Is It Still Possible to Collect Nationally Representative Data in the United States? A Case Study from the CREATE Project

Spencer James

Jeremy Yorgason

Erin Holmes

Dean Busby

Brigham Young University

David Johnson

The Pennsylvania State University

Abstract

The United States is undergoing impressive and transformational social change related to marriage. Social scientists' ability to study such changes are contingent upon being able to collect representative data. Given the expected low response rates in contemporary survey research, it is natural to ask whether it is even possible to still collect high-quality, nationally representative survey data on marriage and family. This paper presents our process of collecting nationally representative data and discusses whether it is still possible to collect nationally representative data in the current data collection climate.

Key Words: Nationally representative, survey research

Is It Still Possible to Collect Nationally Representative Data in the United States? A Case Study from the CREATE Project

The United States is undergoing impressive and transformational social change. These changes hold important implications for the wellbeing of current and future generations. One particularly turbulent aspect for many Americans is the ever shifting role of family life, in large part due to changes in the economy and concomitant increases in income and wealth inequality (Blau, 1998; Ellwood & Jencks, 2004; Piketty, Saez, & Zucman, 2016). An ever increasing proportion of the American population is experiencing delayed and foregone marriage, increasing poverty and nonmarital childbearing, relational instability and high divorce rates, and rising levels of cohabitation, all of which are thought to be tied to global demographic trends encompassed in the second demographic transition (Lesthaeghe, 2010).

Tracking and accurately capturing these changes requires a large undertaking and there have been discussions on the need for new, nationally representative data (House, 2015; Moffitt et al., 2015) and specifically to address emerging questions about American family life (Manning, 2015; Raley, 2015; Seltzer, 2015). Given the increasing complexity and nuances of contemporary family relationships and demographics (Cherlin, 2010), social scientists' ability to respond to such calls to action and the individual and the societal benefits of social science research are both contingent upon our ability to a) collect data that mirror the population, b) draw appropriate conclusions based on a sample's generalizability, c) analyze the data rigorously in accordance with prevailing scientific and statistical standards and d) do all of these things in a timely and orderly fashion, without imposing unnecessary burden on respondents (Moffitt et al., 2015). This is a tall task, as substantial declines in survey responses are partly due to inundation from surveys of varying quality and motive that levy significant costs on respondents (Dutwin et al., 2014).

Given that response rates to current surveys are often in the high single-digits to low teens, it is natural to ask whether it is even possible to still collect high-quality, nationally representative survey data on marriage and other family-related topics (Kohut, Keeter, Doherty, Dimock, & Christian, 2012).

In this paper, we discuss a partial response to these calls, the Couples Relationships and Transition Experiences (CREATE) study, a nationally representative study of newly married couples aged 18-36 married primarily from May-December, 2014. Of course, this study does not respond wholly to the needs discussed by Moffit et al. (2015), as we have neither the resources nor the expertise to mount a data collection effort that would be of general interest across many disciplines, as they call for. We have, however, collected a nationally representative dataset of almost 2200 newlywed couples in a very challenging data environment.

Falling Response Rates

It is, by now, somewhat trite to point out the ability to track and capture changes surrounding contemporary families largely depends on surveys, as there are few other tools at our disposal for assessing national level trends. However, response rates have been falling now for several decades, potentially limiting the ability to draw generalizable conclusions based on these samples (Czajka & Beyler, 2016). Survey response rates used to be quite high in nationally representative surveys employed by social scientists. For instance, the original survey response rate to the National Survey of Families and Households, collected in 1988, was 74% (Wright, 2003). The National Longitudinal Survey of Youth-179 cohort reported a response rate of between 75% and 80% (Frankel et al., 1983) and the 1997 cohort was even higher at 94%

(Moore et al., 2000). Similarly, the original response rate for wave 1 of the National Longitudinal Study of Adolescent Health (Add Health) was 79% (Carolina Population Center, n.d.).

Recent trends in response rates, however, have been uniformly downward (Brick & Williams, 2013; National Research Council, 2013). While government-funded and administered surveys have still managed to retain relatively high response rates for cross-sectional surveys (Czajka & Beyler, 2016), surveys of the general population, even when administered by established institutions such as Pew, have seen dramatic dips in their response rates over time, especially for telephone-based surveys (Keeter, Hatley, Kennedy, & Lau, 2017). In fact, Pew reports that their response rates (AAPOR RR3; (The American Association for Public Opinion Research, 2016)) usually vary from 5% to 15% and suggests that these numbers are in line with other major polling and survey efforts. Kohut, Keeter, Doherty, Dimock, and Christ (2012), for instance, found that the response rates in the studies included in their analysis dropped from 36% in 1997 to just 9% in 2012, and there is no evidence of an upward spike in survey trends in the past 5 years.

Researchers have postulated many reasons for these precipitous declines, commonly classified using three categories: non-contact, refusal, and other. The final category includes issues surrounding language, health, and time spent at home. According to Czaijka (2016), there is a litany of reasons for the increase in nonresponse rates, including the number of dual-earner households, increases in commuting time, advent of cell phones and other phone-related technology such as caller ID, numerous federal surveys, an increase in the use of ‘push’ polling (polls designed to change a respondent’s opinion rather than assess it), the growth of marketing-related ‘spam’ calls and fears of unknown numbers, and the use of communication modes, such as internet surveying, that have not yet been extensively (but are increasingly being) used by researchers. Somewhat counterintuitively, the national Do Not Call Registry does not appear to have had a marked influence on response rates, at least at the state level (Link et al., 2006).

Consequently, there is a need to employ multiple approaches, or modes, of data collection to ensure we capture the largest number of people possible. To do this, many studies have examined the usefulness of various methods, typically with those employing telephone or mail-based surveys reporting lower response rates than those employing internet-based data collection techniques (Chang & Krosnick, 2009; Dillman et al., 2009), although the results are not always consistent.

Responses to Declining Response Rates in Nationally Representative Surveys

In response to this environment of plunging response rates, researchers have pursued several approaches to both reduce survey nonresponse (and, it is hoped, nonresponse bias) as well as finding other ways to collect data that would ideally be collected via random probability samples.

One approach to improving response rates has been to collect data in multiple ways (i.e., modes). It is hoped that by changing the data collection methodology that survey researchers will be able to reach different populations and, when pieced together, obtain a higher response rate than what one would obtain via a single mode. After all, continued reliance upon a single method or more for collecting data-for example, via random digit dialing (RDD)—may increase the number of interviews, but it may not actually reduce nonresponse bias due to similarities across those most likely to respond to a particular survey methodology.

Thus, Kreuter (2013) points out that, at its heart, the goal of mixed-mode surveys is to design the survey methodology in such as way so as to appeal to various subpopulations most likely to respond to particular methodologies. For example, some individuals may be most likely to respond to random digit dialing while others are most likely to respond to a text, email, or an online survey. These modes (RDD, texting, email, online survey), when used together, may be used in hopes of ameliorating known deficiencies of each mode (Czajka & Beyler, 2016).

Accompanying efforts to appeal to various subpopulations via different survey modes have been efforts to reduce respondent burden. Several possible ways to reduce respondent burden have been proposed, varying from shorter questionnaires, spreading out the survey across several subsamples, or combining responses with existing administrative data that can help ‘fill in’ for nonresponse (Kreuter, 2013). And there is evidence that reducing burden may improve response rates. A recent experiment suggested that combining multiple response-inducing techniques, including employing various modes of data collection and even a small cash incentive, can improve response rates (Millar & Dillman, 2011).

Another common response to declining response rates has been to explore the possibility of using non-probability samples more judiciously. The hope is that if we understand nonresponse bias and can therefore correct for it, we may be able to learn as much from “national samples” as “nationally representative” ones. On the one hand, there may be some wisdom in this approach. The statistical law of large numbers suggests that as sample size increases, the amount of bias from nonresponse decreases as the sample size approaches the population size (i.e., $n \rightarrow N$). Thus, if the sample size is sufficiently large, the inferences drawn from such a sample may be reasonably unbiased, especially as the relationship between response rates, sample sizes, and nonresponse bias is not always straightforward nor predictable (Massey & Tourangeau, 2013).

However, the practice of using national samples, often drawn from large panels of people maintained by marketing or commercial survey research firms whose primary customers tend to come from industry rather than academia has resulted in a blurring of the difference between a nationally representative sample, wherein researchers randomly select the participants of their study, and national quota samples, where researchers establish quotas to fill and select study participants but the selection process is not random. In some instances, national convenience samples of marketing panels, wherein the study participants opt into participation on the panel themselves rather than being selected by the researcher, have been called nationally representative (we omit citations here to avoid controversy). As a field, we may be well served to carefully distinguish between nationally representative and national (quota or convenience) samples, as the two are not interchangeable although both may provide useful information.

Some promising recent research in political science on political attitudes and preferences has found that there may be ways to adjust non-probability samples to mirror the gains from probability samples (Wang, Rothschild, Goel, & Gelman, 2015; Zhang et al., 2017). These methods, however, may prove quite difficult to implement in practice because many family surveys do not have the sample size necessary to "partition the data into thousands of demographic cells, estimate [the parameter of interest] at the cell level using a multilevel regression model, and finally aggregate the cell-level estimates in accordance with the target population's demographic composition." (Wang et al., 2015). In the absence of data that make such analyses possible, such promising results regarding the use of external information to inform and weight nonprobability data to approximate data gained using probability samples may be quite limited.

In addition, two studies have examined how various combinations of data collection modes and sampling approaches can influence results. Craig, Hays, Pickard, Cella, Revicki and Reeve (2013) sought to examine how the quality of data varied across panel vendors (firms that enroll and match willing participants to a survey's target audience). Using identical quotas and online surveys across seven panels, they found discordance (20%) between self-reported birth dates and the reported date given when entering the panel. Furthermore, they found another fifth of respondents overlapped across panels (*i.e.*, roughly 20% of respondents in six of the seven panels participated in multiple vendor panels). Additionally, none of the samples adequately represented adults with less than a high school education or those making less than $15,000. Similarly, Chang and Krosnick (2009) compared random digit dialing both nonprobability and probability internet samples. They found that while responses to the nonprobability samples were the most accurate[1], these responses also came from the most biased sample. On the other hand, the internet probability sample demonstrated the ideal blend of sample representative and self-reporting, suggesting that probability samples collected on the internet may be among the most ideal conditions for modern data collection, despite the clear appeal of collecting nonprobability samples and weighting the data *a posteriori*. Referencing specifically nonprobability online samples, Couper made the following observation:

> For academic researchers, nonprobability online samples are often the only affordable option. Such samples are not inherently incorrect, but they increase the risk of inferential errors over probability-based approaches and should be used with caution and with an explicit discussion of the likely inferential limits" (2017, pp. 129–130).

[1] This may make nonprobability samples the ideal place for measure development.

The same cautions can likely be applied to most other types of nonprobability samples, with the possible exception of instances where researchers have sufficiently rich data, both in terms of sample size and measurement, to apply the methods spelled out by Wang et al. (2015).

Given the importance of accurately capturing, analyzing, and understanding the dramatic changes affecting families today, changes that are unlikely to abate in the near future, we present a case study of nationally representative couple data, the Couple Relationships and Transition Experiences (CREATE) study, a sample of 2,186 couples who were married primarily between May and December of 2014.

The Couple Relationships and Transition Experiences (CREATE) Study—Sampling and Weighting Considerations

Sample and Procedures. The CREATE study is a nationally representative survey of newly married young couples. Participants for the study were recruited using a two-stage cluster stratification sample design, with the first stage involving a sample of counties, and the second involving a sample of recent marriages within those selected counties. Counties were selected based on a probability proportion to size (PPS) design. Selection was based on county population size, marriage, divorce, and poverty rates, and the racial-ethnic distribution of the county, with an overselection of counties high in poverty and minority population. The number of marriages selected per county ranged from 40 to 280, depending on these five characteristics. This design yielded a sampling frame of 11,960 marriages across 239 counties. Ten counties did not have at least 40 marriages during the sampling period, leaving the final sampling frame at 11,889 marriages.

In the second stage, marriage record information was used, with assistance from publicly available databases, to locate couples and invite them to participate. To be included in the sample, respondents had to (a) be married and selected into the sample frame (since some marriage applicants did not end up marrying), (b) have at least one partner between 18 and 36 years of age at the start of the study, (c) be a first marriage for at least one of the partners in the dyad, and (d) be living within the U.S. The majority of couples in the study were married during 2014 (90%), with the remainder in 2013 (4%) and 2015 (6%). The study was approved by all appropriate IRB bodies.

Based on the Dillman survey method, potential participants were first contacted by mailed letters that contained a $2.00 bill with an invitation to participate and instructions on how to enroll in the study (Dillman, Smyth, & Christian, 2008). For those that did not respond to the initial invitation, follow-up postal mailings, E-mail invitations, and phone calls were made. As is common with online surveys, participants were asked to read and then acknowledge consent to participate in the study. Participating couples were given a $50.00 Visa gift card upon completion of the survey.

Among the 11,889 couples contacted, 8140 declined participation by either not answering or responding, and 1,220 did not meet inclusion criteria. A total of 2,187 marriages were recruited into the study, drawing a *raw response rate* (AAPOR RR1) of $18.24\% = \frac{1,187}{11,189}$. After dropping ineligible couples, the *adjusted response rate* was $20.50\% = \frac{2,187}{10,669}$. Of the 2,187 marriages, data from both members of the dyad were received in 1,889 (86%) cases, and data from one member of the dyad were received in the remaining 298 (14%) cases.

Additional information gained in the recruiting process allowed us to estimate a more accurate response rate, in accordance with the standards set by prominent survey research organizations (The American Association for Public Opinion Research, 2016). To calculate this, we first estimated the percent of marriages known to be ineligible (i.e., the percent of people who responded but who were not eligible to participate). In total the proportion of known marriages that were ineligible for participation was .48. If we assume that the proportion of ineligibles among those who either refused or did not respond (the unknowns) was similar, then there were an estimated 5,147 couples who were ineligible for our survey. When subtracting these out from the original 11,889, we get an estimated total response rate among eligible households of $32.43\% = \frac{2{,}187}{11{,}889-5{,}147}$ (AAPOR RR4).

Table 1 provides an overview of the number of marriage records we requested, received, and selected for inclusion in the study, along with the number of couples that ultimately participated in the study, broken down by state. Unsurprisingly, the highest response rate of any state came from Utah, where our university is located. However, there is no distinct pattern of response patterns thereafter, as Oregon, Iowa, South Carolina, and Connecticut also reported high (raw) response rates, whereas the lowest response rates came from New York, Nevada, New Jersey, Hawaii, and Florida.

Weighting the data. The sampling design used a self-weighting probability-proportional-to-size (PPS) design within the main sample and the minority oversample. This simplified the weighting for the sample design (design weight) as each respondent was self-weighted. The only design adjustment necessary was to balance the main and minority strata. Persons sampled in the minority county stratum were selected at twice the rate as those in the main sample. These normally would be balanced by assigning a design weight of 1 to those in the main county sample and a weight of 0.5 for those in the minority county oversample. However, because the response rate was lower in the minority county sample than in the main sample, it was necessary to adjust the weight for the oversample to account for the difference in response rates. As a result, we used a weight of .6 for the oversample to adjust for this response rate difference.

Creating sample weights required four steps: 1) creating a response rate weight which adjusted for variability in the response rate from county to county; 2) creating a design weight which adjusted for extreme values; 3) creating a normalized weight which accounted for both the response rate weight and the design weight; and 4) a raking procedure which adjusted existing sample weights based on population characteristics from the Census Bureau. This multi-step weighting process enabled inferences to the population of married couples in the United States. We detail the creation of each of these weights below.

First, to account for variability in the response rate from county to county, we divided the number of couples in which at least one member of the marital dyad completed the survey by the number of marriages selected for contact in the county. To reduce bias, we subtracted the ineligible marriages from the sampled marriages, and used this as the denominator in the response rate equation. The actual response weight was then calculated by taking its inverse. For example, if only ½ of the sampled marriages completed the survey, then the weight would be 2.

Second, an adjustment was made to account for extreme values, such as counties with a very small number of responding couples. For example, counties with a very small number of responding couples would yield a very large weight which will increase the design effect substantially. After examining the distribution of the proportion of marriage records that yielded a complete interview, we added .20 to the proportion responding. This eliminated several high response weights due to the low proportion responding in some counties.

Third, we created a normalized weight so that the weighted N was the same as the unweighted N. To create the final normalized weight, we multiplied the design weight and response rate together using the following calculations:

$$Design\ Weight = 1\ if\ county\ in\ main\ sample$$

$$Design\ Weight = .6\ if\ county\ in\ minority\ oversample$$

Response Weight

$$Response\ Weight = 1 \Big/ \frac{participating\ couple}{sampled\ couples - ineligible\ couples}$$

$$Initial\ Final\ Weight\ (IFweight) = Design\ Weight \times Response\ Weight$$

$$Final\ Weight\ (normalized; Fweight) = IFweight \times \frac{N\ of\ unweighted\ respondents}{N\ of\ weighted\ respondents}$$

The final step in the process of creating the weights involved a raking procedure at the individual level. To do this, we obtained the population parameters for region, age, education and race/ethnicity and then raked until convergence, meaning the existing sampling weights were adjusted based on population characteristics from the Census Bureau, thereby bringing the survey sample into conformity with the population, enabling inferences to the population of married couples in the United States. This procedure is also known as iterative proportional fittings, sample-balancing, or raking ratio estimation.

Responses to the Current Data Collection Environment

Despite our best efforts to execute what we thought was our well conceptualized data collection plan, we realized early on that we needed to adjust our process to maximize our response rate. Consequently, there were multiple instances that required us to deviate from our expected collection plan. For purposes of discussing these, we separate out steps taken during the first (selecting counties) versus second (sampling couples) stage of sampling.

First Stage

As we contacted each county individually, it became apparent that the process of obtaining marriage certificates from the selected counties would not be as straightforward as we had hoped or as much as public records statutes, which typically include marriage certificates, would suggest. Some counties had marriage records online, some could be requested by email, and some could only be requested by U.S. mail. Others could only be requested by a “walk-in” visit, some came only if copies were paid for or some other fee applied, and some, we were told, were simply not available due to privacy laws. Consequently, for those that could only be obtained via “walk-in” we either sent an alumni, a student, or a project PI to that county office to get the records (we paid alumni and students for their travels expenses, and for the records if there was a cost for copies). This involved several multi-day trips across the country. For example, one faculty member traveled to one county in Montana, where a copy of available records were procured after paying for the copies. Another faculty member traveled to three States in the South and East, picking up copies at each county location.

Unfortunately, county marriage certificates in four states were not provided upon request. In two of those states, records were provided after our research team procured special approval by local IRB boards. In one of those cases, two trips were made to that specific county and city offices as well as many emails and discussions with officials about the appropriate use of the marriage records. In the end, we were able to obtain the requested records.

Counties in one state refused to provide county records despite multiple requests and repeated contacts at various levels of state and county government. Fortunately, we discovered after searching the internet that marriages in one of the selected counties were publicized in a local newspaper. We used this information and incorporated it into the sample.

In another instance, the state was forbidden by statute from providing marriage licenses but individual municipalities were not. Thus, we selected multiple municipalities across the state and submitted a FOIA request, all but one of which resulted in the receipt of marriage records for those municipalities. We sent a student to the municipality that chose to not comply to our FOIA request to make an in person request but this too was denied. Ultimately, we chose to enlist the help of our university lawyers who, after making another FOIA request, were able to obtain the information.

In a small number of instances, the relevant government agency either refused or were unable to comply with our requests for marriage records and in these cases we were forced to resample other counties with similar demographic and family patterns, based on the stratification weights obtained during the sample selection process. Together, these strategies ensured that our original sampling frame of 239 counties matched our final sampling frame of 239 counties, despite the fact that a very small number of counties in the final sampling frame were not included in the original sampling frame.

Second Stage

In the second stage, we focused on contacting and enrolling couples whose marriage certificates we had obtained in the first stage. First, it should be noted, that some marriage certificates provided more information than others. Some records contained only the names of the marrying parties and the date of the document whereas other records contained addresses, phone numbers, and ages of parties involved. Given the complexities of locating people with limited information, we sought to maximize the probability of locating and contacting the correct persons via a US mail invitation to the couple to participate. To do this, we searched online databases, both paid and unpaid. Through these, we were able to locate the majority of married couples. Where available, we also purchased voter registration records for the most recent year. These were used to match potential participants with postal addresses and, where available, email and telephone information. We also scoured publicly available social media sites (e.g., Facebook, Twitter, Instagram) for participants' contact information, to the extent that using such information conformed with prevailing legal statutes.

In some instances, multiple addresses were found for the same respondents/couples. In these cases, we mailed invitation letters to multiple potential addresses. Where possible, we contacted participants via email.

Throughout the data collection time, we contacted respondents through as many means as possible, including, as already noted, postal mail, email, social media, landline/cellular telephone numbers, and even texting. Some were more successful than others, with postal mail and email leading the way.

Discussion

In this paper, we ask a simple question: is it still possible to collect a nationally representative dataset? Given some of the complexities in today's data collection environment and the likelihood of these difficulties not only persisting but perhaps even degenerating, this question is key for collecting high-quality data that are most likely to be representative of the populations and people we wish to study. We have discussed some of the extant literature comparing current response rates and differences in data quality between national (quota/convenience) samples and nationally representative ones. We use the CREATE study as a case study and briefly described the process we undertook to collect these data.

To sum, we collected a nationally representative sample of marriages that occurred primarily during May-December 2014 across the United States using a stratified cluster approach to first select counties, from which we obtained marriage certificates, and then selected couples based on the marriage certificates. We obtained a raw response rate of 18% (AAPOR RR1), an adjust response rate of 20.5%, and a refined response rate of 32.4% (AAPOR RR4).

Perhaps the best answer to the question of whether it is still possible to collect a nationally representative sample in family studies is…maybe. While our response rate is higher than many other high-effort studies (Keeter et al., 2017; Pew Research Center, 2016), lending credibility to the potential generalizability of results, we are also collecting data on a well-established societal institution, that of marriage, for which record keeping is widespread and routine, meaning it is possible, at least in principle, to stitch together a sampling frame of all marriages in the country occurring during the specified time period. In other words, if one is careful, each marriage in the United States had an equal probability (accounting, of course, for our oversamples) of selection. The extent to which this is the case for other disciplinary foci may dictate the success of future data collection efforts.

Even if it is possible to garner a sampling frame, it may be worth asking a second question, namely whether the extra effort to collect these data (we're finishing wave 2 of data collection after nearly five years of work) is worthwhile, especially with the widespread availability of select-in panel vendors, whose panels are designed to mirror the United States population and who are in a much better position to deliver the desired number of respondents to any survey in a much quicker time. In our estimation, the time-resources/data quality tradeoff between national vs. nationally representative samples must account for at least three factors. First, the maturity of the field matters. Research on marriage is by now a national past time among family researchers, and the field has many other rich, even if somewhat aging, nationally representative datasets at their disposal, including the National Survey of Families and Households, the National Longitudinal Surveys of Youth (1979, CNLSY, 97), the National Longitudinal Study of Adolescent Health, and a variety of government-funded surveys. Other fields, such as emerging adulthood, with its focus on relationship formation, has fewer nationally representative samples and others, such as those seeking to examine the relationship between family variables and media outcomes, also have comparatively few opportunities to employ large-scale, national-level data. In such fields, national, rather than nationally representative, samples may prove helpful as a next step. Second, the focus of the survey matters. Many nationally representative surveys focus on issues primarily of interest to demographers, economists, and sociologists and there is less focus on process variables and, at least in some instances, measurement of key variables may be less than optimal, as such large surveys are often forced to reduce measurement accuracy in the interest of obtaining data on a large number of topics. Choices about collecting nationally representative data, to the extent that measurement is of critical importance, may need to account for the focus of existing surveys and their strengths and shortcomings.

The extent to which national vs. nationally representative data provide similar answers to research questions of importance in the field has been studied to a certain extent. Although recent efforts by political scientists (Wang et al., 2015; Zhang et al., 2017) have shown promise on the possibility of properly weighting national samples with very large sample sizes, family scientists rarely have the luxury of partitioning the data into thousands of demographic combinations for the simple reason that family scientists rarely have datasets that exceed 5,000[2]. Whether family scientists are able to capitalize on the possibilities of 'big data' or large-scale but unrepresentative data collection methods may determine the extent to which we can draw population-level inferences from non-representative data. Interestingly, another recent piece by Zagheni and Weber (2015) suggest ways to extract signal from the copious noise of non-representative internet data but that rely on assumptions that, in some cases at least, are strong enough to be impractical for family researchers.

To end on a perhaps cynical note, however, it is worth noting that nationally representative data should hold no virtue by itself. After all, it is merely a means to an end. That end, of course, is to maximize the probability that patterns and trends observed in the data gathered from sample respondents is the same or very similar as the patterns and trends that would have been observed had we spoken with each member of the population of interest (i.e., generalizability). A long history of statistical probability and survey research, as outlined in basic research methods textbooks used in thousands of classrooms, supports the claim that this does in fact maximize this probability. However, nationally representative surveys are still subject to many biases and generalizability, even from well-designed and executed surveys and data collection efforts, is not always as high as we may like it to be; nationally representative data collection may be, to employ a colloquially phrase the 'cream of the crap'. Finding ways to utilize alternative methods may enable us to more accurately capture, analyze, and understanding the dramatic changes influencing contemporary families.

[2] To literally pick a number out of the air

References

Blau, F. D. (1998). Trends in the Well-Being of American Women: 1970-1995. In *Journal of Economic Literature* (pp. 112–165). http://doi.org/10.1093/acprof:oso/9780199665853.003.0011

Brick, J. M., & Williams, D. (2013). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 36–59. http://doi.org/10.1177/0002716212456834

Carolina Population Center. (n.d.). Add Health Questions about Field Work. Retrieved September 22, 2017, from http://www.cpc.unc.edu/projects/addhealth/faqs/aboutfieldwork

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, *73*(4), 641–678. http://doi.org/10.1093/poq/nfp075

Cherlin, A. J. (2010). Demographic trends in the United States: A review of research in the 2000s. *Journal of Marriage and the Family*, *72*(3), 403–419. http://doi.org/10.1111/j.1741-3737.2010.00710.x

Couper, M. P. (2017). New Developments in Survey Data Collection. *Annu. Rev. Sociol*, *43*, 121–45. http://doi.org/10.1146/annurev-soc-060116

Craig, B. M., Hays, R. D., Pickard, A. S., Cella, D., Revicki, D. A., & Reeve, B. B. (2013). Comparison of US Panel Vendors for Online Surveys. *Journal of Medical Internet Research*, *15*(11), 3. Retrieved from http://10.0.8.148/jmir.2903

Czajka, J. L., & Beyler, A. (2016). Declining Response Rates in Federal Surveys: Trends and Implications (Background Paper). *Mathematica Policy Research*, *I*(202). Retrieved from https://www.mathematica-mpr.com/our-publications-and-findings/publications/declining-response-rates-in-federal-surveys-trends-and-implications-background-paper

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, *38*(1), 1–18. http://doi.org/10.1016/j.ssresearch.2008.03.007

Dutwin, D., Loft, J. D., Darling, J., Holbrook, A., Johnson, T., Langley, R. E., … Zukerberg, A. (2014). *Current Knowledge and Considerations Regarding Survey Refusals*. Retrieved from https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/RefusalTF_FINAL090814.pdf

Ellwood, D. T., & Jencks, C. (2004). The uneven spread of single-parent families: What do we know? Where do we look for answers? In K. Neckerman (Ed.), *Social Inequality* (pp. 3–78). New York: Russell Sage Foundation.

Frankel, M. R., McWilliams, H. A., & Spencer, B. D. (1983). *NLSY79 Technical Sampling Report*. Retrieved from https://www.nlsinfo.org/sites/nlsinfo.org/files/attachments/130212/NLSY79 Tech Samp Rpt.pdf

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*, 646–675.

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *169*(3), 439–457. http://doi.org/10.1111/j.1467-985X.2006.00423.x

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189. http://doi.org/10.1093/poq/nfn011

House, J. S. (2015). Toward a new nationally representative household panel survey (NRHPS). *Journal of Economic and Social Measurement*, *40*(1), 449–457. http://doi.org/10.3233/JEM-150405

Keeter, S., Hatley, N., Kennedy, C., & Lau, A. (2017). *What Low Response Rates Mean for Telephone Surveys*. Washington, DC. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/12/2017/05/12154630/RDD-Non-response-Full-Report.pdf

Kohut, A., Keeter, S., Doherty, C., Dimock, M., & Christian, L. (2012). *Assessing the representativeness of public opinion surveys*. *Pew Research Center*. Retrieved from http://www.people-press.org/files/legacy-pdf/Assessing the Representativeness of Public Opinion Surveys.pdf

Kreuter, F. (2013). Facing the Nonresponse Challenge. *The Annals of the American Academy of Political and Social Science*, *645*(1), 23–35. http://doi.org/10.1177/0002716212456815

Lesthaeghe, R. J. (2010). The unfolding story of the second demographic transition. *Population and Develoment Review*, *36*(2), 211–251. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20734551

Link, M. W., Mokdad, A. H., Kulp, D., & Hyon, A. (2006). Has the national do not call registry helped or hurt state-level response rates?: A time series analysis. *Public Opinion Quarterly*, *70*(5), 794–809. http://doi.org/10.1093/poq/nfl030

Manning, W. D. (2015). Family formation processes: Assessing the need for a new nationally representative household panel survey in the United States. *Journal of Economic and Social Measurement*, *40*(1–4), 171–193. http://doi.org/10.3233/JEM-150419

Massey, D. S., & Tourangeau, R. (2013). Where Do We Go from Here? Nonresponse and Social Measurement. *The Annals of the American Academy of Political and Social Science*, *645*(1), 222–236. http://doi.org/10.1177/0002716212464191

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, *75*(2), 249–269. http://doi.org/10.1093/poq/nfr003

Moffitt, R., Schoeni, R. F., Brown, C., Chase-Lansdale, P. L., Couper, M. P., Diez-Roux, A. V., … Seltzer, J. A. (2015). Assessing the need for a new nationally representative household panel survey in the United States. *Journal of Economic and Social Measurement*, *40*(1–4), 1–26. http://doi.org/10.3233/JEM-150411

Moore, W., Pedlow, S., Krishnamurty, P., & Wolter, K. (2000). National Longitudinal Survey of Youth 1997 (NLSY97) Technical Sampling Report, *1997*(November).

National Research Council. (2013). *Nonresponse in Social Science Surveys: A Research Agenda*. (R. Tourangeau & T. J. Plewes, Eds.). Washington, D.C.: The National Academis Press.

Pew Research Center. (2016). *Our Survey Methodology in Detail*. Washington, D.C. Retrieved from http://www.pewresearch.org/methodology/u-s-survey-research/our-survey-methodology-in-detail/

Piketty, T., Saez, E., & Zucman, G. (2016). *Distributional national accounts: Methods and estimates for the United States* (NBER Working Paper Series No. 22945). Cambridge, MA. Retrieved from http://www.nber.org/papers/w22945

Raley, R. K. (2015). The Need for Nationally Representative Longitudinal Data for Addressing Key Questions about Family Change. *Journal of Economic and Social Measurement*, *1*(40), 437–445.

Schoeni, R. F., Stafford, F., Mcgonagle, K. A., & Andreski, P. (2013). Response Rates in National Panel Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 60–87. http://doi.org/10.1177/0002716212456363

Seltzer, J. A. (2015). Intergenerational family support processes from young adulthood through later life: Do we need a new national survey? *Journal of Economic and Social Measurement*, *40*(1–4), 231–247. http://doi.org/10.3233/JEM-150403

The American Association for Public Opinion Research. (2016). *Standard definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition.*

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980–991. http://doi.org/10.1016/j.ijforecast.2014.06.001

Wright, D. (2003). *National Survey of Families and Households Wave 3 Field Report*. Madison, WI.

Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, *36*(1), 13–25. http://doi.org/10.1108/IJM-12-2014-0261

Zhang, B., Mildenberger, M., Howe, P. D., Marlon, J., Rosenthal, S., & Leiserowitz, A. (2017). *Quota Sampling Using Facebook Advertisements Can Generate Nationally Representative Opinion Estimates* (New Faces in Political Methodology Conference). State College, PA. Retrieved from http://qssi.psu.edu/new-faces-papers-2017/zhang-quota-sampling-using-facebook-advertisements

Table 1. Number of Marriage Records Requested and Selected from Each State/County.

States (# of Counties)	Certificates Requested	Certificates Received	Certificates Selected[a]	# of Couples in Study
Alabama (5)	300	630	200	34
Alaska (1)	60	80	40	8
Arizona (5)	540	1726	360	78
Arkansas (5)	300	2602	200	33
California (16)	1800	23237	1200	194
Colorado (4)	240	455	159	37
Connecticut (2)	120	80	80	22
Delaware (1)	60	138	40	6
Florida (12)	1020	673	678	86
Georgia (6)	360	399	240	49
Hawaii (1)	120	598	80	9
Idaho (2)	120	5979	80	15
Illinois (7)	780	6805	520	109
Indiana (4)	240	9723	159	42
Iowa (3)	180	160	120	38
Kansas (3)	180	160	120	19
Kentucky (5)	300	330	200	46
Louisiana (6)	360	308	231	42
Maryland (4)	240	972	160	35
Massachusetts (4)	240	147	147	30
Michigan (8)	480	1115	320	76
Minnesota (5)	300	1412	183	42
Mississippi (4)	240	4242	160	28
Missouri (5)	300	245	200	41
Montana (2)	120	143	51	12
Nevada (3)	840	561	560	59
New Jersey (8)	480	441	320	35
New Mexico (5)	420	280	280	55
New York (13)	1200	800	800	76
North Carolina (13)	780	594	520	111

North Dakota (2)	120	120	80	22
Ohio (8)	480	455	319	83
Oklahoma (3)	180	120	120	26
Oregon (2)	120	120	80	28
Pennsylvania (7)	420	322	280	56
Rhode Island (2)	120	80	80	17
South Carolina (5)	300	392	200	57
Tennessee (8)	540	1012	360	78
Texas (21)	1800	1344	1195	177
Utah (2)	120	80	80	37
Vermont (1)	60	60	40	6
Virginia (8)	480	460	320	68
Washington (5)	300	200	200	39
Wisconsin (3)	180	180	120	26

Note: [a] The number of certificates selected per county were based on a weighted stratification. If the number of certificates selected is less than 40 per county, researchers received an incomplete sample of certificates to use.