Transitioning a Random Digit Dialing Health Survey to Address-Based Sampling

Bonnie E. Shook-Sa[1], Lauren Klein Warren[1], David Roe[1],
Brian Head[1], Doug Currivan[1], and Barbara Bibb[1]
[1] RTI International, PO Box 12194, Research Triangle Park, NC 27709

Abstract
The selection of an appropriate sampling frame is dependent on multiple factors, including the target population, coverage of the target population, mode(s) of data collection, anticipated response rates, and impact on data collection costs. For household telephone surveys, two sampling frames are typically considered: dual-frame random digit dialing (RDD) and address-based sampling (ABS). RDD frames comprise samples of landline and wireless phone numbers mapped to the targeted geographic areas. While RDD samples allow for a consistent mode of recruitment, accurately linking wireless phone numbers to geographic areas is problematic. ABS frames are selected from commercially-available versions of the USPS Computerized Delivery Sequence (CDS) file. Geographic targeting is straightforward with an ABS design because the mailing address of each sampled unit is known. However, phone numbers cannot be linked to all sampled addresses which prevents an initial recruitment by telephone for a portion of the frame and poses data collection challenges.

Aligning Forces for Quality: Assessment of Consumer Engagement (AF4Q) is a survey of chronically ill consumers of healthcare residing in targeted geographic markets, ranging in size from single counties to entire states. Previous AF4Q studies were based on traditional and dual-frame RDD designs. Because of the error in linking wireless phone numbers to small geographic areas, the wireless phone sample led to both inefficiencies (phone numbers on the frame that were associated with persons not located within the targeted market) and undercoverage (persons in the targeted market whose phone numbers were not on the sampling frame). To mitigate these concerns, an ABS design was implemented in the current round of data collection, which consisted of sampling three new AF4Q markets. We discuss the challenges associated with each sampling method, and compare key sampling and data collection measures for the RDD and ABS designs.

Key Words: RDD, ABS, sampling efficiency, data collection costs, telephone surveys

1. Introduction

Aligning Forces for Quality: Assessment of Consumer Engagement (AF4Q) is a survey of chronically ill consumers of healthcare residing in geographically-defined markets that are participating in the AF4Q initiative. The goal of the AF4Q initiative is to lift the overall quality of health care in targeted communities, reduce racial and ethnic disparities, and provide models for national reform[1]. The Robert Wood Johnson

[1] More information about the AF4Q initiative can be found here: http://forces4quality.org/.

Foundation sponsors the AF4Q initiative, and the Center for Healthcare and Policy Research at Pennsylvania State University serves as the evaluation team. RTI International conducted the sampling, data collection, and weighting for the second round of the AF4Q.

The target population for the AF4Q consists of chronically ill consumers of healthcare aged 18 and over residing in the 19 AF4Q markets. To be considered chronically ill consumers of healthcare for the purposes of the survey, persons must self-identify as being diagnosed with one or more of the following conditions, and must have seen a healthcare provider regarding at least one of these conditions within the last two years: asthma, heart disease, diabetes, hypertension/high blood pressure, and depression.

To date, two full rounds of AF4Q data collection have been implemented. The AF4Q 1.1 and 1.2 were surveys of the original 15 and 3 additional AF4Q markets at baseline, respectively. Starting in 2011, new cross-sectional samples were selected in each market. The AF4Q 2.1 included the original 15 markets surveyed at baseline plus one new market and a national comparison sample; as with AF4Q 1.2, the AF4Q 2.2 included the 3 additional AF4Q markets. **Figure 1** lists the 19 AF4Q markets by survey implementation. AF4Q markets consist of single counties, collections of contiguous counties, and entire states.

AF4Q 1.1 and 2.1 Markets		AF4Q 1.2 and 2.2 Markets
Puget Sound, WA	South Central Pennsylvania	Albuquerque, NM
Detroit, MI	Greater Cincinnati, OH/KY	Indianapolis, IN
Memphis, TN	Cleveland, OH	Boston, MA
Minneapolis/St. Paul, MN	Kansas City, MO/KS	
Western New York	Minnesota	
Western Michigan	Willamette Valley, OR	
Wisconsin	Oregon, remainder (2.1 only)	
Maine		
Humboldt County, CA		

Figure 1: AF4Q Markets

The AF4Q interview is administered through two survey instruments. First, a randomly selected adult within each participating household is interviewed to determine eligibility for the study based on chronic conditions and care. The full interview is administered to screened and eligible respondents, and can be completed in either English or Spanish. Because of the complex nature of the questionnaire (length and skip patterns), the AF4Q has been restricted to telephone data collection. For this reason, the initial AF4Q data collection efforts were based on a random digit dialing (RDD) sampling frame. The AF4Q 1.1 was based on a traditional (landline only) frame, and the AF4Q 1.2 and 2.1 were based on dual-frame RDD designs. Because of methodological changes to the design between the 1.2 and 2.2 and changes in survey contractors, this paper will be restricted to comparisons between the AF4Q 2.1 dual-frame RDD design and the AF4Q 2.2 Address-Based Sampling (ABS) design. We will discuss each design, the decision to move from an RDD to an ABS design for the AF4Q 2.2, and note key comparisons between the two designs.

2. Sample Designs

2.1 Random Digit Dialing Design

As discussed above, in its first three implementations, the AF4Q was based on an RDD design. In 2011-2012, the AF4Q 2.1 was conducted, leading to approximately 18,000 completed screeners and 5,000 completed interviews across 16 AF4Q markets and the national comparison sample.

The landline component of the frame included geographic stratification within markets to achieve minority oversampling targets. In ten of the AF4Q markets, minority oversampling targets were designed to achieve a disproportionately higher percentage of interviews from minority respondents (Hispanics, non-Hispanic African Americans, and non-Hispanic Asians). Within the landline sample, high and low density minority strata were formed based on the mapped location of phone numbers to geographic areas (census block groups for listed numbers and counties for unlisted numbers). A sample optimization was conducted to allocate the sample across strata, minimizing design effects due to unequal weighting (see Kish, 1965, Valliant et al., 2013), or unequal weighting effects (UWEs), while expecting to achieve the minority oversampling goals.

The cell phone component of the frame was selected based on switch center locations. While rate centers have recently become available and have been shown to be more accurate than switch centers (Marketing Systems Group, 2014), they were not available at the time the sample was selected. The switch center locations associated with each market were targeted and included on the sampling frame. Approximately 25 percent of completed screeners and 20 percent of completed interviews were conducted by cell phone. The lower proportion of interviews conducted by cell phone is due to a lower chronic-illness eligibility rate from cell phone respondents in comparison to landline respondents.

While the RDD design achieved the goals of the study, several challenges were encountered:

- *Coverage Concerns*: The cell phone geographic ineligibility rate ranged from 5.5 percent to 48.5 percent across markets (Couzens et al., 2013), leading to concerns about how well the target population was covered within each market. It was assumed that this error was present in both directions (i.e., not only were we including elements on our sampling frame that were located outside of our targeted geographic areas, but we were also excluding elements from our frame that should have been included). Undercoverage results from both inaccuracies in the assignment of cell phones to markets as well as migration of the target population.
- *Sampling Efficiency Concerns*: During the AF4Q 2.1, it was challenging to achieve the minority oversampling goals with an efficient sample design because of the difficulty pinpointing telephone numbers to small geographic areas. Effective stratification was only possible for listed landline telephone numbers, which could be targeted to the census block group level. Unlisted landline numbers could only be targeted at the county-level, and cell phone numbers could only be targeted at the market level. This led to high UWEs, and shortfalls in minority targets.
- *Weighting Challenges*: Weighting of dual-frame RDD samples is dependent on accurate external benchmark estimates of phone use (i.e., estimates of cell phone only, dual users, and landline only adults in the population). However, the only sub-

state estimates available for these three domains were derived from models using National Health Interview Survey data (Blumberg et al., 2012). The available estimates were not as precise as national estimates of phone use, and no estimates were available for several of the AF4Q sub-state markets.

- *Operational Concerns*: Because of the low geographic eligibility rates encountered on the cell phone sample, the level of effort to reach the target population on the cell frame was quite high in comparison to the landline frame.

2.2 Address-Based Sampling Design

At the conclusion of the AF4Q 2.1, planning began for the AF4Q 2.2, which was conducted in the three additional AF4Q markets in 2013-2014. Due to the challenges encountered during the AF4Q 2.1 described in the previous section, the AF4Q team had an opportunity to make potential improvements by either implementing an RDD design using rate centers instead of switch centers in the cell phone component of the frame, or moving to an ABS design.

ABS frames are based on commercially-available versions of the United States Postal Service's (USPS) Computerized Delivery Sequence (CDS) file. The CDS file is made available to the public through licensing agreements with qualified private companies. The USPS also makes available the No-Stat file, which contains over 8 million primarily rural mailing addresses that supplement the CDS file with both active and vacant addresses that are excluded from the CDS file. The union of the CDS and No-Stat files account for all postal delivery points serviced by the USPS, giving ABS frames near-complete coverage of the household population (Iannacchione, 2011; Shook-Sa et al., 2013).

Unlike RDD frames where frame elements are phone numbers (some of which can be linked to addresses), ABS frame elements are mailing addresses (some of which can be linked to phone numbers). This poses both advantages and disadvantages for the AF4Q. The availability of mailing addresses for each frame element allows for more effective stratification and oversampling of minorities because frame elements can be targeted at small levels of geography. It also allows lead letters to be mailed to all sample cases, and simplifies weighting because sample elements are selected from a single frame (not a dual, overlapping frame as with RDD). However, the lack of phone numbers for all cases poses challenges for a telephone study like the AF4Q. With an ABS design, we would be dependent upon persons whom are residing at sampled addresses for which we cannot determine the phone number to provide us with their phone number or call in to complete the screening interview. This led to concerns both about the timeliness of data collection as well as response rates.

Because the potential advantages associated with ABS were thought to outweigh the challenges, the AF4Q 2.2 was implemented with an ABS design. The sampling frame, comprised of commercially available versions of the CDS and No-Stat files (see Shook-Sa et al., 2013), included geographic stratification for minority oversampling by grouping census block groups into strata based on the proportion of minorities. A two-stage sample of addresses was selected from the stratified ABS frame. The first-stage sample of addresses was selected to achieve the targeted percent minority in each market. Phone numbers were then appended to as many addresses in the first-stage sample as possible (approximately 45 percent). The sample was then further stratified based on phone append status, and addresses were subsampled for initial release. The sample was fielded in two waves to allow for changes in the sample allocation across minority and phone append strata, if needed.

Lead letters were mailed to all sampled addresses. Addresses for which a phone number was matched to the sample were routed to Computer Assisted Telephone Interviewing (CATI) data collection. Addresses for which a phone number could not be matched, and addresses that were matched but were confirmed in CATI to have non-working numbers or numbers associated with the wrong address, were routed to the unmatched portion of the frame. Unmatched addresses received up to three mailings, the first of which included the recruitment instrument, information about the study, and a $2 pre-incentive. The subsequent reminder mailings encouraged persons at sampled addresses to complete the recruitment instrument or call in for screening. Both matched and unmatched cases were offered a $20 promised incentive for completing the full interview, if found eligible. The AF4Q 2.2 design is summarized in **Figure 2**.

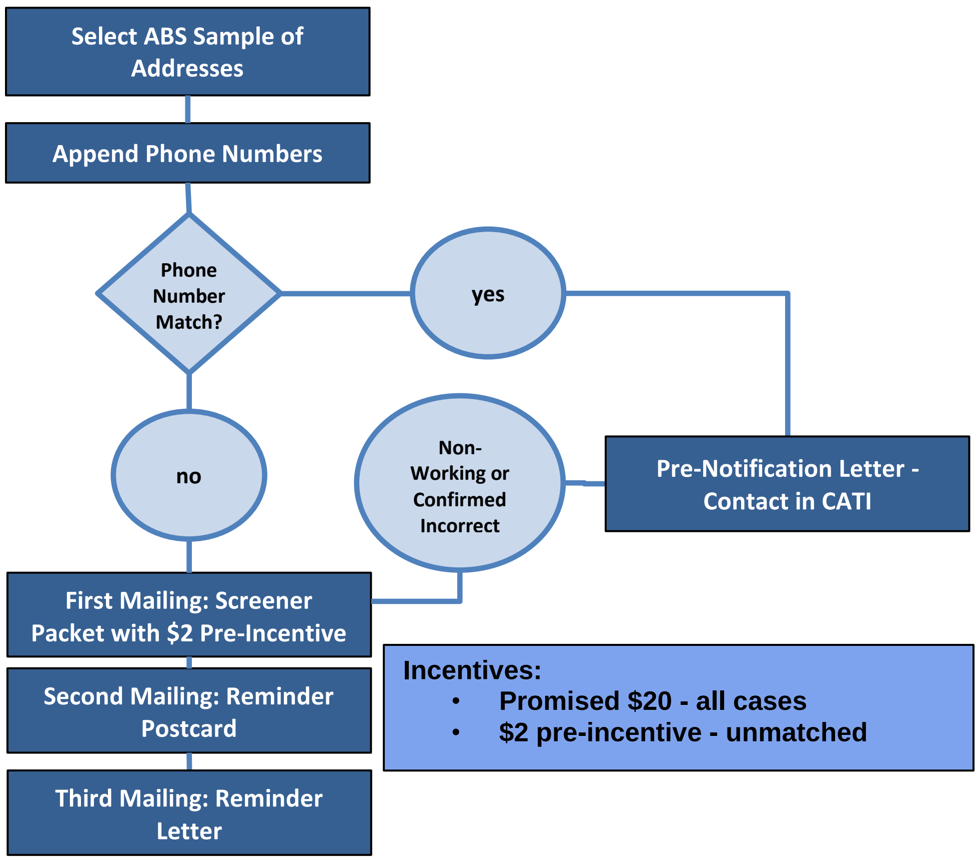

Figure 2: AF4Q 2.2 ABS Design

During the AF4Q 2.2 data collection, about 3,000 screeners and 1,000 interviews were completed. Of the completed interviews, approximately 30 percent were completed with respondents from unmatched addresses, and 70 percent were completed with respondents from matched addresses. As anticipated, the unmatched sample posed the most challenges during data collection. Response rates for the unmatched sample were lower than the matched sample (AAPOR4 response rates of 4.8 percent versus 13.4 percent, respectively). Because phone numbers that can be matched to the ABS frame are comprised primary of landline phone numbers, this led to the potential for nonresponse bias, as response to the survey was correlated with telephone use. It was unclear whether or not the weighting process would lead to the adequate representation of the cell phone population in the AF4Q 2.2 sample (see 2.3.5 for a further evaluation).

2.3 Design Comparisons

Following the AF4Q 2.2 data collection, comparisons were made between the dual-frame RDD (AF4Q 2.1) sample and the ABS (AF4Q 2.2) sample. Frame coverage, data collection costs, sampling efficiency, response rates, and benchmark estimates were computed for both samples. The primary limitation of these comparisons is that the AF4Q 2.1 and 2.2 samples were conducted in different geographic areas, and data collection occurred during different time frames. Despite these limitations, these comparisons between the samples provide insight regarding the advantages and limitations of each design for the AF4Q.

2.3.1 Frame Coverage

For the RDD sample, the landline frame was expected to have reasonable coverage of the target population with landline phone numbers. As previously discussed, the cell phone frame suffered from significant undercoverage of the target population due to inaccuracies in targeting relatively small geographic areas through switch centers. However, even if the sample design were transitioned to rate centers, sizeable undercoverage of the cell phone only population would remain. Previous studies have found state-level error rates of 8 and 10 percent, and county-level error rates of 33 and 40 percent with the use of rate centers for geographic targeting (Speizer et al., 2013; Pew Research, 2014). The literature show that ABS frames comprised of the CDS and No-Stat files provide near-complete coverage of the household population (Iannacchione, 2011).

2.3.2 Data Collection Costs

For both the RDD and ABS designs, we calculated the cost per completed interview. These costs included direct labor charges as well as other direct costs such as the costs of mailings and incentives. ABS completed interviews were approximately 20 percent cheaper than RDD completes, even taking into account costs of additional mailings and pre-incentives that were not incurred with the RDD design.

2.3.3 Sampling Efficiencies

We compared the sampling efficiency between the RDD and ABS samples by evaluating the variability in analysis weights, measured with UWE. With the RDD design, the main contributions to the UWE were differential sampling rates within the landline sample to achieve minority oversampling targets, differential nonresponse across sampling strata, and weight adjustments to account for the dual-frame design. With the exception of adjustments for the dual frame design, these were also the contributors to the UWE for the ABS design. Minority oversampling with the ABS design was expected to be more statistically efficient (i.e. incur less variation in design weights) than the RDD design because all frame members could be allocated into minority strata at a fine level of geography, which was not feasible with RDD. However, the ABS sample suffered from differential nonresponse across phone-append strata, with matched cases responding at higher levels than unmatched cases. This differential nonresponse contributed to increases in the UWE during the nonresponse adjustment.

Because the UWEs within AF4Q markets are highly dependent on the distribution of minorities within the market (i.e., minority oversampling goals can be achieved with a lower UWE in markets where minorities exhibit more geographic clustering than markets where minorities are more geographically dispersed), comparing UWEs across the RDD and ABS markets does not demonstrate which sample design is more statistically efficient. To compare the designs on an even playing field, Warren et al. (2014) conducted a simulation study in which the ABS design was simulated for the 10 AF4Q 2.1 markets that incorporated minority oversampling and were originally conducted via RDD. This simulation allocated the ABS sample within the 10 RDD markets, holding constant the number of interviews and number of minority interviews, assuming sample yield rates would be similar to those observed in the AF4Q 2.2 ABS markets. After allocating the ABS sample within the RDD markets, the UWEs from design weights were increased based on the observed inflation from design to final analysis weights in the three AF4Q 2.2 ABS markets. Effective sample sizes were calculated for both the original (RDD) and simulated (ABS) designs by taking the ratio of the total number of respondents to the UWE. The effective sample size for each design, differences in effective sample sizes, and relative differences are presented in **Table 1**. For all markets except two (Memphis and Cleveland), the ABS simulated effective sample sizes were higher than the RDD observed effective sample sizes. This simulation study concluded that the ABS design was more statistically efficient than the RDD design.

Table 1: Effective Sample Sizes: RDD vs. Simulated ABS

Market	RDD ESS	Simulated ABS ESS	ABS-RDD	% Relative Difference
Minneapolis/St. Paul, MN	86	132	47	54.7
Willamette Valley, OR	122	176	54	44.3
Western New York	79	110	31	39.2
Puget Sound, WA	142	191	48	33.8
Greater Cincinnati, OH/KY	104	128	25	24.0
Kansas City, MO/KS	115	132	17	14.8
Detroit, MI	101	110	9	8.9
National Comparison	183	194	11	6.0
Memphis, TN	106	97	-9	-8.5
Cleveland, OH	118	97	-21	-17.8

ESS = Effective Sample Size
Source: Warren et al., 2014

2.3.4 Response Rates

The literature shows that low response rates can be indicative of nonresponse bias, but that the link between high levels of nonresponse and nonresponse bias are not always present (Groves, 2006). Therefore, response rates can be compared between the two designs, but they are not necessarily indicative of which design produced higher quality estimates.

Table 2 contains the ranked weighted AAPOR4 response rates in the AF4Q 2.1 (RDD) and AF4Q 2.2 (ABS) markets. While the response rates in all market were relatively low (max response rate is 31 percent), the response rates in the RDD markets were higher than the response rates in the ABS markets. With the exception of the national comparison sample, all RDD markets had higher response rates than the ABS markets.

As with the other comparisons, the caveat to the response rate comparison is that the surveys took place during different time frames, and in different geographies.

Table 2: Ranked Weighted Response Rates by Market: AF4Q 2.1 (RDD) and AF4Q 2.2 (ABS)

Market	AAPOR4 Weighted Response Rates (%)
Oregon, remainder	31.0
Western Michigan	29.0
Detroit, MI	27.6
Minnesota	25.9
Cleveland, OH	24.9
Minneapolis/St. Paul, MN	24.3
Wisconsin	22.1
Greater Cincinnati, OH/KY	20.6
Willamette Valley, OR	19.6
Western New York	19.4
Memphis, TN	19.1
Maine	18.8
Humboldt County, CA	17.9
Puget Sound, WA	17.8
South Central Pennsylvania	16.7
Kansas City, MO/KS	16.5
Indianapolis, IN	11.2
Albuquerque, NM	9.9
National Comparison	9.7
Boston, MA	6.4

Blue = AF4Q 2.1 (RDD) markets; Green=AF4Q 2.2 (ABS) markets

2.3.5 Benchmarking

One way to evaluate data quality for a survey is to compare key estimates with gold standard estimates from an external data source (Groves 2006). This is often referred to as benchmarking. We compared key estimates from the AF4Q 2.1 (RDD) and AF4Q 2.2 (ABS) surveys with estimates from the National Health Interview Survey (NHIS) and the National Survey on Drug Use and Health (NSDUH). The NHIS and NSDUH are large, national, in-person surveys that have been conducted since 1957 and 1971, respectively.

The NHIS is sponsored by the National Center for Health Statistics, and data are collected by the US Census Bureau. The NSDUH is sponsored by the Substance Abuse and Mental Health Services Administration, and data are collected by RTI International. Because of the size of these surveys and their rigorous data collection protocols, the NHIS and NSDUH can serve as gold standards for evaluating estimates from the AF4Q surveys.

The AF4Q 2.1, which was conducted via RDD, contained a national comparison sample, which consisted of the complement of the 16 AF4Q markets. Therefore, by combining market-level data with the national comparison sample, national estimates can be computed for the AF4Q 2.1 and compared with national estimates from the NHIS and NSDUH. We compared weighted estimates obtained in the AF4Q 2.1 screening and full interviews to external estimates from the 2011 NHIS. Six estimates from the AF4Q screener were compared with NHIS estimates, and five estimates were also compared to national estimates from the 2010-2011 NSDUH (United States Department of Health and Human Services, 2014). Estimate comparisons are presented in **Table 3**. Most of the estimates track well with the NHIS and NSDUH. There are slight differences in question wordings that could account for some of the small differences in estimates (e.g., the NSDUH referred to hypertension as “high blood pressure”, while the AF4Q used both “hypertension” and “high blood pressure” to describe this chronic condition). The only screener outcome examined with very different responses is health status rated as excellent or very good for the population 18-64. However, this is not necessarily indicative of nonresponse bias. Response options were presented in opposite orders (excellent to poor on the NHIS and NSDUH, poor to excellent on the AF4Q), and mode differences could also confound this comparison.

In addition to screener items, AF4Q interview items were also compared to the NHIS. Estimates for the NHIS were limited to persons reporting that they had ever been told that they had diabetes, hypertension, heart disease, or asthma. Question wordings varied from the AF4Q, and there was no way to limit the NHIS to persons who have seen a health care professional for treatment of their conditions in the last two years to make estimates directly comparable to the AF4Q. AF4Q estimates excluded persons only reporting depression, as there was no depression measure for the NHIS. We first compared the weighted populations for each survey. The restricted NHIS population represents an estimated 98 million persons, while the restricted AF4Q population represents an estimated 69 million persons. The NHIS estimates represent more persons because the NHIS does not have the chronic care restriction. While the estimate comparisons below do not represent exactly the same target population, the populations are as close as possible given the limitations of both surveys. Interview estimates comparisons are also presented in **Table 3**. Most estimates track reasonably well across surveys, with the exceptions of past year flu shot and ER visits in the past 12 months. These differences could potentially be explained by the differences in the representative populations between the surveys (care for chronic condition in the past two years), but there is no way to determine whether or not this contributes to the difference.

While the AF4Q 2.1 estimates tracked reasonably well with the NHIS and NSDUH estimates, there is one limitation to this analysis. Because estimates are weighted, the AF4Q estimates are primarily driven by the national comparison sample, where coverage of the cell phone only population is not as much of a problem as in markets consisting of small geographic areas. For this reason, national estimates might track well with the NHIS while market-level estimates could still exhibit bias.

Table 3: Comparison of AF4Q 2.1 (RDD) Estimates to 2011 NHIS and 2010-2011 NSDUH Estimates

Estimate	AF4Q Estimate (SE)	NHIS Estimate (SE)	NSDUH Estimate (SE)	AF4Q - NHIS	AF4Q - NSDUH
Screener Items					
Self-Reported Health Status as excellent or very good (%)					
18-64	47.1 (1.8)	63.9 (0.4)	63.0 (0.2)	-16.8	-15.9
65+	40.4 (3.3)	41.6 (0.7)	42.6 (0.5)	-1.2	-2.2
Diabetes (%)	11.3 (0.9)	8.9 (0.2)	8.3 (0.1)	2.4	3.0
Hypertension (%)	27.3 (1.3)	29.7 (0.4)	23.4 (0.2)	-2.4	3.9
Heart Disease (%)	5.6 (0.6)	7.4 (0.2)	5.9 (0.1)	-1.8	-0.3
Asthma (%)	10.7 (0.9)	12.6 (0.2)	11.0 (0.1)	-1.9	-0.3
Any Chronic (of the 4) (%)	38.9 (1.5)	42.3 (0.4)	n/a	-3.4	n/a
Interview Items (Limited to persons with diabetes, hypertension, heart disease, or asthma)					
Flu Shot (last year) (%)	58.6 (3.3)	48.0 (0.5)	n/a	10.5	n/a
Flu Spray (last year) (%)	3.1 (1.7)	0.7 (0.1)	n/a	2.5	n/a
Never Smoked (%)	50.5 (3.4)	52.3 (0.5)	n/a	-1.8	n/a
ER Last 12 Months (%)	34.5 (3.2)	26.5 (0.4)	n/a	8.0	n/a
Mean Body Weight (Pounds)	185.2 (3.5)	182.9 (0.4)	n/a	2.4	n/a
Mean Height (Inches)	66.8 (0.3)	66.8 (0.0)	n/a	0.1	n/a

NOTE: Standard errors are in parenthesis

The AF4Q 2.2, which was conducted with an ABS sampling frame, did not include a national comparison sample. Therefore, estimates should not be compared with the national NHIS or NSDUH estimates. However, NHIS produces local estimates for a key AF4Q data quality measure: phone use. As mentioned previously, there were concerns that the AF4Q 2.2 did not adequately represent the cell phone only population due to differential nonresponse between the phone append strata (i.e. unmatched cases responded at much lower rates than matched cases). Therefore, one key estimate to examine is the weighted percent cell phone only in the sample versus the 2012 NHIS (Blumberg, 2013). **Table 4** compares the weighted cell phone only populations in the AF4Q 2.2 (ABS) markets versus the 2012 NHIS. The NHIS only provides sub-state estimates in some areas, and there was not a comparable estimate for the Albuquerque market. The AF4Q 2.2 estimates track quite well with the NHIS estimate in Boston and Indianapolis. The caveat for this analysis is that only one measure could be compared with the NHIS, and this measure is not necessarily indicative of bias in additional estimates.

Table 4: Comparison of AF4Q 2.2 (ABS) Percent Cell Phone Only to 2012 NHIS

Market	AF4Q 2.2	NHIS
Boston	34.4 (2.7)	37.5 (3.6)
Indianapolis	38.3 (2.9)	44.9 (3.3)

NOTE: Standard errors are in parenthesis

As a final comparison between RDD and ABS designs, we obtained publically-available NSDUH 2002-2011 sub-state estimates in the six RDD and the three ABS markets where comparable geographic regions could be defined (United States Department of Health and Human services, 2013). One screener-level outcome could be compared between the two surveys: self-reported health status. From the national comparisons with the RDD sample, we knew that this estimate did not track well with the AF4Q estimated self-reported health status. However, we compared the differences between the AF4Q and NSDUH sub-state estimates at the market level and calculated the mean difference for AF4Q 2.1 (RDD) and AF4Q 2.2 (ABS) markets to see if RDD or ABS markets tended to track closer to the NSDUH estimates. The market-level differences for the percent of persons 18-64 rating their health status as excellent or very good in the AF4Q and the NSDUH are presented in **Table 5**. Mean differences for RDD and ABS markets were -16.6 and -18.1 percent, respectively. There is no evidence that either RDD or ABS estimates aligned more closely with NSDUH estimates for this outcome.

Table 5: Comparison of AF4Q 2.1 and 2.2 Self-Reported Health Status[1] to the 2002-2011 Substate NSDUH Estimates (Persons 18-64)

Market	AF4Q – NSDUH (%)
AF4Q 2.1 (RDD) Markets	
Detroit, MI	-20.7
Memphis. TN	-12.2
Wisconsin	-11.7
Maine	-11.7
Cleveland, OH	-23.3
Kansas City, MO/KS	-19.7
Mean Difference (RDD)	**-16.6**
AF4Q 2.2 (ABS) Markets	
Indianapolis	-20.5
Albuquerque	-12.3
Boston	-21.4
Mean Difference (ABS)	**-18.1**

[1] Percent with self-reported health status of excellent or very good

3. Conclusions

Both the RDD and ABS designs achieved the goals of the AF4Q study. However, key differences were found in terms of frame coverage, data collection costs, sampling efficiency, and response rates. The ABS design offered higher frame coverage, lower costs, and improved sampling efficiency, but the RDD design had higher response rates. The majority of estimates compared between the two designs and the NHIS tracked reasonably well. However, there were limitations of the assessment of data quality for both designs. RDD national estimates were driven by the national comparison sample, so the quality of market-level estimates could only be assessed with a single measure from the NSDUH. Limited estimates were available for comparisons with the ABS markets due to the lack of a national comparison sample. There is not a gold standard method for assessing if the undercoverage of the RDD frame or the low response rates of both designs contributed to bias. While we will continue to evaluate data quality measures, we recommend the ABS design for future implementations of the AF4Q because of the increased coverage, better sampling efficiency, and cost savings provided by this design.

Acknowledgements

The authors would like to acknowledge the contributions of Dennis Scanlon and David Johnson with the Center for Healthcare and Policy Research at Pennsylvania State University. Thanks also to Jill Dever and Marcus Berzofsky with RTI International for their valuable comments.

References

Blumberg, Stephen, Nadarajasundaram Ganesh, Julian Luke, Gilbert Gonzales. 2013. “Wireless Substitution: State-level Estimates From the National Health Interview Survey, 2012.” National Health Statistics Reports: Number 70.

Blumberg, Stephen, Julian Luke, Nadarajasundaram Ganesh, Michael Davern, and Michel H. Boudreaux. 2012. “Wireless Substitution: State-level Estimates From the National Health Interview Survey, 2010-2011.” National Health Statistics Reports: Number 61.

Couzens, Amy, Bonnie Shook-Sa, David Roe, and Derek Stone. 2013. “Evaluation of a targeted dual-frame RDD sample of sub-state populations”. Poster presented at American Association for Public Opinion Research, Boston, MA.

Groves, Robert. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” Public Opinion Quarterly 70:646-675.

Iannacchione, Vincent. 2011. “The Changing Role of Address-Based Sampling in Survey Research.” Public Opinion Quarterly 75:556-575.

Kish, Leslie. 1965. Survey Sampling. New York: John Wiley and Sons.

Marketing Systems Group. "Cellular Sampling – a New Approach – Rate Centers." Available online as of August 20, 2014 at http://www.m-s-g.com/CMS/ServerGallery/MSGWebNew/Documents/GENESYS/whitepapers/CellSampling-RateCenters.pdf

Pew Research. "Cell Phones." Available online as of August 20, 2014 at http://www.people-press.org/methodology/sampling/cell-phones/.

Shook-Sa, Bonnie, Douglas Currivan, Joseph McMichael, and Vincent Iannacchione. 2013. "Extending the coverage of address-based sampling frames: Beyond the USPS computerized delivery sequence file." *Public Opinion Quarterly* 77: 994–1005.

Speizer, Howard. Marcus Berzofsky. Thomas Duffy, Jamie Ridenhour, and Tim Sahr. 2013. "Alternative sample selection and data collection strategies for balancing cell phone response distribution across county/region level geographies in a dual frame (landline/cell) telephone survey". *Presented at AAPOR 68th Annual Conference*, Boston, MA.

United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality. National Survey on Drug Use and Health: 2-Year R-DAS (2002 to 2003, 2004 to 2005, 2006 to 2007, 2008 to 2009, and 2010 to 2011). ICPSR34482-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-06-25. http://doi.org/10.3886/ICPSR34482.v2

United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality. National Survey on Drug Use and Health: 10-Year Substate R-DAS (2002 to 2011). ICPSR34947-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research[distributor], 2013-12-17. http://doi.org/10.3886/ICPSR34947.v1

Valliant, Richard, Jill Dever, and Frauke Kreuter. 2013. "Practical Tools for Designing and Weighting Survey Samples." New York: Springer.

Warren, Lauren Klein, Bonnie Shook-Sa, David Roe, and Brian Head. 2014. "Oversampling Minorities under Random Digit Dialing Versus Address Based Sampling". *Presented at American Association for Public Opinion Research*, Anaheim, CA.