>3,274,533</td>
<td>52,581,138</td>
<td>6</td>
<td>100</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>32,660,626</td>
<td>227,451,792</td>
<td>14</td>
<td>100</td>
<td>14</td>
</tr>
</tbody>
</table>

Source.—2010 American Community Survey.
surname is considerably biased on a range of measures, including demographics, voter registration, and media use (Lee and Ramakrishnan 2002; Dutwin and Herrmann 2004; Ben-Porath et al. 2011). In addition, concerns over the accuracy of surname flags could lead to significant bias (Medina Vidal 2009).

Surveying Hispanics: Language

Language of interview is the second dimension by which Hispanic surveys differ from most other surveys. Many general-population surveys and omnibus surveys are conducted in English only. As shown in table 3, however, nearly 30 percent of Hispanics either do not speak English well (17 percent) or at all (10.7 percent).

Not surprisingly, prior research has found substantial differences between Hispanics who complete surveys in English versus Spanish. For example, Kirkman-Liff and Mondragon (1991) found that Hispanic adults interviewed in Spanish reported being in much worse health and had a more difficult time accessing health care than those interviewed in English. Lee et al. (2008) and Lee and Grant (2009) analyzed Hispanic data in English in the 2003 California Health Interview Survey and found that compared to those interviewed in English, respondents interviewed in Spanish had lower incomes and lower levels of educational attainment and were more likely to report being in fair or poor health and uninsured. Kiley et al. (2010) also found considerable differences on language of interview by education, region, heritage, voter registration, and party identification. A number of other studies have documented significant differences in survey estimates by language as well (Kandula, Lauderdale, and Baker 2007; Viruell-Fuentes et al. 2011).

It is important to note that survey research firms must decide not only whether to interview in Spanish but also how to allocate their often limited team of bilingual interviewers. Bilingual interviewers typically constitute a minority of all interviewers in any given survey research firm and often get paid a higher hourly rate. Some survey firms claim to conduct “instant hand-offs” whereby an English-only interviewer places a respondent on hold and

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaks only English</td>
<td>4,898,848</td>
<td>21.2</td>
</tr>
<tr>
<td>Speaks Spanish and speaks English very well</td>
<td>8,478,956</td>
<td>36.6</td>
</tr>
<tr>
<td>Speaks Spanish and speaks English well</td>
<td>3,361,356</td>
<td>14.5</td>
</tr>
<tr>
<td>Speaks Spanish and speaks English not well</td>
<td>3,935,029</td>
<td>17.0</td>
</tr>
<tr>
<td>Speaks Spanish and does not speak English</td>
<td>2,480,678</td>
<td>10.7</td>
</tr>
<tr>
<td>Total</td>
<td>23,154,867</td>
<td></td>
</tr>
</tbody>
</table>

Source—2010 American Community Survey.
immediately transfers the call to a bilingual interviewer. But it is unclear to what degree this is true in practice. For one thing, it requires that bilingual interviewers are ready to take such calls, which means by definition that they cannot be doing anything other than waiting for these handoffs to occur. This creates a situation where bilingual interviewers not only cost more on an hourly basis, but are more costly to utilize, given that some portion, if not the majority, of their time is spent waiting for a transfer to actually occur.

Most often firms utilize bilingual interviewers to call back households that have been previously dispositioned by English-speaking interviewers as “Spanish-language barriers.” While ideally these households would immediately be called back by a bilingual interviewer, it is operationally challenging, if not impossible, to make an instantaneous callback, since many CATI software programs will not allow an immediate callback in the same daypart. Typically, language barrier callbacks are made in later shifts. Notably, despite numerous call attempts to call households designated “Spanish-language barriers,” many studies will successfully recontact less than half of this sample. Many callbacks result in constant no-answer, busy-signal, or answering-machine dispositions, and thus the opportunity to interview that household—an opportunity that was present when the potential respondent first answered the phone and was greeted by an English-speaking interviewer—is lost to nonresponse. A more effective but operationally complex strategy is to sequester landline Hispanic surname numbers as well as landline and cell-phone high Hispanic incidence exchange numbers so that they can be dialed specifically by a bilingual interviewer from the start, even in a study utilizing an SRS of the general population. Such a strategy will make the most of limited bilingual resources. Bilingual interviewers can then be utilized to dial “Spanish-language barriers” from low- and medium-incidence samples in subsequent call attempts. This strategy is justified by data in table 3, which shows that Hispanics in medium- and low-incidence strata are much less likely to be linguistically isolated than Hispanics in high-incidence areas, particularly Hispanics with a Hispanic surname.

**Surveying Hispanics: Weighting**

Given the vast number of sampling and language-related operational choices in interviewing Hispanics, it follows that there are a plethora of choices to be made with regard to weighting surveys of Hispanics. After all, from a total survey error perspective, weighting is a key tool by which one can reduce bias, albeit at the cost of willingly introducing variance. Weighting has been noted to have the potential for significant impact on estimates of Hispanics, though

---

3. Dayparts refer to segments of time that CATI programs use in order to vary when call attempts are made. A typical program might have three dayparts: (1) weeknight early; (2) weeknight late; and (3) weekend. If a study is to have a maximum of six call attempts, the program will often be set to place only two calls within each daypart.
the research thus far has not isolated specific weighting practices to specific effects on estimates (Perl, Greely, and Gray 2006).

In studies that glean Hispanics from larger samples, there is the question of whether to utilize a “self-weighting” strategy or to specifically weight the data to Hispanic population demographic targets. In the case of the former, the entire sample is weighted and then all non-Hispanics are removed from the data, thereby resulting in a self-weighted sample of Hispanics. Even though the larger data set will match the general-population parameters on those targets to which the data were weighted, such a strategy does not guarantee that Hispanics specifically are representative of the Hispanic population on the same parameters. Yet this strategy of self-weighting is fairly typical in survey research, if only due to the fact that one does not have to expend additional labor in executing a special weighting routine for the Hispanic sample. However, it is unclear to what degree this strategy might result in a weighted sample of Hispanics that fairly accurately matches Hispanic population parameters.

Weighting Hispanics specifically to Hispanic population parameters offers a more precise method of reducing bias. A typical weighting scheme targets age, gender, region, educational attainment, and phone use. Variants may utilize interaction terms and other measures such as population density. All of these measures are identical to those that might be utilized in a general-population framework, with the exception that ethnicity is not included, since the sample is limited only to Hispanics. From a total survey error perspective, however, it is important to determine whether there are parameters particular to Hispanics that might additionally be utilized to reduce potential bias and better match the sample to the population, as suggested by Perl, Greely, and Gray (2006).

The first obvious choice would be heritage; that is, whether the respondent identifies as being Mexican, Puerto Rican, Cuban, or of some other Hispanic background. Given that most Hispanics are Mexican, such a measure might cluster all other groups as South American, Central American, or other Hispanic, rather than try to force the sample to match precisely on the many other different countries of origin. A second possibility is language, specifically the degree to which the respondent speaks English (based on categories provided by the American Community Survey). It is an open issue, however, as to whether weighting should match sample to population on a question that is arguably at least partially attitudinal on the part of the respondent (compared to more particular questions like gender and educational attainment). Alternatively, other potential weighting targets include measures that correlate well with language but are more factually based, such as whether the respondent was born in the United States, and if not, how many years they have lived in the United States.

The Present Studies

Given these major choices facing researchers of Hispanics—namely, decisions regarding sample design, language of interview, and weighting—the present
study attempts to assess the relative bias and variance associated with various choices across these dimensions. The research can be summarized with the following questions:

RQ1: What is the cost in bias and variance in utilizing a non-SRS sampling plan, specifically a plan that is stratified, top market, or surname?
RQ2: What is the cost in bias and variance in conducting a survey of Hispanics that is in English only?
RQ3: What is the cost in bias and variance in executing a self-weighting sample in a general-population framework, compared to a Hispanic-specific weighting plan?
RQ4: What is the cost in bias and variance in executing a weighting procedure that includes “Hispanic-specific” variables in addition to more typical weighting parameters?

Data

The data utilized for this study come from three sources: (1) the 2010 and 2011 Pew Hispanic Center (PHC) National Survey of Latinos (NSL); (2) 12 weeks of the EXCEL omnibus survey administered by Social Science Research Solutions (SSRS); and (3) the 2006, 2008, and 2010 General Social Surveys.

The 2010 NSL was conducted from August 17 through September 19, 2010, among a national sample of 1,375 Hispanic respondents aged 18 and older. The sample comprised a landline component \((n = 710)\) and a cell-phone \((n = 665)\) component. Interviews were conducted in English \((n = 548)\) and Spanish \((n = 827)\). The 2011 NSL was conducted from November 9 through December 7, 2011, and included 1,220 respondents aged 18 and older, 617 from landlines and 603 from cell phones. A total of 546 interviews were conducted in English, and 674 were conducted in Spanish. Both studies were stratified as described earlier in this paper and shown in table 1. The AAPOR RR3 response rate for the 2010 NSL was 25.0 percent overall, while RR3 for the 2011 NSL was 29.6 percent.

In order to investigate the research questions, five data sets were created from the combined 2010–2011 NSL data set, utilizing seven different weights. The stratified data set \((n = 2,595)\) utilized all interviews and was weighted in a typical two-stage method whereby a base weight included the following four steps: (1) correction for the number of adults; (2) weight of .5 for all owners of both a landline and a cell phone; (3) stratification correction whereby respondents from each stratum were weighted to represent the correct percentage of Hispanics living in that stratum; and (4) development of a propensity weight for all low-strata interviews completed using a recontact sample of households interviewed during the prior year for the EXCEL
omnibus (described later). The second step, poststratification, introduced the base weight into a standard raking routine that included education (less than high school, high school diploma, some college, or college degree), gender, age (18–29, 30–49, 50–64, 65+), region, born/years in the United States (U.S. born, foreign born and in the United States 0–10, 11–20, or 21+ years), heritage (Mexican, Puerto Rican, Cuban, Central/South American, or other Hispanic), and phone use (cell phone only, cell phone mostly, mixed, landline mostly, or landline only).

An SRS data set \( (n = 1,352) \) was created by randomly deleting cases within each stratum so that the unweighted data perfectly balanced the percentage of the population estimated to reside in each stratum. The most undersampled stratum was the medium landline stratum, with 136 landline and 331 cell-phone interviews. None of these cases were deleted, but cases from other strata were deleted to match the percentage of the population as reflected in table 1. These data then were weighted using the same procedure as described above, but without a correction for stratification.

A top-market data set \( (n = 446) \) was created by taking the SRS data set and deleting all cases that were not obtained from the ten top Hispanic markets in the United States. Also, a surname data set \( (n = 547) \) was created by deleting all non-surname landline cases from the full data set. Both the top-market and surname data sets then were weighted using the same method utilized in the SRS data set.

To explore differences by language and weighting parameters, two additional weights were generated utilizing two variants of the NSL data. The first data variant is simply the SRS data set described above, while the second deleted all cases from the SRS data set in which the interview was conducted in Spanish \( (n = 594) \). Each of these two data sets was weighted utilizing the full (what we label “enhanced”) weighting procedure described above, while a second weighting procedure did not utilize the two Hispanic-specific parameters of U.S. born/years in the United States and Hispanic heritage.

EXCEL is a national, weekly, dual-frame, bilingual telephone survey of 1,000 adults. Each weekly wave consists of 1,000 interviews, 500 of which are obtained from respondents on their cell phones, and a minimum of 30 interviews are completed in Spanish. EXCEL data for this research were drawn from interviews of Hispanics from 12 consecutive weeks in December 2011 through February 2012, for a total of 991 Hispanic interviews. Just under two-thirds \( (n = 624) \) were conducted in English, and the AAPOR RR3 was nine percent.

From these data, we generated the same eight sets of weighted data described above: (1) stratified (by taking the EXCEL data and randomly deleting cases to mirror the distribution of the NSL stratified design, \( n = 537 \)); (2)
SRS (all cases, \(n = 991\)); (3) top market \((n = 227)\) and surname \((n = 240)\), as well as an English-only variant of the SRS data \((n = 624)\). From these data, the same seven weights were created utilizing the methods noted above (stratified, SRS, top market, surname, all cases enhanced, all cases basic, English-only enhanced, and English-only basic). Two additional weights were generated by taking the full 12 weeks of EXCEL interviews, Hispanic and non-Hispanic alike, weighting to national general-population parameters and then deleting all non-Hispanic interviews (one weight for all cases and one weight for English interviews only).

Finally, the 2006–2010 General Social Surveys were included in these analyses. These data are collected in person and represent a typical high-effort national survey (see the GSS website for more details, http://www3.norc.org/GSS+Website/Download/). While the GSS has historically been an English-only survey, Spanish-language interviewing was added to the design in 2006. While other large-scale publicly available data sets such as the ANES or the NHIS were considered, none of those reviewed both conducted interviews in Spanish and asked the questions needed to replicate the weights noted above. We could not create stratified, top-market, or surname data sets with the GSS since such data were not available. And because the GSS does not ask respondents the number of years they lived in the United States, the enhanced weights included one to account for whether the respondent had been born in the United States without additional breakdowns for how long foreign-born respondents have lived here. The GSS comes with its own set of weights but does not poststratify to national population parameters. We utilized their final weight as a base weight and executed poststratification weights precisely as noted above, thereby generating three weighting procedures: (1) general population weighted, with non-Hispanics then deleted; (2) Hispanic-enhanced, that is, Hispanics weighted to Hispanic population targets including heritage and foreign-born; and (3) Hispanic-basic, that is, Hispanics weighted to Hispanic population targets without heritage and foreign-born targets, across two sets of data (all Hispanics, \(n = 852\), and English only, \(n = 515\)). See the row headers of tables 4 and 5 as a summary of each combination of data and weights.

**Method**

The study compares each weighted sample to others on measures of bias and variance. Variance is measured by design effect, a common metric that indicates the degree to which weighting inflates the standard errors compared to a simple random sample. Bias is measured simply as the difference between the point estimates attained in the survey and “gold-standard” point estimates, the parameter estimates as reported by the 2010 American Community Study. Using gold standards from these data means that the variables we can compare
are limited. Nevertheless, there are enough to provide a clear picture of bias in each sampling design, language strategy, and weighting scheme.

Using nativity measures for the respondent and the mother and father of the respondent, we created measures of individuals being of first generation (respondent foreign born) and third generation (both parents U.S. born). The measure of first generation was not used to compare the bias of the various sample design data sets, since all the data sets were weighted to this parameter. However, variation can and did occur in the degree to which samples were third generation. We also compared first- and third-generation Hispanics since we were interested in the degree to which the self-weighted sample and the Hispanic-weighted sample that did not include born/years in the United States were biased on this measure. In a similar vein, we compared the percent of Mexicans to assess the extent of bias in this measure in the self-weighted omnibus sample versus the basic Hispanic weight.

Another important point of comparison between the data sets and weighting schemes is the proportion of respondents who said they could speak English very well. This question is asked differently in our data compared to the gold standard. Specifically, the ACS asks “How well does this person speak English?” while both EXCEL and the NSL ask “How well would you say you can carry on a conversation in English?” (the GSS does not ask this question). The response options, however, are identical, and historically the
Table 5. Variance and Bias by Weight, Language of Interview, and Sample Design

<table>
<thead>
<tr>
<th>Study</th>
<th>ACS estimate</th>
<th>General population English (column a)</th>
<th>Hispanic English (column b)</th>
<th>Hispanic-enhanced English (column c)</th>
<th>General-population all (column d)</th>
<th>Hispanic all (column e)</th>
<th>Hispanic-enhanced all (column f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design effect</td>
<td>EXCEL</td>
<td>1.37</td>
<td>1.76</td>
<td>2.62</td>
<td>1.20</td>
<td>1.15</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>GSS</td>
<td>1.38</td>
<td>1.52</td>
<td>1.84</td>
<td>1.42</td>
<td>1.43</td>
<td>1.43</td>
</tr>
<tr>
<td>Age 18–29</td>
<td>EXCEL</td>
<td>29.6%</td>
<td>12.6%***</td>
<td>-0.3%</td>
<td>-0.2%</td>
<td>7.5%***</td>
<td>-0.3%</td>
</tr>
<tr>
<td>No H.S. Diploma</td>
<td>EXCEL</td>
<td>36.9%</td>
<td>-17.4%***</td>
<td>0.0%</td>
<td>-1.2%</td>
<td>-0.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Mexican</td>
<td>EXCEL</td>
<td>63.6%</td>
<td>-10.7%***</td>
<td>-1.7%</td>
<td>0.0%</td>
<td>-2.8%**</td>
<td>-4.2%**</td>
</tr>
<tr>
<td>English very well</td>
<td>EXCEL</td>
<td>54.2%</td>
<td>25.0%***</td>
<td>26.4%***</td>
<td>16.5%***</td>
<td>-2.6%</td>
<td>-1.6%</td>
</tr>
<tr>
<td>1st generation</td>
<td>EXCEL</td>
<td>57.7%</td>
<td>11.8%***</td>
<td>-11.8%***</td>
<td>-26.8%***</td>
<td>-0.1%</td>
<td>-3.6%**</td>
</tr>
<tr>
<td>3rd generation</td>
<td>EXCEL</td>
<td>22.2%</td>
<td>11.4%***</td>
<td>20.5%***</td>
<td>3.6%***</td>
<td>-1.9%</td>
<td>-2.4%</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>ACS estimate</th>
<th>General population English (column a)</th>
<th>Hispanic English (column b)</th>
<th>Hispanic-enhanced English (column c)</th>
<th>General-population all (column d)</th>
<th>Hispanic all (column e)</th>
<th>Hispanic-enhanced all (column f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>EXCEL 44.0%</td>
<td>-6.8%**</td>
<td>-6.1%**</td>
<td>-4.5%</td>
<td>-5.2%**</td>
<td>-0.6%</td>
<td>0.1%</td>
</tr>
<tr>
<td></td>
<td>GSS -3.9%</td>
<td>0.7%</td>
<td>4.0%</td>
<td>2.8%</td>
<td>4.2%**</td>
<td>4.3%**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NSL -4.4%</td>
<td>-4.4%</td>
<td>7.5%**</td>
<td>-0.6%</td>
<td>4.3%*</td>
<td>4.3%*</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>EXCEL 63.0%</td>
<td>2.6%</td>
<td>0.6%</td>
<td>2.5%</td>
<td>-1.8%</td>
<td>-2.7%</td>
<td>-2.2%</td>
</tr>
<tr>
<td></td>
<td>GSS 2.8%</td>
<td>-1.4%</td>
<td>2.0%</td>
<td>3.7%*</td>
<td>3.0%</td>
<td>3.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NSL -6.1%*</td>
<td>-6.1%*</td>
<td>-2.3%</td>
<td>-3.3%</td>
<td>-3.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average bias</td>
<td>EXCEL</td>
<td>10.9%&lt;sup&gt;c,d,e,f&lt;/sup&gt;</td>
<td>9.2%&lt;sup&gt;d,e,f&lt;/sup&gt;</td>
<td>3.4%&lt;sup&gt;a,b,c,f&lt;/sup&gt;</td>
<td>3.1%&lt;sup&gt;a,b,f&lt;/sup&gt;</td>
<td>1.9%&lt;sup&gt;a,b,c&lt;/sup&gt;</td>
<td>1.0%&lt;sup&gt;a,b,c,d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>GSS 7.1%&lt;sup&gt;b,c,d,e,f&lt;/sup&gt;</td>
<td>4.1%&lt;sup&gt;a,c,e,f&lt;/sup&gt;</td>
<td>1.1%&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>2.0%&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>1.3%&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>1.2%&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NSL 10.6%&lt;sup&gt;c,e,f&lt;/sup&gt;</td>
<td>2.8%&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.8%&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

**Note.**—Superscript letters in the last three rows of the table denote significant statistical difference between the contents of marked cells and the contents of other cells in the same row. For example: in the row for Average bias: EXCEL, the superscripts c, d, e, and f next to 10.9% indicate that there is a statistically significant difference between 10.9% and the figures in the same row, columns c, d, e, and f. *<i>p</i> < .05; **<i>p</i> < .01; ***<i>p</i> < .001
PHC surveys have matched quite well on this question compared to the gold
standard (Hakimzadeh and Cohn 2007; Kiley et al. 2010).

Just as there is no point in comparing first generation on the samples com-
parison (in table 5), we also do not make other comparisons to variables that
were used as poststratification parameters. Metrics that are used in weighting
can be useful points of comparison to understand the degree to which the
self-weighting samples may be biased, since those samples were weighted
to demographic parameters for the entire population, not just Hispanics. As
such, we utilize gold-standard Hispanic variables for age (percent of respond-
ents aged 18–29 years old) and education (percent not attaining a high school
diploma) to measure potential bias in the self-weighting sample.

Again, because there are few gold standards available that reside in all the
data utilized in these analyses, we were limited in additional potential compar-
isons. Nevertheless, we were able to make comparisons using two additional
variables included in each of the data sets: marital status (percent married with
spouse present) and employment (percent working at least part-time).

**Results**

**SAMPLE COMPARISONS**

According to table 4, which provides results of the comparisons by sampling
design, the RDD weight attains the smallest design effect in both studies, and
understandably, the increase in design effect for the stratified weight is due
to the need to adjust for the disproportionate stratification of the sample. As
reflected by the design effects in columns a and d, the use of the landline-only
surname sample is associated with a nearly 200 percent increase in design
effect in comparison to the RDD design effect in the NSL data. The top-market
weight attains only a slightly higher design effect than the stratified weight in
the NSL data, and is in fact lower in the EXCEL data.

On average, the surname sample exhibits over twice the bias of the strati-
fied sample and significantly more bias than other samples as well, with the
exception of the top-market sample, as illustrated in figure 1. The surname
design does a particularly poor job of attaining third-generation Hispanics and
Hispanics who speak English very well. With both data sources averaged, the
top-market design is approximately six percent less likely to be married, third
generation, or able to speak English very well. None of the samples is within
six percentage points of being able to speak English very well.

**LANGUAGE AND WEIGHTING COMPARISONS**

As shown in table 5, the English-only surveys of Hispanics introduce a sig-
nificantly greater amount of weighting variance to get such samples to match
weighting parameters. While the English-only general-population weight’s
design effect is only marginally larger than the general-population (all) data
in *EXCEL* data (and in fact slightly lower for the GSS data), the English-only Hispanic design effect is 1.5 times larger than the full *EXCEL* sample and the English-only enhanced Hispanic design effect is 2.2 times greater in *EXCEL*. In the GSS data, the pattern is similar but not as pronounced. The design effect did not significantly differ for the full sample in comparing the general-population, Hispanic, and enhanced Hispanic weights, in any data source; however, the design effect increased in the English-only samples from 1.37 for the general-population weight to 2.62 for the enhanced Hispanic weight in the *EXCEL* data; it increased similarly from 1.38 to 1.84 in the GSS data, and increased substantially as well in the NSL.

There is a clear trend toward less bias as the samples become dual language and utilize the enhanced weights (see figure 2). On average, bias associated with the general-population English-only sample and weight was over 10 percent, across all measures, in *EXCEL* data, and 7.1 percent in GSS data. While age is a weighting variable, the general-population weight in *EXCEL* produces data for Hispanics that have 12.6 percent more respondents in the 18-to-29 age range than data for Hispanics based on the ACS, and 17.4 percent less likely to be without a high school diploma. GSS data follow the same pattern (7.7 percent too young and 14.2 percent less likely to be without a high school diploma). The general-population English-only weight attained either the most or second-most bias of any weight on every measure. As expected, all English-only weights performed quite poorly at attaining a weighted sample with regard to the language measure. Specifically, across all three English-only weights and the two available sample sources on this measure, and compared to the ACS point estimate, the data included 21 percent more respondents reporting being able to speak English very well. And without the correction for “born/years in the United States” that was made...
for the Hispanic-enhanced English-only weight, the generation variable was also substantially biased across all three data sources. Generally speaking, the bias for marital status was significantly reduced when more precise weighting was applied, from English-only to all-Hispanic weights and from general-population to enhanced-Hispanic weights, within EXCEL and GSS data. That said, there was no clear trend in the reduction of bias for employment status, with the basic Hispanic English-only weight and the general-population (all) weight attaining the least amount of bias. Finally, we find no bias in the percent of Mexicans in the enhanced weights, unsurprising given that this was a specific weighting parameter. However, we do find a large degree of bias in the English-only general-population weight, in both EXCEL and GSS data.

**Discussion and Conclusions**

Hispanic research has utilized a wide range of sampling designs and operational procedures—including general-population samples with less than fully specified weights, English-only samples, and non-SRS sampling methods—without any real understanding of the potential error inherent in any single design choice or combination of design choices. This article set out to document the potential costs of these design choices in order to establish parameters for best practices for Hispanic research, broadly defined as techniques that minimize survey error, specifically variance and bias. We analyzed three very different data sets, the goal being that if the results from each are consistent with one another, then we can be all the more confident that the results are not an artifact of a feature of any one data source. Overall, despite the differences
in methodologies—a short-field, low-cost RDD survey (EXCEL); an in-person, high-response rate survey (GSS); and a Hispanic-specific study falling in the middle in terms of effort and field length (NSL)—the patterns and results are highly similar, with only a few exceptions, across all the estimates tested, in both measured variance and bias. Below, we summarize our findings and consider the implications for Hispanic research.

One common technique in Hispanic research is the culling of Hispanic interviews from omnibus or other general-population surveys. The reasons for this are straightforward: The cost per interview in utilizing this method is often no greater than the cost for interviews of the general population, and there is a wide range of publicly available large-sample surveys of the general population, such as the ANES, the GSS, and the National Health Interview Survey. Notably, when conducted as part of a dual-language survey, Hispanic interviews are attained from a simple random sample and therefore have a great deal of validity in terms of sampling. That said, researchers who utilize general-population surveys for analyses of Hispanics should consider creating weights specific to Hispanic-population parameters, since our analysis found considerably more bias when using the general-population weights provided by the surveys we analyzed, compared to the Hispanic-specific weights generated for our analysis.

It is interesting to note that since simple random telephone samples typically underrepresent younger respondents, cell-phone-only households, and ethnic minorities, including Hispanics, general-population weighting results in larger weights given to precisely these three groups. But because Hispanics are comparatively young and tend to use only cell phones, general-population weights overcorrect for these two parameters within Hispanic samples in particular. Indeed, our data find that the English-only general-population weight produces data showing 13 percent younger Hispanics, and 3 percent more who use only cell phones. The general-population (all) weight is similarly 8 percent too young and a considerable 10 percent greater in cell-phone-only use.

Of greater concern is the use of English-only samples for Hispanic research, particularly when Hispanics are culled from surveys of the general population. Not only will the weighting in English-only surveys of Hispanics have to “work harder” to balance the samples, leading to significantly higher design effects, but weighting alone will not be fully sufficient in eliminating the bias inherent in English-only samples. The exception to this was our comparison of our enhanced Hispanic weighting routine. With this weight, we failed to find significant average bias compared to surveys that interviewed in both English and Spanish. However, the design effects of the English-only enhanced Hispanic weight were excessive, at an average of 2.5 across all three data sources.

Overall, we recommend the use of Hispanic-specific weighting parameters for Hispanic research, including Hispanic-specific point estimates such as
heritage and born in the United States/years in the United States. Even in the samples that included English and Spanish interviewing, while there was no overall significant drop in bias, there were estimate-specific measures, including a considerable improvement in heritage (percent Mexican), the ability to speak English, and generation. Overall, best practices with regard to Hispanic research include bilingual interviewing and weighting to Hispanic-specific population parameters.

Our study also highlights the danger in utilizing surname samples, in both the substantial amount of variance that is introduced in trying to correct the bias inherent in such a sample, and then the failure of weighting to fully correct for bias. Even though we utilized our enhanced weighting procedure in the analysis, the surname sample was still, on average, 6 percent short of third-generation respondents and short of respondents who can speak English very well by the same amount.

It is interesting to note that the top-market design, though still more biased on some measures than the SRS samples, performed significantly better than the surname design, and with a degree of variance on par with the stratified design. This points to the viability of utilizing such a design for in-person studies of Hispanics, though with a survey incidence on par with the NSL stratified design, and half the coverage, there is no reason to execute a top-market design via telephone.

By focusing on gold standards and analyzing three very different data sets, we were not able to extend our analysis to any attitudinal or behavioral variables such as health measures or political behaviors. That said, there are numerous articles, many of which are noted in this paper, that already link differences in such variables by those metrics utilized here, particularly those that happen to show the greatest degree of bias, such as generation and language use. By extension, our substantial findings on these measures should naturally impact variables highly correlated with these two measures.

In the past decade, Hispanic growth made up more than half of the nation’s overall increase in population. In addition, Hispanics continue to be one of the fastest-growing racial/ethnic groups in the United States in key measures such as purchasing power and the percentage of new registered voters. That means interest in Hispanic research will only continue to grow. Therefore, it is imperative to understand the relative costs in bias and variance of the multitude of research designs available to survey researchers. Our study finds that there are no longer any quick fixes. In the past, before the age of cell phones and before the recent influx of new immigrants, surname or English-only designs might have sufficed. However, with the size and diversity of today’s U.S. Hispanic population, researchers must be very careful to enroll sample designs, language strategies, and weighting procedures that will accurately represent that population.
Appendix. Questions Used in the Analysis

(NSL, EXCEL, and GSS) What is your age?

(NSL, EXCEL) What is the last grade or class that you completed in school?

(GSS) What is the highest grade in elementary school or high school that you finished and got credit for? Did you ever get a high school diploma or a GED certificate? Did you ever complete one or more years of college for credit—not including schooling such as business college, technical or vocational school? How many years did you complete? Do you have any college degrees? What degree or degrees?

(NSL, EXCEL) Just to confirm, are you, yourself, of Hispanic or Latino origin or descent? Now I want to ask you about you and your family’s heritage. Are you Mexican, Puerto Rican, Cuban, Dominican, Salvadoran, or are you and your ancestors from another country?

(GSS) Are you Spanish, Hispanic, or Latino/Latina? Which group are you from?

(NSL, EXCEL) Would you say you can carry on a conversation in English, both understanding and speaking—very well, pretty well, just a little, or not at all?

(NSL, EXCEL) Were you born on the island of Puerto Rico, in the United States, or in another country?

(GSS) Were you born in this country?

(NSL, EXCEL) How many years have you lived in the (continental) United States?

(NSL, EXCEL) Was your mother born on the island of Puerto Rico, in the United States, or in another country?

Was your father born on the island of Puerto Rico, in the United States, or in another country?

(GSS) Were both your parents born in this country? (all combinations recorded)

(NSL, EXCEL) Are you currently married, do you have a partner, are you widowed, divorced, or separated, or have you never been married?

(GSS)
Are you currently—married, widowed, divorced, separated, or have you never been married?

(NSL, EXCEL)
Are you now employed full-time, part-time, or not employed?

(GSS)
Last week were you working full-time, part-time, going to school, keeping house, or what?

(NSL, EXCEL, GSS)
Record language of interview.

References


