Combining Probability and Nonprobability Samples to form Efficient Hybrid Estimates: An Evaluation of the Common Support Assumption

Jill A. Dever
RTI International, Washington, DC

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract
Nonprobability surveys, those without a defined random sampling scheme, are becoming more prevalent. These studies can offer faster results at less cost than many probability surveys, especially for targeting important subpopulations. This can be an attractive option given the continual challenge of doing more with less, as survey costs continue to rise and response rates to plummet. Nonprobability surveys alone, however, may not fit the needs (purpose) of Federal statistical agencies where population inference is critical.

Nonprobability samples may best serve to enhance the representativeness of certain domains within probability samples. For example, if locating and interviewing a required number of subpopulation members is resource prohibitive, data from a targeted nonprobability survey may lower coverage bias exhibited in a probability survey. In this situation, the question is how to best combine information from both sources.

Our research searches for an answer to this question through an evaluation of hybrid estimation methods currently in use that combine probability and nonprobability data. Methods that employ generalized analysis weights (i.e., one set of weights for all analyses) are the focus because they enable other survey researchers and policy makers to analyze the data. The goal is to identify procedures that maximize the strength of each data source to produce hybrid estimates with the low mean square error.

The details presented here focus on the propensity score adjusted (PSA) nonprobability weights needed prior to combining the data sources, and the common support assumption critical to hybrid estimation and PSA weighting. Empirical results suggest that the support common to the probability and nonprobability data should inform the PSA weights, resulting in lower bias within the nonprobability estimates. Though matching techniques are used, additional research is needed to identify robust methods to evaluate the common support assumption.

Introduction
Probability sampling has been "the rule" with surveys intended for population inference of totals, means and more complex estimators. Published by Neyman (1934), random sampling theory is used to extrapolate information obtained for a sample to the target population through analysis weights. For example, the Horvitz-Thompson estimator (1951) of a population total, $t_{Uy} = \sum_U y_i$, for some variable y is

$$\hat{t}_y = \sum_s d_k y_k \qquad (1)$$

where U is the universe or target population, s is the sample set; $d_k = \pi_k^{-1}$, the inverse selection probability for the k^{th} sample member calculated with respect to the sampling design; and y_k is the corresponding value of y for the sample member. Sampling theory was developed with the assumption that data are collected in its entirety—no nonresponse, no coverage error, etc. However, surveys today are fraught with several interesting challenges.

Many surveys with a probability-based sample design have been criticized as being economically inefficient. Nonresponse has been discussed for many years as a growing challenge. For example, Keeter et al. (2017) note that response rates for telephone polls in the U.S. have decreased steadily in the period from 1997 to 2012, leveling off at around 9 percent through 2016. Moreover, the relationship between pre-interview or promised incentives and increased response rates is mixed, suggesting that monetary incentives could exacerbate study costs without benefiting the analyses (Mercer et al. 2015). Response rates are further affected if persons are members of hard-to-survey (hard-to-reach) populations (Tourangeau 2014), either because such identifying information is not available on the list frame for targeted sampling or the inability to recruit them into the study. For example, address-based sampling frames are stated as having nearly complete coverage of the U.S. household population, but the viability of the auxiliary information appended to these lists is also mixed (Iannacchione 2011; Harter et al. 2016). Though low response is not necessarily indicative of bias (Groves & Peytcheva 2008), a small number of interviews overall or for important subgroups will lower precision for the estimates and power for statistical tests. Finally, if list frames are not readily available, extensive amount of time and funds may be needed to construct the lists such as with an in-person household survey using area probability sampling (Valliant et al. 2018).

Surveys conducted in the absence of a reproducible, probability-based sample design (i.e., nonprobability sampling) are not new. A few more recent examples of nonprobability survey data collection include surveillance of HIV and hepatitis C infected groups (Solomon et al. 2017), estimates of injection drug users (Wu et al. 2017), estimated price index for 22 countries using web-scraped data (Cavallo & Rigobon 2016), e-cigarette usage (Kim et al. 2015), opinions of political issues (Clement 2016; Conway et al.2015; Dropp & Nyhan 2016), social stability indicators (Kleinman 2014), and public health events (Harris et al. 2014). Prevalence of nonprobability surveys in recent years is notable as they tend to produce data requiring fewer resources (Baker et al. 2013), and proliferation of big data sources and software to analyze generous amounts of data (Japec et al. 2015), to name just a few reasons.

Research to date has focused primarily on the use of nonprobability sampling as the stand-alone source of survey responses for population inference. The utility of the methods to generate the univariate population estimates, such as propensity score adjustments (PSAs) detailed in the next section, are mixed. For example, both Mercer et al. (2018) and Valliant (2018) stress the use of effective auxiliary information in creating nonprobability analysis weights; however, the weighting methods (e.g., raking, PSA) were unable to remove all biases from the population estimates. Dutwin & Buskirk (2017) demonstrated that some techniques could be effective (e.g., sample matching) but they did not attain consistency across univariate statistics. Studies comparing the estimated relationship between variables from probability and nonprobability surveys has shown "greater" but not complete consistency than the univariate statistics (Pasek 2015).

Noting the challenges stated for probability surveys, nonprobability samples may better serve to enhance the representativeness of certain domains within probability samples such as those hard-to-survey populations instead of providing data for population inference alone. Researchers have proposed several methods for combining the two sources to form what we call *hybrid estimates*. The research presented in the paper focuses on a preliminary step for hybrid estimation associated with the derived nonprobability weights. Namely, we evaluate one of the critical assumptions of PSA weights known as common support. Techniques to date for constructing hybrid estimates are briefly summarized in the *Background* section along with details of the common support assumption. The *Methods* section contains our approach for evaluating the common support assumption from a survey of adults' opinions of marijuana use in the U.S. (*National Estimates of Marijuana Use in the U.S. Adult Population* section). *Preliminary* results are presented from the analyses. We provide conclusions of the results and next steps in hybrid estimation research in the last section of the paper.

Background

Hybrid estimation is the label we use to describe statistics generated from a combined data file containing probability-based and nonprobability sample cases. These data are combined in such a way as to maximize the information from each source. This is similar to the goal in dual-frame estimation where probability samples from different but possibly overlapping sampling frames are combined to generate a single set of population estimates (Lohr and Raghunathan 2017; Mecatti 2007; Lohr & Rao 2006). There are several differences, however, between hybrid and dual-frame estimation.

Figure 1 provides a visual description of a general hybrid estimation protocol. Data on a probability sample are collected, resulting in final analysis weights adjusted to reduce bias. Adjustments applied to the inverse selection probability weights that reflect the probability sample design may include those to address nonresponse bias and coverage such as a weighting class adjustment or calibration to known population totals (see, e.g., Valliant et al. 2018; Kott 2006). At this point, population inference can be made from this sole source with the fully-adjusted, probability-based analysis weights.

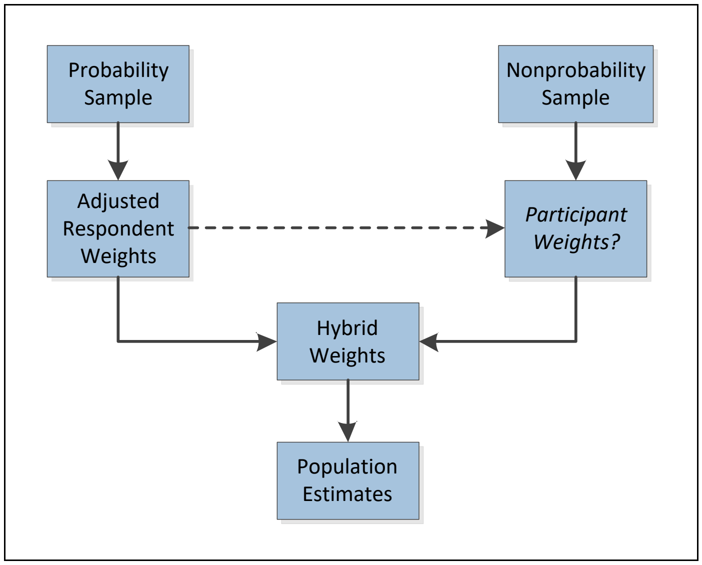

Figure 1. Generic hybrid estimation system

The right-hand side of Figure 1 focuses on the data from the nonprobability sample. A summary of several nonprobability designs that could generate these data are found in Baker et al. (2013). Most nonprobability designs do not automatically produce a selection weight; an exception may be sampling from an opt-in (nonprobability) panel where panel weights were already created (Callegaro et al. 2014). Researchers desiring population estimates from the nonprobability sample alone or in combination with the probability sample face questions such as: "are nonprobability analysis weights needed?" and "if weights are needed, how should they be constructed?" Let us forgo answers to these questions for now and assume that the nonprobability sample alone, like the probability sample, could be used to produce population estimates.

Prior to generating hybrid estimates, the input weights are further adjusted so that the when combined the data project to the intended target population for the study. The step is also used in dual frame estimation. Continuing with equation (1), a two-frame composite estimate of total y is

$$\hat{t}_y = \sum_{s_A} w_{Ak} y_{Ak} + \sum_{s_{A\cap B}} \lambda_k w_{Ak} y_{Ak} + \sum_{s_{B\cap A}} (1-\lambda_k) w_{Bk} y_{Bk} + \sum_{s_B} w_{Bk} y_{Bk} \qquad (2)$$

where s_A and s_B are the samples unique to frames A and B and to the corresponding populations, $s_{A\cap B}$ is the sample from frame A that is also covered by frame B, $s_{B\cap A}$ is the sample from frame B that is also covered by frame A, w_{Ak} and w_{Bk} are the analysis weights for the respective samples, and y_{Ak} and y_{Bk} are the y-values obtained from the A and B samples. The term $\lambda_k (\leq 1)$ is referred to as a composite factor that addresses any overlap of the populations and may account for precision and bias from each sample (Lohr & Raghunathan 2017; Brick et al. 2011; Merkouris 2010). Dual-frame estimation may also be achieved through generalized regression estimation where sample estimates are calibrated to population totals or population estimates (Merkouris 2010); we leave this approach to forthcoming publications.

Inherent in the use of dual-frame estimation is the assumption that the estimates from each survey are (relatively) unbiased measures for the same target population. For dual-frame estimation with data from two probability-based surveys, this means $E\left(\sum_{s_A} + \sum_{s_{A\cap}}\right) \approx t$ and $E\left(\sum_{s_{B\cap A}} w\ \ y + \sum_{s_B} w\ \ y\right) \approx t$, where E denotes the theoretic expectation of the expression within the parentheses (see, e.g., Särndal et al. 2003). Thus, effective auxiliary information (e.g., population control totals) and weight adjustments are needed to remove biases so that nonparticipants are considered missing at random (MAR) and the expectation assumption holds. Data used for estimation from the multiple samples are collected with the same instrument and ideally with the same mode of data collection. Finally, the composite factor (or "glue" in the combined data file) should minimize the measure square error (=bias^2 + variance) for at least a key set of estimates, because the single set of analysis weights is never expected to be ideal for all analyses for a survey (Valliant et al. 2018).

A similar approach and the same set of assumptions as discussed for probability-based, dual-frame estimation is needed for hybrid estimation. Nonprobability survey data alone, however, may not produce reasonable population estimates with w_k=1. Without controlling how participants join the study, many argue that some type of adjustment is needed to control selection bias (Lee & Valliant 2009; Valliant & Dever 2011). Additionally, the nonprobability sample, even after some statistical adjustments, may not cover the entire target population (Valliant 2018).

Given such challenges with the nonprobability data, we return to the questions above. For convenience and as an orientation to the research presented here, we respond to the questions with the answer: "yes participants weights will be produced for the nonprobability data via propensity scores estimated with the probability sample." Note that PSAs are not the only option to calculate nonprobability participant weights, hence the dotted line in Figure 1. Other examples include sample matching where probability weights are borrowed for the nonprobability records in the matched set (Dutwin & Buskirk 2017; Elliott & Valliant 2017) and weight calibration (Kott 2017; Lee & Valliant 2009; Dever & Brown 2016). Instead, we concentrate on PSA for the research presented here and withhold evaluation of the participant weight methodology for a subsequent evaluation.

Binary regression is used to estimate the probability of being in the nonprobability sample with information from a reference survey. In the context of hybrid estimation, the reference survey is generally a probability-based survey of the target population under study. The binary dependent variable is coded "1" for the nonprobability records and "0" for the probability records. Input weights for the nonprobability records are set to 1, while the fully-adjusted analysis weights are used for the probability sample. The inverse of the resulting propensities for the nonprobability sample then forms the participant weights. Note that the direct use of the propensities for the weights is preferred over their use in forming weighting classes for subsequent adjustments (Valliant & Dever 2011). The resulting nonprobability weights then may be calibrated to population totals; if the reference survey is also the source of the calibration totals, then the adjustment is applied in one step. Additional information on propensity weighting can be found in, for example, Valliant, Dever, & Kreuter (2018) and Valliant & Dever (2018).

Several assumptions are inherent in the use of PSAs for nonprobability estimation (Valliant & Dever 2011). Like dual-frame estimation, the binary regression model covariates (and, most likely, in a subsequent calibration adjustment) should be collected using identical questionnaire items. Nonparticipants in both surveys should be MAR. The probability sample should produce unbiased estimates and have sufficient size to estimate the propensity scores. Study participants should be included in only one sample, i.e., no overlap in the sample files. Finally, both samples—or portions of the samples—should cover the same portion of the population, referred to as common support (Valliant & Dever 2018; Stuart et al. 2011; Stuart 2010). **Figure 2a** demonstrates when two samples have common support in that their estimated distributions completely overlap. In this instance, the population total in expression (2) is recast as

$$\hat{t}_y = \sum_{s_{A\cap B}} \lambda_k w_{Ak} y_{Ak} + \sum_{s_{B\cap A}} (1-\lambda_k) w_{Bk} y_{Bk}$$

because there are no units uniquely associated with a particular source. In other words, s_A and s_B are empty sets and corresponding components in (2) are zero.

Conversely, **Figure 2b** shows when the common support assumption fails. For example, if the probability sample includes respondents who do not have access to the internet, the source for the opt-in (nonprobability) convenience sample, and their responses are unique from those with access, then the probability distribution will cover a portion of the population not accessible by the nonprobability design. This is represented in Figure 2b as the left-skewed distribution of the probability sample.

The nonprobability survey may capture sample members not included in the probability sample. For example, if the opt-in convenience sample includes fraudulent participants (such as those who participate multiple times for additional incentive money with or without "bot" assistance; see, for example, Teitcher et al. 2015), then the convenience sample distribution would contain ineligibles. The nonprobability sample may also include hard-to-survey population members such as those with a relatively rare characteristic (e.g., young adult current users of marijuana). Either condition is represented in Figure 2b by the right-skewed distribution of the convenience (nonprobability) sample. Assuming the ineligibles are identified post-data collection, then the estimated population total would be estimated with expression (2).

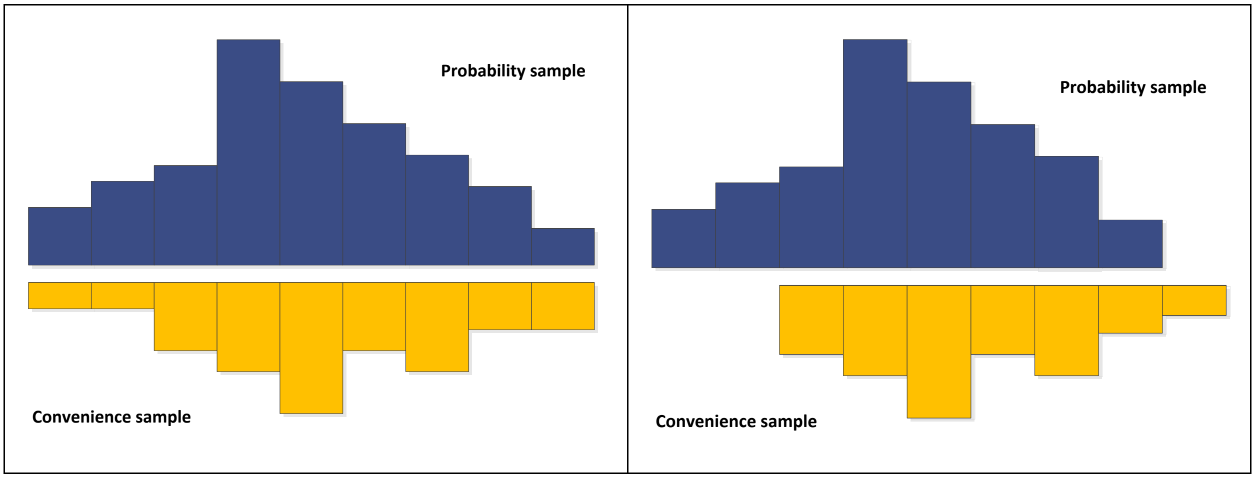

Figure 2a. Common support **Figure 2b. Common support violations**

If the “Figure 2b” condition holds, then the estimated propensities will not effectively represent the likelihood of being “selected” into the nonprobability sample if the “Figure 2a” condition is assumed. Thus, an evaluation of the common support assumption is a necessary first step for hybrid estimation. The question remains for how best to estimate the common support.

To date several options have been used to compare probability and nonprobability samples. For example, demographic and geographic characteristics (age, race/ethnicity, region) for the two samples are contrasted to determine if defined groups are missing from either source (Dutwin & Buskirk 2017; Yeager et al. 2011). Quota sampling of nonprobability survey participants is one remedy for such misalignments (Hayes et al. 2015; Brick 2011). However, differences in these characteristics may not translate into differences for the estimates desired from the surveys. Research on the effects of survey nonresponse, for example, suggests that descriptive auxiliary information alone may not provide a sufficient adjustment to the weights to remove bias in the survey estimates (Peytchev et al. 2018; Mercer et al. 2017; Krueger & West 2014; Lee & Valliant 2009).

Another example is sample matching (Baker et al. 2013; Stuart 2010; Rivers & Bailey 2009). Here models are used to identify similar records across the two sources. Similarity is defined by some distance measure between the probability and nonprobability propensities. Matches can be defined as a unique 1-to-1, 1-to-many, or many-to-many, and the matches need not include all records from each survey but typically do. Unlike the geo-/demographic comparisons, a model-based approach afforded by techniques such as sample matching can incorporate a variety of covariates and interaction terms, much like the PSA models themselves.

As with our choice of PSA for calculation of the participant weights, we focus on sample matching as the method used to evaluate the common support assumption and the effects of ignoring the results from the evaluation. Further consideration of the mechanics to best quantify common support is saved for another day. Analyses are conducted on data collected from two surveys—one with a defined probability sample design and one without—using the same questionnaire to assess opinions of marijuana use in the United States as described in the next section.

National Estimates of Marijuana Use in the U.S. Adult Population

National Marijuana Study. RTI International (RTI) funded a survey in the U.S. called the *2016 National Marijuana Beliefs and Behaviors Study* (NMS). The purpose of the NMS was to assess adults’ thoughts on the potential health risks and medicinal benefits of marijuana. Evaluations were made in context of policies on medical and recreational marijuana in place for their reported state of residence at the time of the survey.
`

The NMS is best described as a survey with a dual-frame, stratified design. One sample was selected from an address-based sampling frame (ABS) where one adult resident per household was interviewed. States were classified into four groups by the marijuana statutes prior to sampling address lines: legal for recreational use, legal for medical use only with liberal access, legal for medical use only with restricted access, and not legal. ABS sample members were recruited by mail and asked to complete the 20-minute interview via the web. A $5 cash incentive along with detailed information about the NMS accompanied the recruitment notice. A reminder postcard was sent to households without a completed interview. The third and final mailing included a $2 incentive, the hardcopy questionnaire with return envelope, and details for accessing the online version of the questionnaire; sample members were asked to complete either version of the questionnaire.

A total of 14,110 addresses were selected from the ABS frame for the NMS (**Table 1**). Roughly equal proportions of addresses were selected from three strata groups—recreational use, medical use (liberal and restricted access), and not legal—to enable comparative analyses. Allocation to the two medical-use strata was in proportion to their adult populations.

Base weight accounting for the ABS stratified simple random sampling were adjusted for nonresponse via generalize exponential modeling with SUDAAN's WTADJUST procedure (RTI 2012). Model covariates included characteristics associated with the geographical location of the households. Nonresponse adjusted weights were then calibrated to population control totals by gender, age, race, education, and strata obtained from American Community Survey (ACS) public-use microdata sample (https://www.census.gov/programs-surveys/acs/technical-documentation/pums.html).

Overall, the ABS portion of the study resulted in a 14.5 percent RR2 weighted response rate, 13.2 percent unweighted (AAPOR 2016). Response rates were slightly higher for those living in the Midwest portion of the U.S. and in states where marijuana for medical needs was legal.

Table 1. Address-based sample, respondent counts, and response rates by state-level domains: 2016 National Marijuana Beliefs and Behaviors Study

	Sample		Respondent		Response Rate (%)[a]	
Domain	**n**	**%[b]**	**n**	**%[b]**	**Unweighted[c]**	**Weighted[d]**
Total	14,110	100.0	1,867	100.0	13.2	14.5
Region of the U.S.						
Northeast	1,792	12.7	203	10.9	11.3	13.6
Midwest	2,205	15.6	302	16.2	13.7	16.4
South	4,028	28.5	497	26.6	12.3	14.3
West	6,085	43.1	865	46.3	14.2	13.5
State Marijuana-use Statute						
Medical	4,703	33.3	710	38.0	15.1	16.8
Recreational, Liberal Access	2,800	19.8	340	18.2	12.1	14.5
Recreational, Restricted	1,941	13.8	223	11.9	11.5	13.4
Not Legal	4,666	33.1	594	31.8	12.7	14.7

[a] RR2 AAPOR response rate formula (AAPOR 2016).
[b] Unweighted proportion relative to the total in the column.
[c] The unweighted response rate is calculated as the number of respondents over the number of sample members by row.
[d] The weighted response rate is calculated as the weighted number of respondents over the weighted number of sample members by row using the base weight (inverse probability of selection).

Participants for the second of two samples were recruited from social media sites (e.g., Facebook) through advertisements and friend referrals. Recruitment was unrestricted and tailored to a set of distributions as in quota sampling. Perspective participants were screened for eligibility (age and residence within the U.S.) and viability of an email address prior to enrollment. Respondents received a $10 online gift card after completing the interview and after being verified as not having completed a prior interview. The social media sample was used for the NMS to address the projected shortage of adult respondents in younger age groups required for the analyses, and to ensure a sizeable number of interviews with marijuana users.

The NMS protocol captured 11,263 adults from the social media (SM) sites (**Table 2**). A higher proportion was recruited from the Western U.S. and, like the ABS sample, from states where marijuana for medical needs was legal. Additional protocols were administered to identify those enrolling multiple times, these protocols screened out 56.1

percent of the recruited adults. Retention rates were comparable across region and marijuana laws except for states with medical-restrictive policies.

Table 2. Social media screened sample and count/proportion retained for analyses by state-level domains: 2016 National Marijuana Beliefs and Behaviors Study

	Screened		Respondents		
Domain	**n**	**%[a]**	**n**	**%[a]**	**% screened[b]**
Total	11,263	100.0	4,943	100.0	43.9
Region of the U.S.					
Northeast	1,120	9.9	497	10.1	44.4
Midwest	1,339	11.9	618	12.5	46.2
South	2,798	24.8	1,159	23.4	41.4
West	6,006	53.3	2,669	54.0	44.4
State Marijuana-use Statute					
Recreational	4,416	39.2	2,020	40.9	45.7
Medical, Liberal Access	2,365	21.0	1,059	21.4	44.8
Medical, Restricted	1,435	12.7	478	9.7	33.3
Not Legal	3,047	27.1	1,386	28.0	45.5

[a] Unweighted percent relative to the total in the column.
[b] The unweighted completion rate is calculated as the number of respondents over the number of screened by row.

Questionnaires for the NMS ABS mail, ABS web, and SM samples were identical and included questions on demographic characteristics (gender, age, race/ethnicity), highest education attained, type of insurance coverage, family composition, access to the internet, and smoking status in the last 30 days. Other questions asked if they regularly vote in local/national elections, how they classify themselves on a liberal vs. conservative scale, whether they have ever used marijuana, and their opinions of marijuana use for medical and recreational purposes.

Other National Surveys. The NMS is not the only source of estimates for marijuana usage among adults in the United States. Examples include:

- The National Survey on Drug Use and Health (NSDUH) is a multistage, area probability survey with interviews conducted at sampled households. NSDUH marijuana (or hashish) questions include ever used, age at first use, time when used last, usage in past 12 months/30 days, ease of purchase, and effects on quality of life/lifestyle. Additional information on the NSDUH design and questionnaires is currently found at https://www.datafiles.samhsa.gov/study-series/national-survey-drug-use-and-health-nsduh-nid13517.
- A Gallup Poll was conducted via telephone on a random sample of households (McCarthy 2016). Questions include ever used and current use patterns, along with items to characterize religiosity.
- Another probability-based dual-frame telephone survey was the Yahoo News/Marist Poll: Weed & The American Family (http://maristpoll.marist.edu/yahoo-newsmarist-poll/).

Methods
The study evaluation began with a descriptive comparison of the characteristics within the probability and nonprobability samples. Characteristics included demographic and geographic variables, along with measures associated with marijuana use. Substantively meaningful differences were of particular interest for use in the sample matching algorithm.

Sample matching was conducted with the R MatchIt package (Ho et al. 2018; Stuart et al. 2011). The software allows a variety of matching algorithms and does not require a match for all records in either sample. Nearest neighbor algorithm was used for this demonstration with no caliper to allow a liberal definition of a match. A variety of models were evaluated to determine the sensitivity of the results. For brevity, two are discussed here. The first model incorporated only the 4-level stratum variable to capture state policies on marijuana usage (recreational, medical with liberal access, medical with restricted access, and not legal). The second sample matching model included the 4-level stratum variable and 11 other covariates: gender, age, race/ethnicity, education, employment status, regularly votes in local/national elections, type of insurance, children in the household, parent of child aged

18-21 years, liberal/conservative views, and smoked or vaped in the last 30 days. Model covariates were identified via the descriptive evaluation. As a precursor to the results discussed in the next section, the two models identified differing sets of probability sample respondents that were not part of the common support; matches for all nonprobability survey participants were identified but not required. The proportion of the target population covered by the two subsets was estimated from the unadjusted analysis weights to determine the extent of the unmatched samples.

The propensity score model incorporated all 12 variables used in the full sample matching model. From this model, propensity scores were calculated for the SM sample using three scenarios.

1) The first logistic model used all participant records from both surveys, effectively ignoring the results from the sample matching procedures.
2) The second model used the results from the first sample matching algorithm, which used an abbreviated list of matching covariates.
3) The third PSA model incorporated the enhanced sample matching model results. Using the same PSA model across the scenarios afforded the comparison of the sample matching results.

The resulting SM PSA weights were poststratified to the ACS population controls for stratum, gender, age group, and education to determine if the calibration adjustment lowered the mean square error for the estimates.

We then calculated a series of estimates with the SM analysis weights. Comparisons against the ABS survey estimates as well as national estimates from other surveys noted above were made to identify the relative magnitude of the estimated bias.

Results

Table 3 contains the unweighted proportion of cases by survey sample for a set of characteristics. For example, distributions by state stratum were similar for the ABS and SM samples; the highest proportion of participants for both samples came from states that allow recreational marijuana use. A higher proportion of Hispanic participants, those aged 34 years or less, and households with children were obtained from the SM sample, compared with non-Hispanic White and 65+ year old participants in the ABS sample. The ABS sample had a roughly equal distribution by gender, while the SM sample contained a higher proportion of females. The SM sample yielded a higher proportion of those who have ever used marijuana as expected, along with a higher proportion of those with very/somewhat liberal views. Therefore, we suspected that the support for each respondent set was slightly different.

Table 3. Unweighted Percent Distribution by Sample Source and Respondent Characteristics: 2016 National Marijuana Beliefs and Behaviors Study

	Unwtd Percent[a]			**Unwtd Percent[a]**	
	ABS	**SM**		**ABS**	**SM**
Characteristics	**n=1,867**	**n=4,943**	**Characteristics**	**n=1,867**	**n=4,943**
State stratum			Race/ethnicity		
Recreational	37.9	40.9	Hispanic	5.6	14.4
Medical, liberal	18.2	21.4	NH White	82.8	73.9
Medical, restrictive	11.8	9.7	NH Black	5.7	3.7
Not legal	32.1	28.0	NH Other	5.8	7.2
Age category (years)			Employment status		
18 - 24	4.0	12.2	(Self) Employed	57.6	52.1
25 - 34	12.5	25.7	Out of work	2.0	5.5
35 - 44	16.1	18.9	Homemaker	3.6	12.7
45 - 54	17.9	14.5	Student	2.6	5.9
55 - 64	23.1	17.6	Retired	26.6	14.3
65+	26.4	11.2	Unable to Work	3.6	8.5

(*continued*)

Table 3. Unweighted Percent Distribution by Sample Source and Respondent Characteristics: 2016 National Marijuana Beliefs and Behaviors Study (continued)

	Unwtd Percent[a]			Unwtd Percent[a]	
	ABS	SM		ABS	SM
Characteristics	n=1,867	n=4,943	Characteristics	n=1,867	n=4,943
Education			Gender		
No HS diploma/GED	2.3	2.4	Female	54.3	71.5
HS diploma/GED	16.2	15.1	Male	45.7	26.4
Some college/Assoc	35.6	44.6	(Nearly) Always votes		
Bachelors	23.1	21.9	Yes	77.9	67.3
Masters or higher	22.8	15.6	No	19.2	31.1
Private insurance			Very/somewhat liberal views		
Yes	55.5	43.8	Yes	30.5	44.4
No	38.8	54.0	No	64.4	52.3
Child in household			Child aged 18-21 years		
Yes	26.5	41.9	Yes	11.9	10.7
No	72.1	57.7	No	86.7	89.2
Internet access?			User of social media?		
Yes	87.1	94.3	Yes	68.2	96.8
No	6.4	3.1	No	26.1	1.2
Smoked/vaped (last 30 days)			Ever used marijuana		
Yes	15.5	41.1	Yes	56.5	69.2
No	81.9	58.8	No	41.2	29.6

Note: ABS = address-based sample; SM = social media sample; NH = non-Hispanic.
[a] Unweighted column percentages may sum to less than 100 because of item nonresponse.

Table 4 contains the proportion of ABS respondents that were found to have at least one match within the SM sample by model. For ease of discussion, the first model is labeled as "none" to signify that a sample matching algorithm was not used. The augmented model (#3) identified a match for all SM records from only approximately 67.0 percent of the ABS sample cases, accounting for an estimated 73.7 percent of the population. The less restrictive model (#2) could match approximately 92 percent of the ABS sample to at least one SM participant, accounting for an estimated 89.0 percent of the population.

Table 4. Unweighted and weighted rate for ABS respondents to one or more SM participants: 2016 National Marijuana Beliefs and Behaviors Study

Sample matching model		ABS match rate (%) Unweighted	Weighted	Est'd Population Size
1	none	100.0	100.0	243,702,042
2	Strata (4)	92.1	89.0	216,809,591
3	Strata (4) + 11 characteristics[a]	67.0	73.7	179,548,699

Note: ABS = address-based sample; SM = social media sample.
a Characteristics include gender, age, race/ethnicity, education, employment status, regularly votes in local/national elections, type of insurance, children in the household, parent of child aged 18-21 years, liberal/conservative views, and smoked or vaped in the last 30 days.

Table 5 provides descriptive statistics for the final ABS analysis weights. Details for the PSA SM and final (PSA plus calibration) weights are also included. The design effect of the final ABS weights, also known as the unequal weighting effect or UWE, is 3.34. The variability of all the SM weights are all higher than this level; sample matching Model 3 produced a noticeable reduction in the UWE (4.43) that was further contained through poststratification (3.89). The comparative increase in the effective sample size was also notable across the SM weights. The weight variability resulted in efficiencies of the simple random sample design with at most 25.7 percent of the actual participant size of 4,943 compared with 29.9 percent in the ABS sample.

Table 5. Statistics for ABS and SM weights: 2016 National Marijuana Beliefs and Behaviors Study

Sample	Weight[a]	UWE[b]	Est'd Population[c]	Resp (n)	Effective size[d] n	pct
ABS	Final analysis weight	3.34	243,702,042	1,867	559	29.9
Social Media	PSA weight, Model 1	5.15	106.3	4,943	960	19.4
	PSA weight, Model 2	5.13	94.7		964	19.5
	PSA weight, Model 3	4.43	79.1		1,116	22.6
	Final weight, Model 1	4.39	100.0		1,125	22.8
	Final weight, Model 2	4.43	100.0		1,116	22.6
	Final weight, Model 3	3.89	100.0		1,272	25.7

Note: ABS = address-based sample; PSA = propensity score adjustment; Resp (n) = unweighted respondent count
[a] Sample matching models are defined in Table 4. The final weight is defined as the PSA weight calibrated to population controls.
[b] UWE (unequal weighting effect) = design effect calculated as the standard deviation of the weights divided by the mean weight.
[c] The social media values are shown as percentages where the numerator is the sum of the weights divided by the estimated population size from the ABS sample.
[d] Effective sample size = respondent count / UWE.

A series of point estimates was calculated from each source of NMS responses with the analysis weights listed in Table 5. The corresponding estimates were calculated from the ABS data using the final analysis weights for the full sample and for each of the two subsets identified with the sample matching algorithm. To answer the question on whether weighting was needed for the SM sample, **Table 6** displays a comparison of a few ABS-only full sample estimates (here used as the gold standard) with the corresponding unweighted SM-only sample estimates. Here we see that the unadjusted SM estimates are much larger than the weighted ABS estimates.

Table 6. Weighted ABS and Unweighted SM Marijuana Estimates: 2016 National Marijuana Beliefs and Behaviors Study

Source	Ever used	Support for medical use	Support for recreational use
ABS final analysis weight	56.6	77.0	47.7
SM, unweighted	70.0	85.7	60.5
Difference (SM – ABS)	13.4	8.7	12.9

Note: ABS = address-based sampling; SM = social media

Table 7 provides a comparison of various ABS and SM weighted estimates overall. The full-sample ABS estimates are used as the gold standard for comparison against estimates using the other weights. Here we see that 56.6 percent of adults are estimated to have ever used marijuana, while 77.0 percent and 47.7 percent of U.S. adults support marijuana use for medical and recreational purposes, respectively. The corresponding estimates were calculated from the ABS data using the final analysis weights for the full sample and for each of the two subsets identified with the sample matching algorithm. ABS Model 2 (the 92-percent subset; see Table 4) estimates for ever used and support for medical use were similar to the gold standard values; however, the percent of adults who support recreational use is estimated to be 2.6 percentage points higher. The similarities also existed for the full-sample and ABS Model 3 (the 67-percent subset of the full ABS sample) "support for medical use" estimates; the ever-use and recreational-use were much higher than the gold standard estimates. Calibration of the analysis weights within the two subsets did not change the estimates in a substantively meaningful way. Statistical tests indicate that only the support for recreational use estimate for Model-3 was significantly different from the gold standard ($p\text{-value}<0.05$).

Therefore, the Model-3 subset does not meet the MAR condition that will inform the common support evaluation with the SM estimates and the "s_A" component in expression (2) is not necessarily zero.

The "ever used" SM estimates in Table 7 are closest to the full-sample ABS estimates with the Model-3 data. The other analysis variables remain significantly different with or without a calibration adjustment after PSA. Overall, the calibration does appear beneficial for the point estimates regardless of the sample matching model used.

Table 7. Weighted ABS and SM Marijuana Estimates: 2016 National Marijuana Beliefs and Behaviors Study

Sample	Weight[a]	Ever used[b]	Support for medical use[b]	Support for recreational use[b]
ABS	Final analysis weight	**56.6**	**77.0**	**47.7**
	Model 2	-0.5	0.5	2.6
	Model 3	3.8	0.2	8.0 *
	Calibrated, Model 2	-0.6	0.3	2.4
	Calibrated, Model 3	4.2	1.6	7.9 *
Social Media	PSA weight, Model 1	-1.2	6.4 *	8.3 *
	PSA weight, Model 2	-1.3	6.4 *	8.3 *
	PSA weight, Model 3	-0.2	7.6 *	9.7 *
	Final weight, Model 1	-0.9	5.5 *	7.1 *
	Final weight, Model 2	-1.0	5.6 *	7.0 *
	Final weight, Model 3	0.7	6.1 *	8.4 *

Note: ABS = address-based sampling
* Statistically different from 0 with p-value<0.05.
[a] Sample matching models are defined in Table 4.
[b] Population estimates for the ABS sample calculated with the final analysis weight are considered the "gold standard". All other population estimates are shown as relative difference, i.e., population estimate – gold standard (ABS).

Table 8. Weighted ABS and SM Marijuana Estimates by Respondent Characteristics: 2016 National Marijuana Beliefs and Behaviors Study

		Respondent Characteristics					
		Females			Retirees		
Sample	Weight	Ever	Med	Rec	Ever	Med	Rec
ABS	Final analysis weight	**52.2**	**73.2**	**45.3**	**38.0**	**68.9**	**32.0**
	Model 2	-0.2	-3.9	2.2	1.7	-3.9	-3.1
	Model 3	4.6	-1.7	6.7	1.7	-2.8	-0.8
	Calibrated, Model 2	-0.2	-4.0	2.0	1.7	-3.7	-3.2
	Calibrated, Model 3	3.8	-1.4	5.6	3.7	-1.5	0.2
Social Media	PSA weight, Model 1	0.9	12.3	6.9	10.9	9.9	10.8
	PSA weight, Model 2	1.1	12.3	7.0	10.7	10.0	10.5
	PSA weight, Model 3	2.3	13.1	8.9	12.8	10.8	12.0
	Final weight, Model 1	1.0	11.7	5.8	12.8	10.4	10.9
	Final weight, Model 2	1.0	11.8	5.7	12.4	10.5	10.5
	Final weight, Model 3	2.4	12.3	7.3	14.7	11.3	12.9

Note: ABS = address-based sampling; Ever = ever used marijuana; Med = support for medical use; Rec = support for recreational use

Table 8 repeats the analyses shown in Table 7 for two subgroups. Much of the discussion for Table 7 is also repeated for the results in Table 8. Even though gender is included as a marginal covariate in the calibration models, the positive bias still exists. This result suggests further evaluation of the propensity and calibration model is needed, likely with the addition of interaction terms and other covariates identified via a regression tree analysis (Valliant & Dever 2018).

Table 9 contains the estimated proportion of adults in the U.S. who have ever used marijuana from various sources. The proportion of adults who have ever used marijuana estimated from the NMS ABS-only sample is much higher than similar estimates from the other probability-based surveys. We also note the limited comparability of the estimates for the three external surveys. Differences in the estimates for all surveys in Table 9 may be attributed to differences in data collection mode, sample composition, weighting, and question wording. Ideally these comparisons would be made among surveys with the same essential survey conditions.

Only one of the surveys examined, the 2017 Yahoo News/Marist Poll, had questions on opinions of approved marijuana uses. The medical-use estimate from the NMS ABS sample in Table 9 was the smaller of the two. Conversely, the recreational-use estimates from NMS and the probability-based telephone poll were comparable.

Results shown in Table 9 suggest that additional weight adjustments for the NMS ABS survey may be warranted prior to the common-support evaluation. Such adjustments may include calibration to population estimates of marijuana use to align NMS with reliable national surveys (Dever & Valliant 2016). Considering the differences in the three (and possibly other relevant) surveys, source for the calibration controls should be based on precision and commonality of the survey questions.

Table 9. Marijuana Usage Estimates for Adults in the U.S. by Survey Source

	Ever used		Support for medical use		Support for recreational use	
Survey[a]	**pct**	**se**	**pct**	**se**	**pct**	**se**
2016 NMS ABS	56.6	2.13	77.0	1.94	47.7	2.23
2016 NSDUH (CBHSQ 2017)	47.0	0.35	na		na	
2017 Yahoo News/Marist Poll (2017)	48.0	1.48	83.0	1.48	49.0	1.48
2016 Gallup Poll (McCarthy 2016)	43.0	nr	na		na	

Note: pct = weighted percentage; se = weighted standard error; nr = not reported; na = not available.
[a] Surveys other than the 2016 National Marijuana Beliefs and Behaviors Study (NMS), Address-based Sample (ABS) are described in the 'National Estimates of Marijuana Use' section.

Conclusions and Future Research

Composite / hybrid estimation, and propensity score adjustment (PSA) methods all rely on common support to produce appropriate population estimates. The results presented here are preliminary. However, they suggest that even slight violations of the assumption can introduce additional bias into the nonprobability estimates. Once determined, this overlap informs the records used in the PSA models and the non-zero components of the estimator shown in expression (2). However, consistent methods are needed to test for common support. Additional research in this area is already underway that includes a simulation study instead of reliance on just a single study.

Our research also demonstrated that some bias reduction in the nonprobability estimates can occur with the constrained PSA estimation space, as created with sample matching Model 3. This issue affects not only hybrid estimation but situations where only data from a nonprobability survey are available. Consequently, this benefit should be examined with other methods used to create the nonprobability participant weights, in addition to an evaluation of the PSA model.

Hybrid estimation is a fruitful area of research for years to come. Paramount is the question on the creation of the lambda factor that glues the samples into one data file for population inferences. Second, is whether inclusion of the nonprobability sample is beneficial because mean square errors may be larger for hybrid estimates relative to probability estimates alone. Informed by guidance for these questions (because "it depends" is likely the answer), researchers may one day soon be able to design hybrid estimation system where appropriate. With these best practices, sample size and allocation to the survey sources may be attainable. This will be in stark contrast to situations of today that are more happenstance both with regards to sample management of both surveys and how to maximize information from the resulting data. That said, let's get to work.

References

American Association for Public Opinion Research (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 9th edition. AAPOR. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., et al. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, ***1***:90–143. http://jssam.oxfordjournals.org/content/1/2/90.full.pdf+html.

Brick, J.M. (2011). The Future of Survey Sampling. *Public Opinion Quarterly*, ***75***(5):872–888. https://doi.org/10.1093/poq/nfr045

Brick, J.M., Cervantes, I.F., Lee, S. and Norman, G. (2011). Nonsampling errors in dual frame telephone surveys. *Survey Methodology*, ***37***:1–12.

Callegaro, M., Baker, R.P., Bethlehem, J., Göritz, A.S., Krosnick, J.A., and Lavrakas, P.J. (Eds.; 2014). *Online Panel Research: A Data Quality Perspective*. United Kingdom: John Wiley & Sons, Ltd.

Cavallo, A. and Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, ***30*** (2):151-78.

Center for Behavioral Health Statistics and Quality (CBHSQ; 2017). 2016 National Survey on Drug Use and Health: Detailed Tables. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Clement, S. (2016). How the Washington Post-SurveyMonkey 50-state poll was conducted. *The Washington Post*, 6 Sep 2016.

Conway, B.A., Kenski, K. and Wang, D. (2015). The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. *Journal of Computer-Mediated Communication*, ***20***(4):363-380.

Dever, J.A., and Brown, D. (2016). Estimated-control calibration for non-probability surveys. Presented at AAPOR 71st annual conference, Austin, TX.

Dever, J.A., and Valliant, R. (2016). General regression estimation adjusted for undercoverage and estimated control totals. *Journal of Survey Statistics and Methodology*, ***4***(3):289-318.

Dropp, K. and Nyhan, B. (2016). The 2016 Race: It Lives. Birtherism Is Diminished but Far from Dead. *The New York Times*, 23 Sep 2016. Accessed 25 Sep 2016.

Dutwin, D. and Buskirk, T.D. (2017). Apples to Oranges or Gala versus Golden Delicious?: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples. *Public Opinion Quarterly*, ***81***(S1):213–239.

Elliott, M.R. and Valliant, R. (2017). Inference for Nonprobability Samples. *Statistical Science*, ***32***(2):249–264.

Groves, R.M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. Public Opinion Quarterly, 72(2):167–189. https://doi.org/10.1093/poq/nfn011.

Harris, J.K., Mansour, R., Choucair, B., Olson, J., Nissen, C., and Bhatt, J. (2014). Health Department Use of Social Media to Identify Foodborne Illness--Chicago, Illinois, 2013-2014. *Morbidity and Mortality Weekly Report*, ***63***(32):681-685.

Harter, R., Battaglia, M.P., Buskirk, T.D., Dillman, D.A., English, N., et al. (2016). *Address-based Sampling*. Prepared for AAPOR Council by the Task Force on Address-based Sampling (January 7, 2016), https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx#3.3%20Auxiliary%20Data%20Quality.

Hays, R.D., Liu, H., and Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behavior Research Methods*, ***47***(3):685–690. https://link.springer.com/article/10.3758/s13428-015-0617-9

Ho, D., Imai, K., King, G., Stuart, E., and Whitworth, A. (2018). *MatchIt: Nonparametric Preprocessing for Parametric Causal Inference*, 9 Jan 2018. R package version 3.0.2. https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf.

Horvitz, D.G. and Thompson, D.J. (1951). A Generalization of Sampling without Replacement from a Finite Universe. *Journal of the American Statistical Association*, ***47***:663-685.

Iannacchione, V.G. (2011). Research synthesis: The Changing Role of Address-Based Sampling in Surveys. *Public Opinion Quarterly*, ***75***(3):556-576. https://academic.oup.com/poq/article/75/3/556/1834922

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P. et al (2015). Big Data in Survey Research: AAPOR Task Force Report. *Public Opinion Quarterly*, ***79***(4):839–880. https://doi.org/10.1093/poq/nfv039

Keeter, S., Hatley, N., Kennedy, C., and Lau, A. (2017). "What Low Response Rates Mean for Telephone Surveys: Telephone polls still provide accurate data on a wide range of social, demographic and political variables, but some weaknesses persist." Pew Research Center Series on Numbers, Facts and Trends Shaping Your World, 15 May 2017. http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/, accessed 17 Apr 2018.

Kim, A.E., Hopper, T., Simpson, S., Nonnemaker, J., Lieberman, A.J., et al. (2015). Using Twitter Data to Gain Insights into E-cigarette Marketing and Locations of Use: An Infoveillance Study. *J Med Internet Res*, ***17***(11):e251.

Kleinman, A. (2014). Facebook Can Predict with Scary Accuracy If Your Relationship Will Last. *The Huffington Post*, 14 Feb 2014.

Kott, P.S. (2017). A Partially Successful Attempt to Integrate a Web-Recruited Cohort into an Address-Based Sample. Presented at NISS/WSS Workshop on Inference from Nonprobability Samples, 25 Sep 2017 in Washington, DC. https://www.niss.org/sites/default/files/event_attachments/KottINPS92517.pdf.

Kott, P.S. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Survey Methodology*, ***32***(2): 133–142.

Krueger, B.S. and West, B.T. (2014). Assessing the Potential of Paradata and Other Auxiliary Data for Nonresponse Adjustments. *Public Opinion Quarterly*, ***78***(4):795–831, https://doi.org/10.1093/poq/nfu040.

Lee, S. and Valliant, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research*, ***37***:319-343.

Lohr, S.L. and Raghunathan, T.E. (2017). Combining Survey Data with Other Data Sources. *Statistical Science*, ***32***(2):293-312.

Lohr, S. and Rao, J.N.K. (2006). Estimation in Multiple-Frame Surveys. *Journal of the American Statistical Association*, ***101***(475):1019-1030. https://doi.org/10.1198/016214506000000195

McCarthy, J. (2016). One in Eight U.S. Adults Say They Smoke Marijuana. Gallup Poll Social Series Report, 8 Aug 2016, http://news.gallup.com/poll/194195/adults-say-smoke-marijuana.aspx. Accessed 6 May 2018.

Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, ***33***(2): 151-157.

Mercer, A., Caporaso, A., Cantor, D., and Townsend, R. (2015). How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly*, ***79***(1):105–129. https://doi.org/10.1093/poq/nfu059.

Mercer, A.W., Kreuter, F., Keeter, S., and Stuart, E.A. (2017). Theory and Practice in Nonprobability Surveys: Parallels Between Causal Inference and Survey Inference. *Public Opinion Quarterly*, ***81***:250–279. https://academic.oup.com/poq/article-abstract/81/S1/250/3749176

Mercer, A., Lau, A., and Kennedy, C. (2018). *For Weighting Online Opt-In Samples, What Matters Most? The right variables make a big difference for accuracy. Complex statistical methods, not so much.* Pew Research Center, 26 Jan 2018. www.pewresearch.org/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/, accessed 26 Apr 2018.

Merkouris, T. (2010). Combining information from multiple surveys by using regression for efficient small domain estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, ***72***:27–48.

Pasek, J. (2015). When will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence. *International Journal of Public Opinion Research*, 28(2):269–291. https://doi.org/10.1093/ijpor/edv016

Peytchev, A., Presser, S., and Zhang, M. (2018). Improving Traditional Nonresponse Bias Adjustments: Combining Statistical Properties with Social Theory. *Journal of Survey Statistics and Methodology*, smx035, https://doi.org/10.1093/jssam/smx035.

Research Triangle Institute (2012). *SUDAAN Language Manual*, Volumes 1 and 2, Release 11. Research Triangle Park, NC: RTI.

Rivers, D., and Bailey, D. (2009). Inference from Matched Samples in the 2008 U.S. National Elections. In the Proceedings of the Joint Statistical Meetings, *American Statistical Association*: 627–639. http://ww2.amstat.org/sections/SRMS/Proceedings/y2009/Files/303309.pdf

Särndal, C.E., Swensson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling*. New York, NY: Springer-Verlag, Inc.

Solomon, S.S., McFall, A.M., Lucas, G.M., Srikrishnan, A.K., Kumar, M.S., Anand, S., Quinn, T.C., Celentano, D.D., and Mehta, S.H. (2017). Respondent-driven sampling for identification of HIV- and HCV-infected people who inject drugs and men who have sex with men in India: A cross-sectional, community-based analysis. *PLOS Medicine* (28 Nov 2017), https://doi.org/10.1371/journal.pmed.1002460.

Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, ***25***(1):1-21. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943670/
Stuart, E.A., Cole, S.R., Bradshaw, C.P. and Leaf, P.J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society*, Series A (Statistics in Society), ***174***(2):369-386.
Stuart, E.A., King, G., Imai, K., and Ho, D. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, ***42***(8):1-28. http://www.jstatsoft.org/v42/i08.
Teitcher, J.E.F., Bockting, W.O., Bauermeister, J.A., Hoefer, C.J., Miner, M.H., and Klitzman, R.L. (2015). Detecting, Preventing, and Responding to "Fraudsters" in Internet Research: Ethics and Tradeoffs. *Journal of Law, Medicine & Ethics*, ***43***(1):116–133. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4669957/
Tourangeau, R. (2014). Defining hard-to-survey populations. In Tourangeau R., Edwards, B., Johnson, T.P., Wolter, K.M., and Bates, N. (Eds.), *Hard-to-Survey Populations*, Chapter 1. United Kingdom: Cambridge University Press.
Valliant, R. (2018). Comparing Alternatives for Estimation from Nonprobability Samples. Proceedings from the *Federal Committee on Statistical Methodology* (FCSM) Research Conference, 7-9 Mar 2018, Washington, DC.
Valliant, R. and Dever, J.A. (2018). *Survey Weights: A Step-by-step Guide to Calculation*. College Station: Stata Press.
Valliant, R. and Dever, J.A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods and Research*, ***40***:105-137.
Valliant, R., Dever, J.A., and Kreuter, F. (2018). *Practical tools for designing and weighting sample surveys* (2^{nd} ed.). New York: Springer.
Wu, J., Crawford, F.W., Raag, M., Heimer, R., and Uusküla, A. (2017). Using data from respondent-driven sampling studies to estimate the number of people who inject drugs: Application to the Kohtla-Järve region of Estonia. *PLOS One* (2 Nov 2017), https://doi.org/10.1371/journal.pone.0185711.
Yahoo News/Marist Poll (2017). Report: Weed and the American Family, 17 Apr 2017, http://maristpoll.marist.edu/wp-content/misc/Yahoo...Family%20(1).pdf. Accessed 05 May 2018.
Yeager, D.S., Krosnick, J.A., Chang, L., Javitz, H.S., Levendusky, M.S., Simpser, A., and Wang, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, ***75***(4):709–747.