

Bias Reduction through Rural Coverage for the AmeriSpeak Panel

Steven Pedlow¹, Jie Zhao¹

¹NORC at the University of Chicago, 55 E. Monroe, Chicago, IL 60603

Abstract

The NORC AmeriSpeak Panel is a household, multi-client panel to support NORC's mission to deliver reliable data and rigorous analysis to guide critical programmatic, business, and policy decisions. AmeriSpeak uses the NORC 2010 National Sampling Frame to draw an address-based nationally representative sample of all Americans with full urban and rural coverage.

The rural coverage provided by the NORC 2010 National Sampling Frame allows the NORC AmeriSpeak Panel to have the most coverage of the U.S. population among multi-client panels in the U.S. While there is great difficulty in determining the coverage achieved by opt-in non-probability panels and by phone surveys, the U.S. postal address database used for address-based sampling (ABS) provides city-style addresses for 92 percent of the U.S. population. This excludes post office boxes and rural route addresses that can receive mail, but are not reachable through in-person or telephone modes. In areas not well covered by the postal address database, the NORC 2010 National Sampling Frame enhances the ABS frame with in-person listing. We estimate that this increases the coverage by five percentage points.

We also attempt to reduce bias by including these households in rural areas that would be missed by ABS studies. In the first year of AmeriSpeak recruitment, 508 out of 7,881 (6.4 percent) of the recruited households were recruited in "rural" segments that were not well covered by the U.S. postal address database's city-style addresses, and were therefore listed in-person. Our focus in this paper is a comparison of AmeriSpeak estimates with and without this rural component against national benchmarks. Our main question is, "For which variables does the rural coverage improve the representativeness of AmeriSpeak?"

1. Introduction to AmeriSpeak

Funded and operated by NORC at the University of Chicago, AmeriSpeak® is a probability-based panel designed to be representative of the U.S. household population. Randomly selected U.S. households are sampled with a known, non-zero probability of selection from the NORC National Sampling Frame, and then invited to join the AmeriSpeak panel by U.S. mail, telephone, and field interviewers (face-to-face). AmeriSpeak panelists participate in NORC studies or studies conducted by NORC on behalf of governmental agencies, academic researchers, and media and commercial organizations.

As a probability-based panel, AmeriSpeak is designed to be a nationally representative panel of U.S. adults. During its initial year in 2015, 7,752 households were recruited, and

the goal for 2016 is to reach a panel size of 17,500 households. Future plans include the possibility to reach a panel size of 30,000 and also include the development of custom online panels that can include specific population samples such as military veterans and low-income households.

There are three benefits to the AmeriSpeak methodology. First, higher coverage is obtained by using rural addresses listed in-person as well as the city-style addresses in the United States Postal Service (USPS) Delivery Sequence File (DSF). This provides a higher coverage than phone surveys, even those with cell coverage. Technically, mail surveys can increase coverage by using P.O. boxes and other rural addresses in the DSF that we drop because they cannot be visited. Our complete listings in selected rural areas provide more coverage than the partial listings in the DSF, though. The second benefit is that AmeriSpeak provides a higher response rate because our in-person recruitment determines the overall weighted response rate. There is a variance cost due to the subsampling for in-person recruitment, however. Third, NORC (Dennis et al., 2015) presented evidence at the 2015 AAPOR Annual Meeting in Hollywood, Florida that our in-person recruiting does result in a more representative web panel. Higher response rates mean that there is potentially less self-selection bias. The topic for this talk is whether the higher coverage due to full rural representation for AmeriSpeak also provides a more representative sample.

The recruitment process for AmeriSpeak occurs in two stages: 1) initial recruitment by mail and phone; and 2) nonresponse follow-up (NRFU) in-person recruitment. The initial recruitment packet is sent by mail and we send four separate USPS mail-based contacts. We also match the selected addresses to phone numbers and recruit with this mode as well. Email addresses were not very productive in 2015, so this contact mode was dropped for 2016. Beginning with the recruitment package, sampled households are provided the information necessary to register online at the AmeriSpeak panel portal or by calling the study's toll-free telephone number for phone-based registration. Once they agree to join AmeriSpeak, they fill out a required profile survey and then they are eligible for client surveys. The profile survey is meant to introduce taking AmeriSpeak surveys, make it fun, and these surveys will help us target certain client surveys. After a certain amount of effort and time, we select a subset of non-respondents for in-person follow-up. Mail and phone recruiting is continued from the central office only for those not selected for in-person follow-up.

After one adult in the household becomes an active AmeriSpeak panel member, other adults in the household are invited to join AmeriSpeak. These additional adults are required to complete the same registration and introduction survey before becoming an active AmeriSpeak panel member. After the initial profiling is completed via the introduction survey, subsequent NORC surveys invite the panelist to provide additional background information about public affairs, health services, health conditions, financial services, and computer and internet usage (among other topics).

1.1 AmeriSpeak Samples So Far

In the January through June 2015 time frame, AmeriSpeak fielded NORC-designed surveys collecting background data from panelists with respect to political attitudes and behavior, public affairs, health conditions, use of health services, and financial status. In June, 2015, AmeriSpeak began client studies with the launch of AmeriSpeak Omnibus. AmeriSpeak has completed or is currently conducting studies either commissioned or funded by the Robert Wood Johnson Foundation, Kaiser Family Foundation, Stanford

University, Northwestern University, Consumer Reports, and the Associated Press. NORC is currently including AmeriSpeak as a sample source in proposals to RFPs issued by Federal agencies.

AmeriSpeak panel members typically participate in AmeriSpeak web-based or phone-based studies two to three times a month. Surveys are usually 10 to 20 minutes in length; however, longer surveys are permitted with the allowance for additional incentives to recognize the time spent by AmeriSpeak panelists. AmeriSpeak panel members who do not have internet access for survey taking are enrolled in the telephone mode for client studies and profile surveys, while the internet users are enrolled in the web mode.

1.2 AmeriSpeak Response Rates

The cumulative AAPOR RR3 (response rate) for the panel recruitment between October 2014 and August 2015 (this includes two pilots and the first year of full recruitment) is 36.6 percent (weighted to take into account selection probabilities). The estimated cumulative AAPOR RR3 for client surveys is 13 percent to 20 percent (varying according to study parameters and taking into account all sources of non-response including panel recruitment, panel household attrition, and survey participation). Table 1 gives some key statistics for the AmeriSpeak panelists as of August 2015:

Table 1: AmeriSpeak Key Statistics for the 2014-2015 recruited households

Recruited via the non-response follow-up	48 percent
Prefer Telephone data collection	24 percent
Cell-Phone only or Cell-Phone Mostly	71 percent
African-American households	19 percent
Hispanic households	13 percent
Household Income less than \$35,000	33 percent

2. AmeriSpeak Sampling Strategy

Web panels can be recruited with a variety of probability and non-probability methods. Among the probability methods that can be used are address-based sampling methods using the U.S. Postal Service (USPS) database called the Delivery Sequence File (DSF) alone. Based upon our NORC National Sampling Frame research, we believe such a methodology can deliver a 92 coverage rate of U.S. households. Iannacchione et al. (2007) estimates the DSF coverage at 93 percent based on a smaller sample of segments. Random-digit dialing methods can also deliver a probability sample of households, though the coverage is much less than ABS sampling (Barron et al, 2010), and is largely unknown for cell-phone samples. There are also a variety of web panels that have non-probability sources that have low coverage.

AmeriSpeak households are selected initially from NORC's National Sampling Frame, an area probability sample funded and managed by NORC and used for all national in-person studies at NORC, including the General Social Survey funded by the National Science Foundation and the Survey of Consumer Finances sponsored by the U.S. Federal Reserve Board in cooperation with the U.S. Treasury Department. The General Social Survey (GSS) uses the National Sampling Frame every other year and the Survey of Consumer Finances (SCF) uses the National Sampling Frame every three years. Other projects that have used the 2010 National Sampling Frame include AmeriSpeak, the

Landmark Spirituality and Health Survey, and the third round of the National Social Life, Health, and Aging Project (NSHAP).

NORC designs a new National Sampling Frame after every Decennial Census. NORC's National Sampling Frame is designed to provide 97 percent sample coverage by supplementing the USPS DSF with in-person listings in rural areas where DSF coverage is not at least 90 percent. The NORC National Sampling Frame contains almost 3 million households, including over 80,000 rural households added through the in-person listing of households.

The USPS DSF is used as the frame of addresses wherever it is determined to be complete enough. Some addresses in the DSF are not locatable or geocodeable; these include post office boxes and rural route addresses. We refer to such addresses as non-city-style addresses. These addresses are mailable, but can't be matched to phone numbers or visited in person, and they can't be tracked in a multi-mode survey such as AmeriSpeak. We receive monthly updates of the post office addresses, so we are able to update the National Sampling Frame in urban areas before major sample draws, but we do not update the in-person listings.

For the 2014-2015 AmeriSpeak recruitment, a stratified random sampling approach was used to select sample units from the National Sampling Frame, including an oversample of households in Census tracts or block groups higher in young adults, Hispanics, or African-Americans. AmeriSpeak attempts to recruit all English-speaking members age 18 and older in the sampled households. Spanish-language recruitment was introduced in February 2016.

2.1 The NORC National Sampling Frame

The 2010 National Sampling Frame uses a two-stage probability sample design to select a representative sample of households in the United States. At the first stage, the sampling unit is a National Sampling Frame Area (NFA), which is either an entire metropolitan area (made up of one or more counties) or a county (some counties were combined so that each NFA contains a population of at least 10,000). The 38 largest NFAs with a population of at least 1,543,728 (0.5 percent of the 2010 Census U.S. population) were selected with certainty; these areas have a high population density, and are dominated by tracts with city-style addresses. These areas contain 56 percent of the population within 8 percent of the geographic area of the U.S. The remaining areas were stratified into areas where city-style addresses predominate (we consider these "urban" areas) and the remaining areas which are less likely to have city-style addresses. The latter stratum ("rural" areas) comprises 81 percent of the geographic area, but only 14 percent of the U.S. population.

Within the selected National Sampling Frame Areas, the second stage sampling unit is a segment, defined either in terms of Census tracts (in "urban" areas well covered by the DSF) or block groups (in "rural" areas where DSF coverage might not be sufficient), containing at least 300 housing units according to the 2010 Census. A stratified probability sample of 1,514 segments was selected with probability proportional to size. We determined whether or not the DSF could be used for each of these segments. We compared the total number of city-style addresses in the DSF against 2010 Decennial Census counts. If the coverage was less than 90 percent, we used in-person listing. Since we didn't want to list entire census tracts, if a census tract had less than 90 percent

coverage, we selected a random block group and re-checked coverage for just the block group. The number of block groups that required in-person listing was 121 (8 percent).

2.2 AmeriSpeak Coverage

As discussed above, using our National Sampling Frame counts and census counts, we estimate that the city-style addresses alone cover approximately 92 percent of U.S. housing units. This ignores non-city-style addresses such as post office boxes and rural route addresses. We estimate that sample coverage is improved 5 percentage points to 97 percent as a result of the in-person listing.

We estimate the undercoverage at 3 percent. This is because, for cost-efficiency, we use the DSF even if we estimate its coverage is only 90 percent. We estimate the human error during in-person listing at 1 percent, and there are also errors in the DSF. Finally, time lags occur in both the DSF and listing. While we have the ability to update the DSF monthly, we do not update our in-person listings due to cost constraints (these get updated every decade when a new National Sampling Frame is designed).

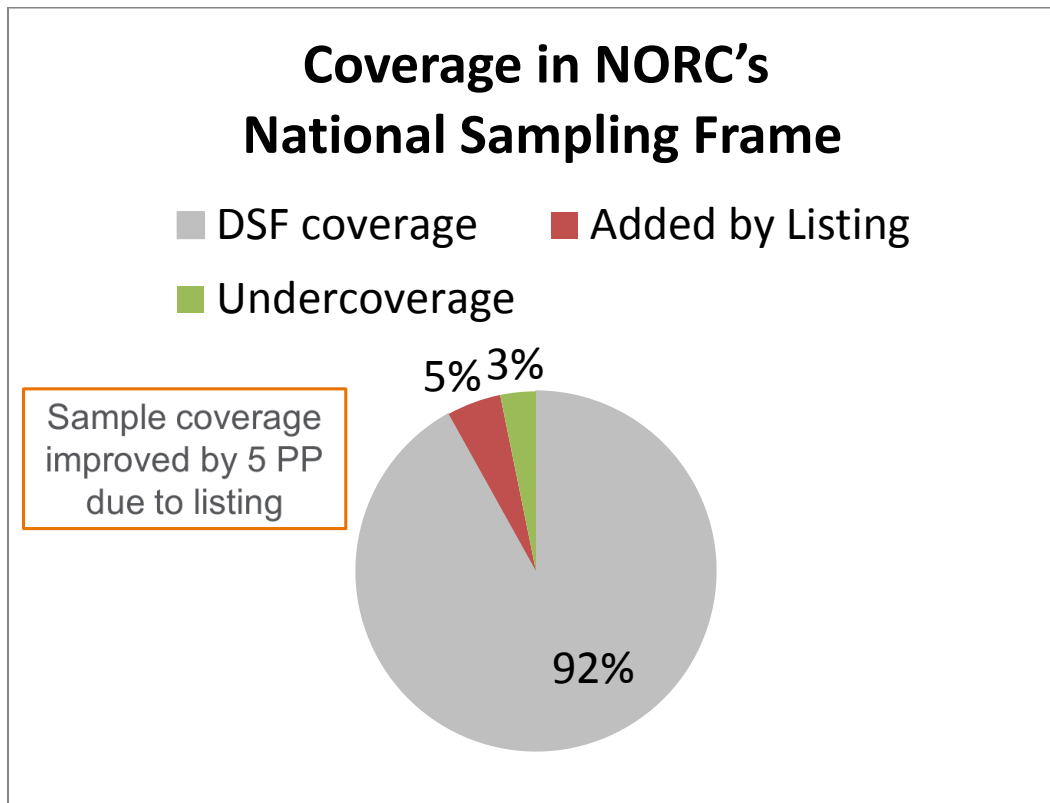


Figure 1: Coverage in NORC's National Sampling Frame

3. Rural Recruit Results

Dennis et al. (2015) showed that the non-response follow-up (NRFU) in-person recruitment did increase the representativeness of the AmeriSpeak panel. The focus of this paper is on the rural coverage, and whether the additional in-person listing, added to the DSF portion of the recruited sample, increases the representativeness of the AmeriSpeak panel.

We compared AmeriSpeak against benchmarks on a number of variables, all of which are included in this paper. To determine if the rural coverage increased the representativeness of AmeriSpeak, we also split the AmeriSpeak panel into two parts: those recruited through the DSF; and those recruited through our listed address sample frame.

All of the following graphs have four bars in this order: DSF, AmeriSpeak (all cases together), an appropriate national Benchmark, and List. All of the estimates reported here are unweighted. The AmeriSpeak bar is a weighted average of the DSF and List bars, but is much closer to the DSF bar because most AmeriSpeak participants (93.6 percent) were recruited through the DSF address frame.

This ordering of the four bars makes it easy to see whether the rural portion increases representativeness or not. To do so, you only have to look at the left three bars, in which the total AmeriSpeak sample estimate is between the DSF-only estimate and the national benchmark. If these three bars are monotonic (↗ or ↘), then the total AmeriSpeak sample estimate is between the DSF-only estimate and the national benchmark, which means that the rural panelists have moved the AmeriSpeak average closer to the national benchmark for that variable.

3.1 Demographics

Figure 2 looks at the demographics, but this figure does not speak to whether the rural coverage improves the representativeness of the AmeriSpeak panel. On the left, we see that the rural sample has a higher percentage of females, which may be due to the in-person recruitment methodology. Most surveys, unless they are very controlled, tend to have more female respondents than male respondents due to availability (at home) and/or response rate issues. Figure 2 also shows that the Hispanic percentage in AmeriSpeak is lower than the national Census benchmark, especially in the rural areas. It is not surprising that the rural areas would be lower in Hispanics, but AmeriSpeak is working to

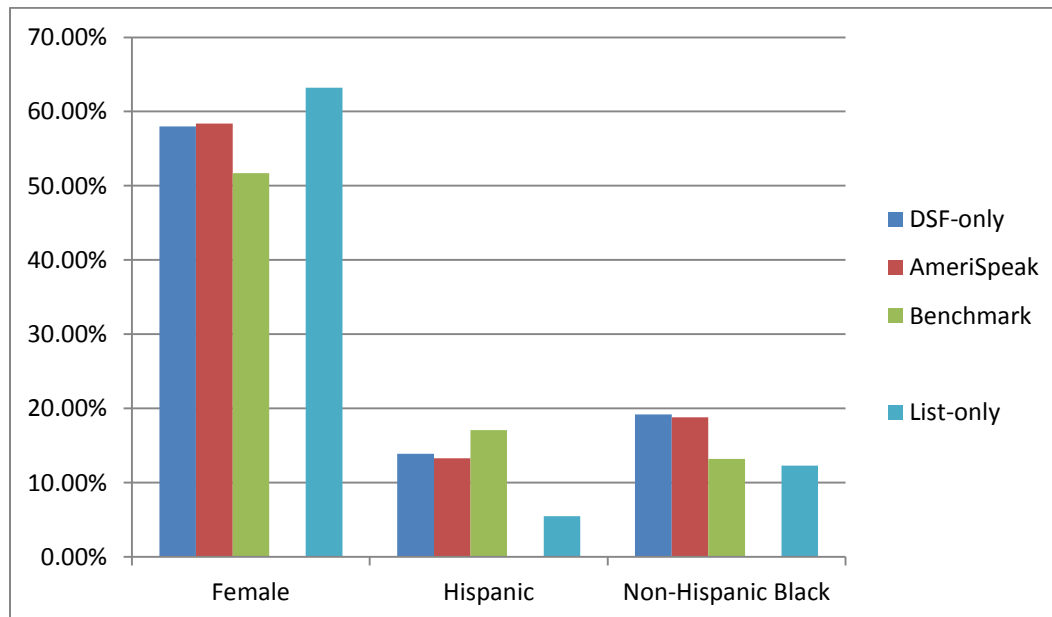


Figure 2: Demographics: Gender, Hispanic origin, and Non-Hispanic Blacks

erase its shortfall in Hispanics by adding Spanish-language recruitment materials in 2016. Non-Hispanic Blacks have been oversampled by AmeriSpeak, resulting in a percentage of panelists higher than the national Census benchmark. Again, it is not surprising that the rural areas would be lower in non-Hispanic Blacks.

Figure 3 looks at more demographics: age and marital status. Just like for minority race/ethnicity groups, AmeriSpeak oversamples areas high in young adults. As a result, AmeriSpeak has a slightly higher percentage of panelists in the 18-34 age range. Figure 3 also shows there are fewer young adults in the rural areas, though there is not much of a shortfall there. AmeriSpeak does have a higher response rate for senior citizens aged 65 and up, and this is shown in Figure 3 by a surplus in this age group. The surplus is even larger in the rural areas, which is due to the older demographic trend in rural areas. Marital status is the first variable for which we can actually determine whether the rural sample improves AmeriSpeak representativeness, and it does. We have a lower percentage of married panelists in AmeriSpeak, but the rural percentage married is closer to the national Census benchmark and reduces the shortfall from 6.7 percent to 6.1 percent.

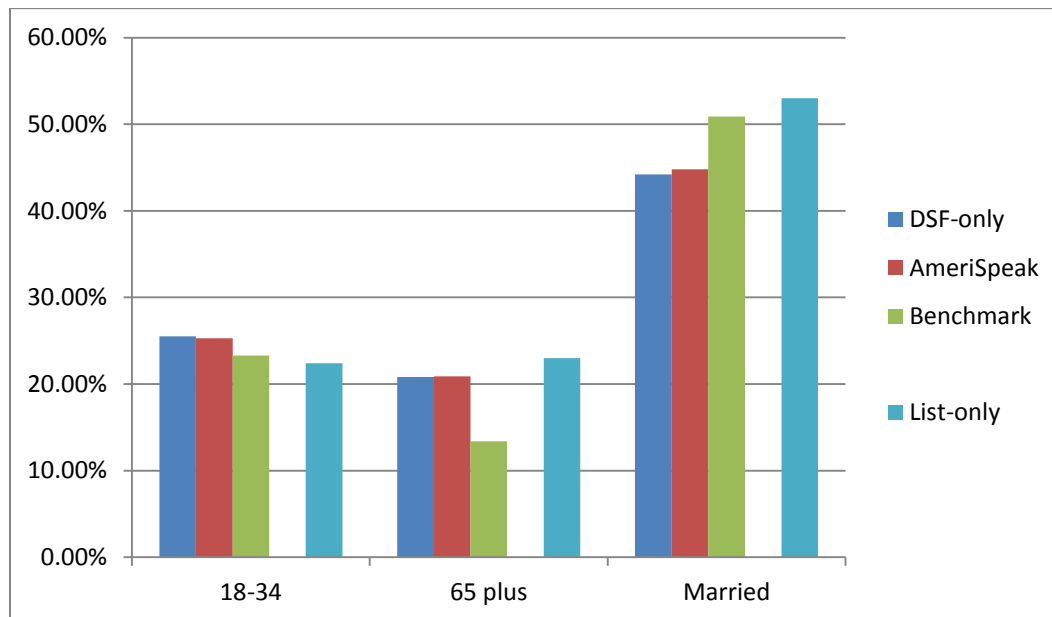


Figure 3: Demographics: Age and Marital Status

3.2 Income and Education

Figure 4 shows the results for education and income variables. For education, the rural coverage improves representativeness on both the upper and lower ends. For both ends of the scale (High School or Less and College Graduates), the rural portion is close to the national benchmarks and makes the AmeriSpeak sample more representative. Many surveys, especially those involving electronic data collection, find that they recruit a sample that has a higher than average education level. The rural coverage reduces the shortfall in the category High School or Less from 10.3 percent to 9.8 percent. Similarly, the rural coverage reduces the surplus in College Graduates from 8.0 percent to 7.5 percent.

For income, the rural coverage improves representativeness for lower income households, but not for higher income households. The rural sample includes many lower income households, reducing the shortfall in households with a household income under \$35,000 from 1.8% to 1.1%. However, the shortfall in households with a household income above \$50,000 rises from 5.0% to 5.7%.

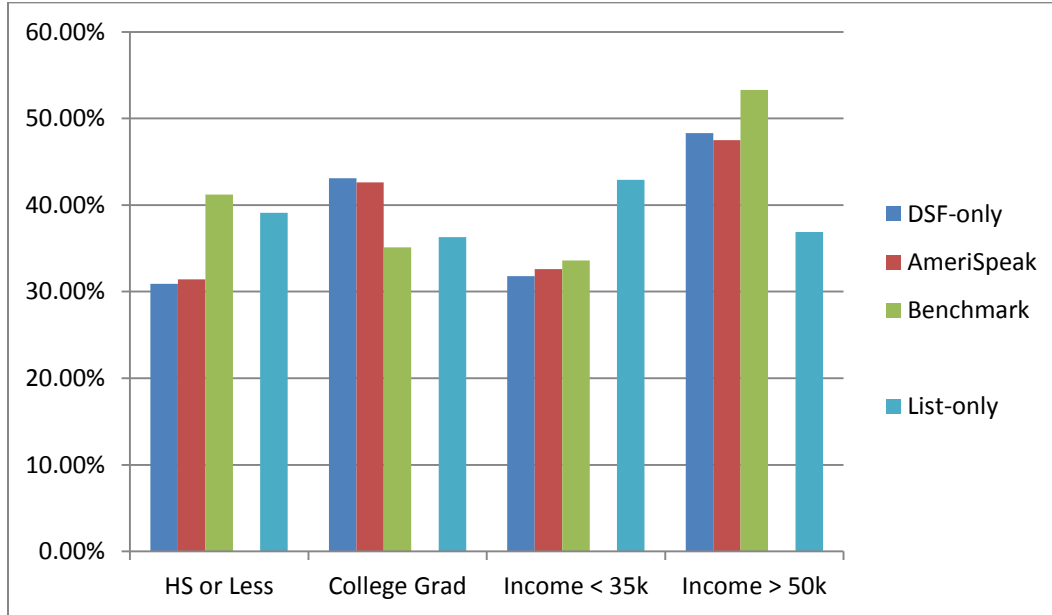


Figure 4: Education and Income

3.3 Household Size and Ownership

Figure 5 shows variables on household size and ownership status. AmeriSpeak has a shortfall of single-person households and it is not surprising that rural households are

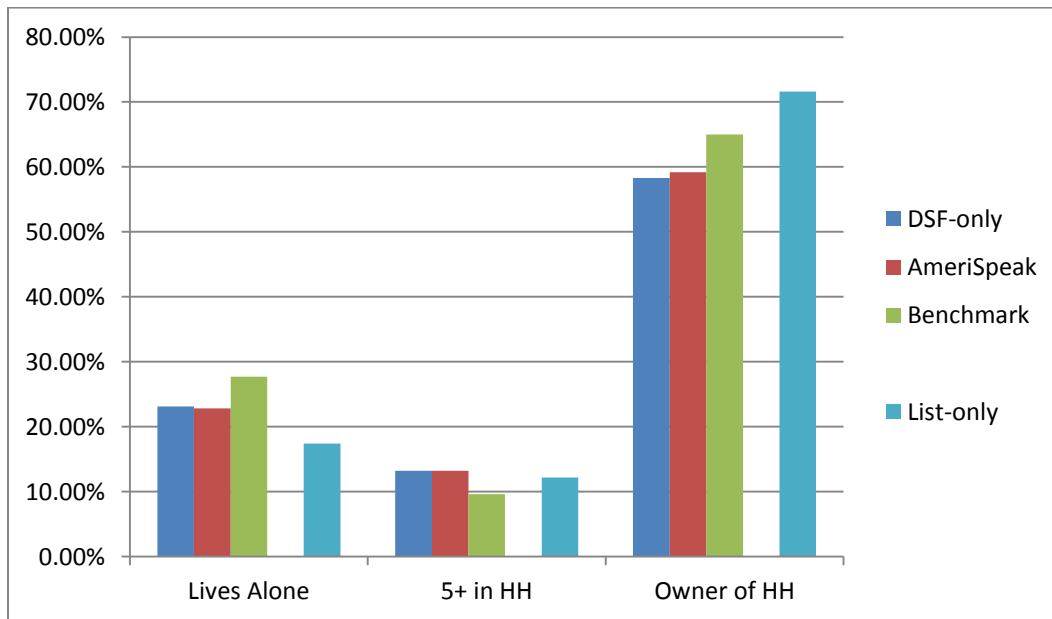


Figure 5: Household Size and Household Ownership Status

less likely to be single-person households. With rural coverage, the shortfall increases from 4.6 percent to 4.9 percent. Rural households are slightly less likely (by 1 percent) to be large households, defined as five or more persons, but the rural coverage has no impact on the 3.6 percent surplus of large households in AmeriSpeak.

Rural coverage does improve the representativeness on home ownership status, however. AmeriSpeak has a shortfall of home owners, but the rural coverage decreases this shortfall from 6.7 percent to 5.8 percent.

3.4 Internet and Phone Service

Figure 6 shows whether or not the household has internet access at home (AmeriSpeak allows panelists to request that they can complete AmeriSpeak surveys by telephone, so we are not restricted to internet-only panelists), as well as what type of phone service the household has. Blumberg and Lake (2016) has the latest national benchmarks used for these variables, which are taken from the National Health Interview Survey. While AmeriSpeak includes non-internet households, there is still a shortfall in non-internet households, but rural coverage decreases this shortfall from 6.6 percent to 6.3 percent. Rural households are more likely to not have phone service, but the AmeriSpeak shortfall is 1.9 percent with or without rural coverage. AmeriSpeak has a surplus of landline-only households and a shortfall of cell-only households, but the rural coverage improves the representativeness of AmeriSpeak on both. For landline-only households, the surplus is reduced from 5.2 percent to 4.9 percent; for cell-only households, the shortfall is reduced from 4.9 percent to 4.2 percent.

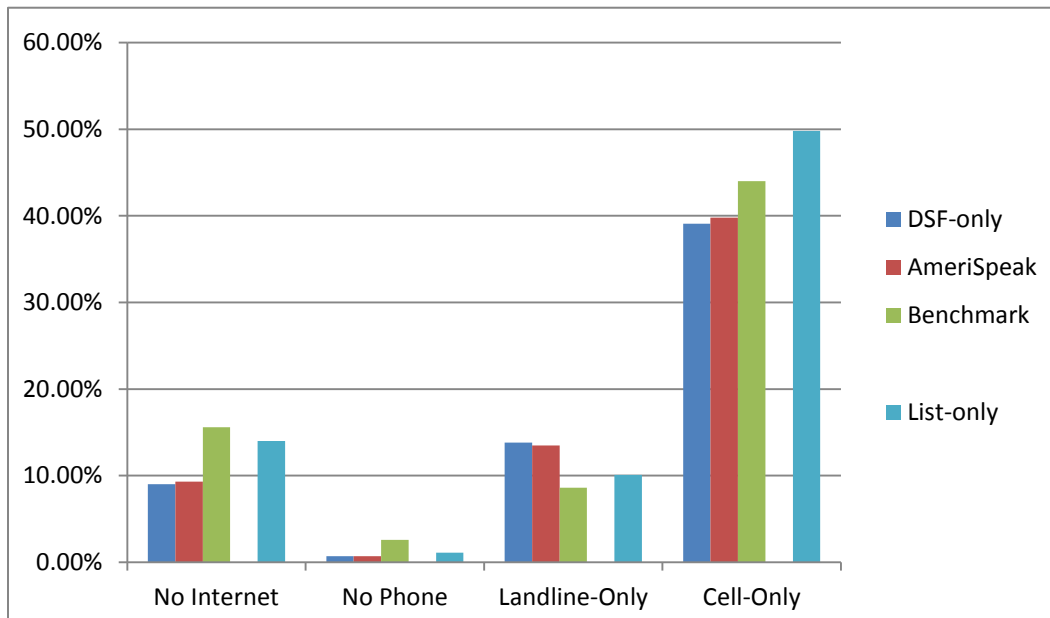


Figure 6: Internet and Phone Status

3.5 Political Affiliation

Figure 7 shows how our AmeriSpeak panel compares in political affiliation with national Gallup poll numbers that we used as our benchmark. The AmeriSpeak panel overall has more liberal panelists and fewer moderate panelists while the percentage of conservative panelists is very close to the Gallup benchmark. AmeriSpeak's rural coverage does improve the representativeness on liberal voters (surplus is reduced from 6.7 percent to 6.3 percent), but not on moderate (shortfall increases from 6.5 percent to 6.6 percent) and conservative voters (rural voters are more conservative, but AmeriSpeak is very close to the Gallup benchmark with or without the rural coverage).

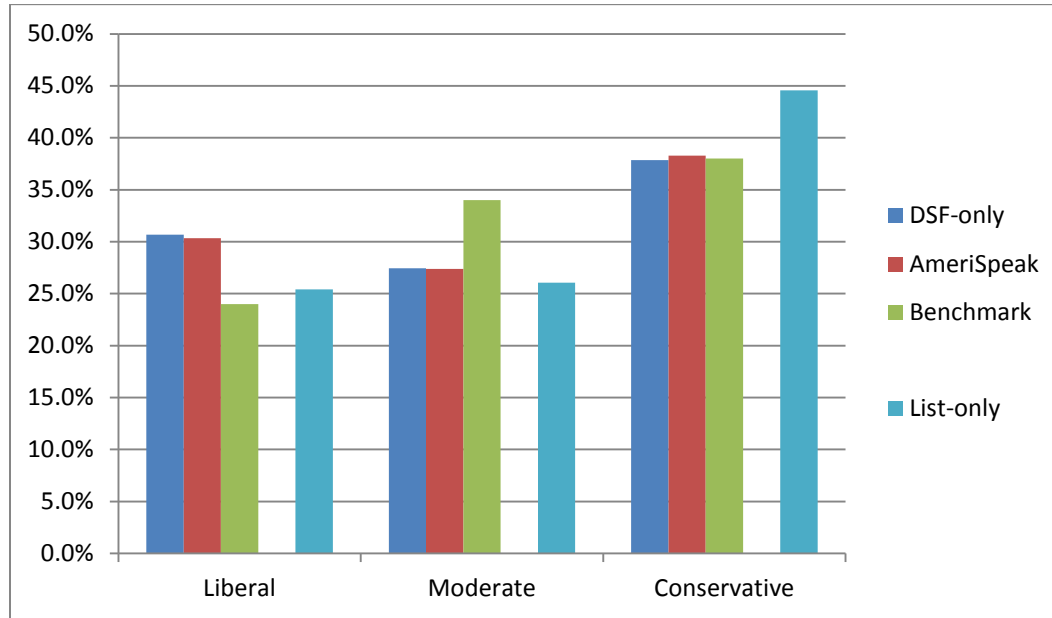


Figure 7: Political Affiliation

3.6 Employment and Pet Ownership

Figure 8 below shows our last two comparisons on employment and pet ownership. Rural inclusion did not improve AmeriSpeak's representativeness on either of these two variables. AmeriSpeak has a slightly lower employment rate than the Bureau of Labor Statistics, but the rural coverage increases the shortfall from 1.3 percent to 1.9 percent. For pet ownership, AmeriSpeak has a slightly higher rate and the rural coverage increases the surplus from 2.1 percent to 3.1 percent.

4. Conclusions and Future Work

NORC's AmeriSpeak web panel uses a full area-probability design to improve coverage. It uses in-person recruiting to improve the weighted response rate and also to reduce self-selection bias, but only as non-response follow-up to reduce costs. The added rural coverage does improve representativeness in Education, Low Income households, Internet and Telephone Usage, Marital Status and Home Ownership Status. The added rural coverage does not improve representativeness in Gender, Senior Citizens, High Income households, Household Size, Employment Status, and Pet Owners. For political affiliation, the rural coverage does move the liberal and conservative percentages towards

the benchmark, but not the moderate percentage. This could be because of a difference in the questions, though.

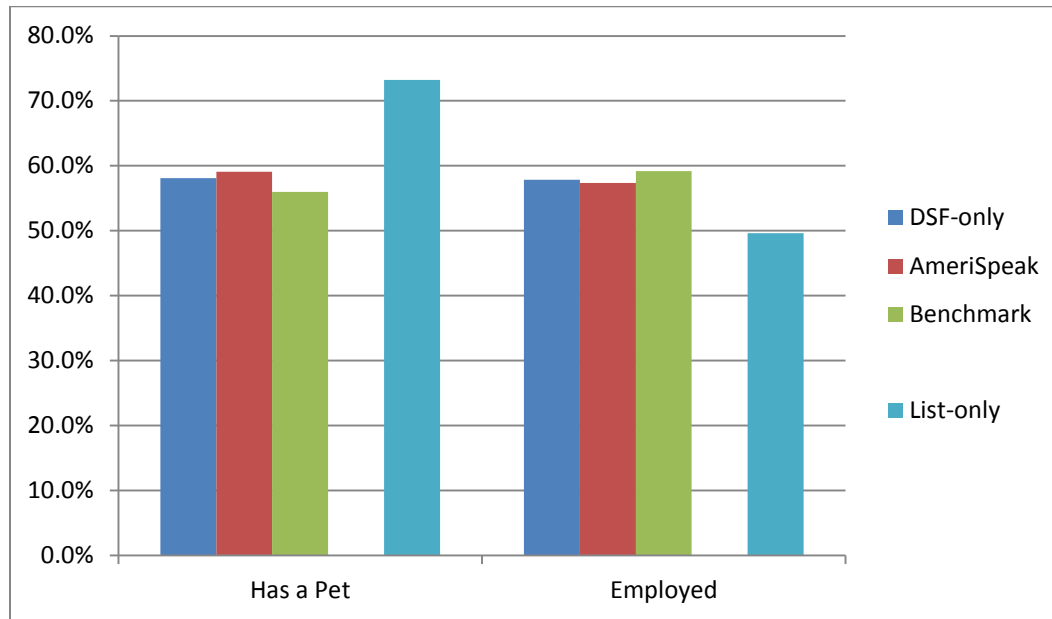


Figure 8: Having a Pet and Employment Status

Excluding race/ethnicity and young adults due to our oversampling of these groups, we examined a total of nineteen categories within twelve variables. For twelve (63 percent) of these categories, the AmeriSpeak estimate is closer to the national benchmark with the rural coverage than it would be without. While a sign test is not significant ($p = .18$), the gains in education, income, and telephone status do seem important.

The gains are certainly small, which is mostly because only six percent of our cases are within the rural segments and it could be argued that the rural coverage gains shown here are not sufficient to justify the cost. However, these costs are split among many projects at NORC, so AmeriSpeak is not responsible for the full cost, making these gains defensible.

References

- Barron, M., Kelly, J., Montgomery, R., Singleton, J., Shin, H., and Skalland, B., Tao, X., and Wolter, K. (2010). More on the Extent of Undercoverage in RDD Telephone Surveys Due to the Omission of 0-Banks. *Survey Practice*, Vol 3, Number 2. Accessed September 2016 at <http://www.surveypractice.org/index.php/SurveyPractice/article/viewFile/124/pdf>.
- Blumberg, S. and Luke, J. (2016). Wireless substitution: Early release of estimates from the National Health Interview Survey, January–June 2015. National Center for Health Statistics. December 2015. Accessed September 2016 at <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201512.pdf>

- Dennis, J.M., Ganesh, N., Fontes, A., Chew, K., and Lavrakas, P. (2015). Boosting Probability-Based Web Survey Response Rates via Nonresponse Follow-Up. AAPOR 2015 Presentation, Hollywood, FL. Accessed September 2016 at http://www.aapor.org/AAPOR_Main/media/AnnualMeetingProceedings/2015/C7-2-Fontes.pdf.
- Iannacchione, V., Morton, K., McMichael, J., Cunningham, D., Cajka, J., and Chromy, J. (2007). Comparing the Coverage of a Household Sampling Frame Based on Mailing Addresses to a Frame Based on Field Enumeration. Proceedings of the American Statistical Association Section on Survey Research Methods. Accessed September 2016 at <http://ww2.amstat.org/sections/srms/Proceedings>.