FIT FOR PURPOSE IN ACTION: DESIGN, IMPLEMENTATION, AND EVALUATION OF THE NATIONAL INTERNET FLU SURVEY

JILL A. DEVER[*],

RTI International, 701 13th St NW, Suite 750, Washington, DC 20005-3967, USA

ASHLEY AMAYA,

RTI International, 701 13th St NW, Suite 750, Washington, DC 20005-3967, USA

ANUP SRIVASTAV,

National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329, USA and Leidos Inc., 11951 Freedom Drive, Reston, VA 20190, USA

PENG-JUN LU,

National Center for Immunization and Respiratory Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329, USA

JESSICA ROYCROFT,

RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC, 27709-2194, USA

MARSHICA STANLEY,

RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC, 27709-2194, USA

M. CHRISTOPHER STRINGER,

formerly at RTI International, is with the U.S. Census Bureau, 4600 Silver Hill Road, Hillcrest Heights, MD 20746, USA

MICHAEL G. BOSTWICK,

formerly at RTI International, is with Squarespace, 8 Clarkson St, New York, NY 10014, USA

STACIE M. GREBY,

National Center for Immunization and Respiratory Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329, USA

TAMMY A. SANTIBANEZ,

National Center for Immunization and Respiratory Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329, USA

WALTER W. WILLIAMS

[*]Address correspondence to Jill A. Dever, RTI International, 701 13th St. NW, Suite 750, Washington, DC 20005-3967, USA; jdever@rti.org.

National Center for Immunization and Respiratory Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30329, USA

Abstract

Researchers strive to design and implement high-quality surveys to maximize the utility of the data collected. The definitions of quality and usefulness, however, vary from survey to survey and depend on the analytic needs. Survey teams must evaluate the trade-offs of various decisions, such as when results are needed and their required level of precision, in addition to practical constraints like budget, before finalizing the design. Characteristics within the concept of fit for purpose (FfP) can provide the framework for considering the trade-offs. Furthermore, this tool can enable an evaluation of quality for the resulting estimates. Implementation of a FfP framework in this context, however, is not straightforward.

In this article, we provide the reader with a glimpse of a FfP framework in action for obtaining estimates on early season influenza vaccination coverage estimates and on knowledge, attitudes, behaviors, and barriers related to influenza and influenza prevention among civilian noninstitutionalized adults aged 18 years and older in the United States. The result is the National Internet Flu Survey (NIFS), an annual, two-week internet survey sponsored by the US Centers for Disease Control and Prevention. In addition to critical design decisions, we use the established NIFS FfP framework to discuss the quality of the NIFS in meeting the intended objectives. We highlight aspects that work well and other survey traits requiring further evaluation. Differences found in comparing the NIFS to the National Flu Survey, the National Health Interview Survey, and Behavioral Risk Factor Surveillance System are discussed via their respective FfP characteristics. The findings presented here highlight the importance of the FfP framework for designing surveys, defining data quality, and providing a set a metrics used to advertise the intended use of the survey data and results.

Keywords

1. INTRODUCTION

Survey researchers, public health leaders, policy makers, and the like face many challenges to obtain current information on their population and topic of interest. Often, tough decisions must be made to address a series of challenges. Project teams, in both the public and private sectors, weigh the benefits and limitations of one survey design over another, acknowledging that there is no perfect approach. Take as one example, response rates.

Ever-decreasing response rates have been cited in the literature for many years (Williams and Brick 2018). For example, Meyer, Mok, and Sullivan (2015) discuss response rate trends in household surveys, and Tourangeau (2019) discusses those for opt-in web surveys and enrollment in web panels. Keeter, Hatley, Kennedy, and Lau (2017) show the declines in response rates for telephone surveys over the years, a trend that appears to be stabilizing at

relatively low levels. Nonresponse bias affects the resulting estimates if those who respond to the survey request are not a random representation of the target population (see Kalton and Maligalig 1991). Low response rates may increase the likelihood of nonresponse bias in at least some of the survey estimates, though this is not guaranteed (Groves 2006; Groves and Peytcheva 2008; Brick and Tourangeau 2017). If low response rates are not anticipated, decreased precision in the estimates will likely result with fewer than desired responses to analyze. If anticipated, the increased sample release will increase not only the survey budget but also the time needed to collect data from enough respondents to meet the analytic needs. Thus, survey researchers must weigh various options and their consequences during the survey's design phase.

Timeliness is another factor at play. Many researchers require fast access to data to inform decisions and next steps in real time. For some, data even a few months old could produce estimates with limited generalizability and therefore limited impact to address important issues. For example, the US Centers for Disease Control and Prevention (CDC) conducts a short survey early in the influenza season to estimate early season influenza vaccination coverage and to assess knowledge, attitudes, behaviors, and barriers related to influenza and influenza prevention (Lu, Srivastav, Santibanez, Stringer, Bostwick, et al. 2017). Such information is needed to inform official communications on the importance of continued vaccination throughout the influenza season. Timeliness needs also may inform the choice of a sampling methodology and the source for sample members. For example, an area household survey requires months of preparation and planning before an interviewer is able to knock on the first door (Valliant, Dever, and Kreuter 2018). Conversely, capturing survey participants through an opt-in web survey is relatively fast; population estimates from these and other nonprobability sampling designs, however, can have sizable levels of selection and other biases that limit their utility (Valliant and Dever 2011; Tourangeau, Conrad, and Couper 2013; Dutwin and Buskirk 2017; Mercer, Lau, and Kennedy 2018; Valliant 2018).

Many surveys, including those funded by a country's government, must be conducted in what can be less than ideal conditions (e.g., insufficient time or budget), as the data need outweigh the challenges. The challenges, as hinted in the previous discussion, are not independent and may not be addressed individually. Instead, researchers must weigh various design options after prioritizing the survey conditions. One such framework to guide discussions and the ultimate decisions is referred to as "fit for purpose."

Thinking through any framework can be difficult—there are many interdependent decisions and competing needs; however, application is possible. In this article, we outline a fit for purpose (FfP) framework to address practical constraints inherent in surveys (section 2). In section 3, we present one relatively straightforward case study on early season vaccination coverage with the CDC's National Internet Flu Survey (NIFS) to illustrate the use and challenges of FfP in design and evaluation of a survey. In the final section, we summarize key points in using FfP and discuss a few limitations of our case study.

2. FIT FOR PURPOSE FRAMEWORK

The needs and constraints of each survey dictate the design and subsequent analyses. For example, one research team requires population data as quickly as possible to assess the public's misunderstanding of an infectious disease. Another team can only afford a small survey to capture opinions on the likelihood of infection. A third team is primarily concerned about internal validity of an experiment on the effectiveness of educational material on correcting misinformation. Thus, there is no universal survey design that will fit all needs. Each team must define the components for their survey, along with their relative importance to meeting the research objectives within practical constraints. Essentially, they must specify the conditions that fit the purpose of their research.

Dr. Robert Santos (2014) appropriately summarized a FfP framework as an important tool to provide "the balance researchers seek between available resources, the rigor of research design and implementation, and the nature of the insights needed to effectively address the research questions." This framework typically contains many inter-related components. For our purposes, we borrow from the literature on quality, total survey error, and total survey quality (Biemer and Lyberg 2003; Dever and Valliant 2014; Biemer 2016, 2010) with origins beginning with Deming (1944) (see also Groves and Lyberg 2010 and Groves 1989). Specifically, Statistics Canada's Quality Assurance Framework (2017) includes six interconnected components that we use to define the FfP framework, listed in order of importance for our case study:

1. *Timeliness* means that estimates generated from the data are specific to the desired time period and are available when needed. A survey designed to produce population estimates for an official government communication with data collected four weeks prior may well fit the timeliness needs. In contrast, a design requiring months to collect, process, and analyze the data (such as many large-scale government surveys conducted in the United States) could result in estimates that are years out of date.

2. *Accessibility* means that data are available using an appropriate data collection methodology and without undue burden on the participants or survey resources. For example, a sample from an existing probability-based web panel may serve well to assess knowledge and opinions on a recently enacted policy. This panel may not serve well if the survey requires estimates for those without internet access or for a subgroup of internet users not well covered by the panel. Similarly, if a single sampling frame does not contain all subgroups within a target population, then researchers may seek supplemental sources for a multiframe design (Lohr and Raghunathan 2017) or even redefine the population of interest.

3. *Relevance* means that the data meet the specified needs for analyses. Questionnaire design is one critical aspect of this component. Survey questions should be developed or borrowed from other sources to capture the construct of interest (Bradburn, Sudman, and Wansink 2004). Pre-existing data without responses to the key questions naturally would not meet the survey objectives.

Borrowing questions from other sources is critical when comparing estimates from the new survey with those sources to assess quality or adjusting survey weights to population estimates from the sources to reduce bias, as with estimated-control calibration (Dever and Valliant 2016; Valliant and Dever 2018).

4.   *Interpretability* means that the data, estimates, and survey design are easily analyzed and understood within the context of similar research. Said another way, the essential survey conditions and the resulting estimates should align with existing theory. For example, the mode with which responses are obtained should align with the subject matter; as in the case of sensitive questions, the best practice is to provide a self-administered questionnaire without the aid of an interviewer (Tourangeau and Yan 2007; Kreuter, Presser, and Tourangeau 2008). For surveys conducted at set intervals, such as longitudinal or repeated cross-sectional surveys, existing theory may also be related to design consistency across the cycles of the survey to evaluate trends. Thus, consistency of questions, sample source, sampling method, and data collection activities should be considered for surveys conducted over time.

5.   *Accuracy* means data and estimates are aligned with the intended target population within acceptable levels of precision and bias. Coverage bias—or any type of bias that negatively affects who provides data for the survey or the data provided—is an important factor to consider here. Researchers should assess the coverage properties for candidate sampling frames or other sources used to obtain the participants. For example, surveys often underrepresent individuals on the ends of the age spectrum, racial minorities, and other groups generally less likely to be interviewed (Abraham, Maitland, and Bianchi 2006; Sterrett, Malato, Benz, Tompson, and English 2017). If candidate data sources allow access to different target population members, then selecting participants from multiple frames, such as with dual-frame random-digit-dial surveys (DFRDD), may be the FfP approach (Lohr and Raghunathan 2017).

6.   *Coherence* means that estimates are consistent with other sources such as a similar survey of the same target population or associated administrative records. Coherency with other data sources is also referred to as external validity when the external sources are considered a gold standard or in some fashion superior. For example, incoherence with a gold standard might suggest non-negligible biases. Coherence is aided through harmonization with the essential survey conditions— questionnaire items, data collection mode, target population, and the like; misalignment of the survey components before data are collected introduces possible confounders such that differences may be explained by bias, survey conditions such as data collection mode, or their interactions.

3.   CASE STUDY: THE NATIONAL INTERNET FLUSURVEY

In this section, we provide the reader with a glimpse into the decision process of the FfP framework in action to collect information on influenza vaccination. The goals of this survey

are to estimate early season influenza vaccination coverage and assess knowledge, attitudes, behaviors, and barriers (KABBs) related to influenza and influenza prevention among the civilian, noninstitutionalized adult population aged 18 years and older in the United States. Estimates were desired for the full target population and for important domains by age group, race/ethnicity, and presence of one or more medical conditions putting the person at high risk for serious influenza complications (Lu, O'Halloran, Ding, Srivastav, and Williams 2016). The result of the decision process was the creation of a new survey: the National Internet Flu Survey (NIFS).

The NIFS is an annual internet survey sponsored by the CDC and conducted within a two-week period on a random sample from a probability-based internet panel. The CDC uses estimates from the NIFS to inform official communications to health care providers and the public on the importance of continued vaccination throughout the influenza season. The CDC also uses the NIFS to assess changes over time within important population subgroups as an indication for how mid-season and end-of-season influenza vaccination rates and severity of population infections might fare. Lower than anticipated early season rates based on relative differences across subgroups within a season or trends across time could be emphasized in the official communications.

In summary, the CDC needed estimates quickly to be relevant (a critical need as discussed further in this section), but they also required that the survey and resulting data be "good enough" so that policy makers, health care providers, and the public are confident in the results (i.e., face validity). This push-pull is at the heart of FfP for survey design, implementation, and evaluation.

We could have chosen other studies to highlight the use of FfP, but the NIFS is a good case study for several reasons. For example, the NIFS goals and target population discussed here are relatively straightforward. A more complex design would have required additional background material at the expense of the FfP discussion. Second, the topic area within public health is commonplace, thus limiting the need to orient the reader to the subject of influenza vaccination. Third, other surveys collect similar information on influenza vaccination because of its importance; therefore, multiple gold standards are available for comparison.

We begin this case study with a discussion of available data sources that might have met the needs of the CDC to produce early season influenza vaccination estimates. Though no satisfactory data source was identified, this evaluation identified key traits for an ideal survey. Next, we summarize and evaluate the NIFS design decisions linked to the six FfP criteria using three years of NIFS data (2014–2016), where appropriate. We conclude this section with a discussion of how well the NIFS succeeded in meeting the intended FfP.

3.1 Available Sources of Influenza Estimates

The CDC implemented the Rapid Flu Survey (RFS) in November 2010 and again in March 2011 within twenty local areas and a sample of the rest of the United States. The purpose of the RFS was to assess whether local areas could use within-season vaccination coverage and associated KABB estimates to adjust their influenza programs within the season and

to measure change in vaccination coverage at the end of the season. The CDC found the November 2010 estimates of influenza vaccination coverage by age, race/ethnicity, and place of vaccination to be very useful for the National Influenza Vaccination Week (NIVW) campaign. The NIVW, held each year in December, presents material to the public that emphasizes the importance of vaccination throughout the season. Though vaccination coverage estimates were available early in the seasons, the local areas could not react quickly enough to apply needed changes within the influenza season in reaction to the survey results. The RFS was a probability-based DFRDD sample selected from landline and cell phone frames. The CDC conducted the survey again the following year (November 2011 and March 2012), renamed the National Flu Survey (RFS-NFS), with a sample to cover the entire United States, not specific to the local areas (CDC 2012). Because of cost, the CDC sought alternatives for early season vaccination estimates and discontinued the RFS-NFS in 2012.

Historically, influenza vaccination coverage in the United States has been measured through large-scale annual surveys of the US population, such as the National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System (BRFSS). The NHIS is a nationally representative probability-based survey that collects information about the health and health care of the civilian, noninstitutionalized population in the United States continuously throughout a calendar year (NCHS 2017). Data on vaccination status and sociodemographic characteristics are self-reported during a face-to-face interview. The NHIS involves a complex sampling design with stratification, clustering, and multistage sampling. The NHIS questionnaire contains items about receipt of vaccinations recommended for adults asked of (directly or via proxy) one randomly selected adult 18 years of age or older within each family in the household.

The BRFSS is a probability-based DFRDD telephone survey of the noninstitutionalized population aged 18 years of age or older in the United States (NCCDPHP 2013; CDC 2017). State health departments contract out the conduct of the survey, in collaboration with CDC, to collect uniform, state-specific data on self-reported preventive health practices (such as influenza vaccination) and risk behaviors linked to chronic diseases, injuries, and preventable infectious diseases.

Although NHIS and BRFSS could be used to assess early influenza vaccination coverage for a specific season, the data they generate are not available until after the season. Therefore, these estimates were not useful for the NIVW communication strategy that reports on vaccination coverage through early November of the current influenza season. Furthermore, these large-scale surveys were not designed to capture the breadth of information—such as respondents' experiences with the influenza vaccine and reasons for or against getting vaccinated.

3.2 The NIFS Design Decisions

Noting the challenges in using pre-existing data sources for their needs, the CDC decided to create a new survey within a cost-constrained environment. Table 1 contains a brief comparison of the survey components for the available data sources—RFS-NFS, NHIS, and BRFSS—with the NIFS. This summary provides an overview of the NIFS FfP criterion

described and evaluated in the subsequent sections. As shown in table 1, the CDC borrowed characteristics from the pre-existing studies that were FfP.

3.2.1   Timeliness-estimates are available in the time required.—Timeliness was the most important FfP component for the RFS-NFS and thus was retained for the NIFS. As noted in table 1, a design such as the NHIS and BRFSS with their complex sample designs and delays in publishing estimates would not be feasible for the NIVW schedule. To meet the challenge, the NIFS included three key features to maximize the timeliness of results.

First, the yearly NIFS questionnaire was designed to be completed in less than 10 minutes with topics including current influenza vaccination status, reasons for (not) getting vaccinated, and KABBs on influenza and influenza vaccination. RTI's Questionnaire Appraisal System was used to uncover quality issues in the NIFS instrument that could affect respondent comprehension or increase burden (Willis and Lessler 1999).

The NIFS was only offered in English. While this may have introduced bias if non-English speakers are vaccinated at different rates than English speakers, adding a Spanish translation would have delayed the start of data collection. As of 2016, the US Census Bureau estimates that almost 60 percent of Spanish-speaking US residents 5 years of age or older report speaking English "very well" (US Census Bureau 2017). Adjustments for undercoverage of Spanish-speaking adults and other subgroups are discussed further in section 3.2.5.

Second, email invitations were sent to all NIFS sample members to request completion of the questionnaire via the web. The proliferation of mobile technology (smartphones, tablets) and concerns for data quality suggest that surveys be administered through best practices and optimized for multiple devices (Couper 2008). Consequently, the NIFS questionnaire was implemented through software that reformats the screen to accommodate the respondent's detected device, ensuring no delays in their participation. The distribution by device type in 2016 was 65.8 percent for computers, 22.3 percent for smartphones, and 11.5 percent for tablets. Participation with a mobile device grew over the three years evaluated in the case study, with the largest increase in the smartphone usage category (9.3 percentage points).

Third, public health reminders to get an influenza vaccination begin as early as September each year. To capture early season estimates after vaccines were made available (typically by August) and adults had enough time to get a vaccination, the CDC determined late October/early November as an optimal survey window. A two-week data collection period was deemed the most feasible compromise between the availability of vaccines and the time needed to generate materials for the NIVW.

Data collection for the NIFS surveys ran fourteen days, spanning the end of October and the second week of November each year, much like the RFS-NFS. As discussed in more detail later, response rates, estimated nonresponse bias, and precision of key estimates combined to meet the analytic objectives. Estimates were produced within two weeks after the end of data collection.

3.2.2 Accessibility—sample member data are easily obtained at a reasonable cost.—Timely access to the target population of civilian, noninstitutionalized adults in the United States 18 years of age and older was a central need for the NIFS. Therefore, it was important to access a pre-existing frame with information readily available for sampling and with current contact information to facilitate the short data collection window. The project team devised a sampling design described later on to enable the selection of the sample just one week prior to the start of data collection.

The RFS-NFS demonstrated the utility of early season vaccination estimates for the NIVW from data obtained within a two-week data collection window. However, switching from an interviewer-administered telephone mode to a web mode was important for lowering costs. A pre-existing probability-based sampling frame with access to a sufficiently large sample of adults in the United States through an internet panel was needed to estimate national early season vaccination coverage.

The NIFS sampling frame was generated from the GfK KnowledgePanel, a large-scale online panel based on a representative random sample of the US population, that has been in existence for many years. This panel is refreshed periodically with new members recruited from a random sample selected from an address-based sampling (ABS) frame. Address-based sampling frames are noted for having "nearly complete coverage" of the US residential population (Iannacchione 2011; Shook-Sa, Currivan, McMichael, and Iannacchione 2013; Valliant, Hubbard, Lee, and Chang 2014; Harter, Battaglia, Buskirk, Dillman, English, et al. 2016). The NIFS frame was quickly developed from the existing panel with information collected during the recruitment process after excluding ineligible panelists such as adults not proficient in English. Frame sizes ranged from 40,227 to 42,075 adults across three years of the NIFS evaluated in this research.

The NIFS uses a single-stage stratified sample with twelve mutually exclusive design strata defined by the interaction of age and race/ethnicity (table 2). The samples were selected independently each year with probabilities proportional to the panel weights. The panel weights reflect the probability of selection into the panel plus calibration adjustments to the most recent March supplement of the Current Population Survey (CPS), a monthly household survey of the civilian noninstitutionalized population in the United States conducted in English and Spanish to address coverage bias (Kott 2006, 2016; Valliant et al. 2018). Sample size and allocation to strata were calculated to achieve the analytic objectives and minimize the variation in the survey weights that could lower precision. The target number of respondents by year were 4,025 in both 2014 and 2015 and 4,159 in 2016; the change in 2016 was attributed to increased funding. The resulting sample sizes were inflated to account for estimated nonresponse based on historical information from similar studies in the first year and from prior rounds of the NIFS in later years (table 2).

3.2.3 Relevance—estimates are pertinent to the needs at hand.—Relevance is related to two aspects noted here. First, the NIFS estimates needed to represent an early season snapshot of the influenza vaccination coverage within the US adult population and to identify those subgroups with relatively low vaccination rates. Therefore, the CDC identified two important domains for analyses:

1. age group (18–49 years, 50–64 years, and 65 years)

2. race/ethnicity (Hispanic, non-Hispanic [NH] white only, NH black only, and NH other/multiple races).

As noted for the NIFS accessibility trait, we controlled the distribution age and race/ethnicity during sampling. GfK captured this and other information during the KnowledgePanel recruitment process and through periodic updates from panel members.

A third domain of importance included adults 18–64 years of age with medical conditions linked to increased risk of serious influenza complications (e.g., asthma, heart disease, diabetes, anemia; see Lu et al. 2016). The project team excluded the high-risk domain from the design because of concerns that the desired construct and the measure calculated from the frame information were not aligned and because inclusion of this domain could reduce precision of the overall estimates. Additionally, estimates from the 2013–2014 BRFSS indicated that nearly 20 percent of adults 18–64 years of age have a high-risk medical condition, suggesting that enough domain members would be attained without additional stratification (O'Halloran, Lu, Williams, Bridges, and Singleton 2016).

Second, the CDC desired information on KABBs that affect an adult's likelihood of becoming vaccinated in the current influenza season. However, except for the RFS-NFS, other studies that collect information on influenza vaccination did not capture data for KABB estimation. In the short NIFS questionnaire, KABB questions were included to enable estimates on the reasons for (not) getting an influenza vaccination, the likelihood that the respondent will get vaccinated, the influenza vaccine effectiveness and safety, and the chance of getting influenza if not vaccinated.

3.2.4 Interpretability–the resulting estimates are understandable and applicable.—Consistency across years of the survey was one metric deemed to quantify the interpretability and stability of the NIFS estimates. This stability would also allow comparisons across time to evaluate influenza vaccination change overall and by the key subgroups. Additionally, the project team desired an easily reproducible survey design for ease in implementation and analyses of the resulting responses. We discuss consistency of the completion rates and the estimates across three cycles of NIFS.

Completion rate by year.: A completion rate, per the AAPOR Standard Definitions (2016), is a response rate for "a particular survey invitation sent to eligible panel members" and does not (nor can it) account for rates associated with panel recruitment. As shown in table 3, completion rates increased across the years of the survey.

Survey estimates by year.: The primary NIFS question of interest was whether the respondent had been vaccinated by the time of the interview (table 1). Table 4 shows the 95 percent confidence intervals for the estimated early season vaccination rates across year. Though some variations are seen, the relative difference in the rates are comparable across year overall and within subgroup, suggesting stability of the data collection methods. Precision of the estimates are also stable, a topic we revisit under the next FfP criterion.

3.2.5 Accuracy–estimates align with the intended target population.—

Accuracy is also known as mean square error (MSE; Walther and Moore 2005; Lohr 2010). Estimates with a relatively small MSE, calculated as the sum of variance and squared bias, are preferred. Therefore, we evaluated the accuracy of the NIFS estimates by assessing the variance and the bias of the key variable of interest—current year vaccination. We also assessed the weighting methodology because the weights can decrease the precision of the estimates. Moreover, while effective weights may reduce or eliminate bias, poorly constructed weights may add to the observed bias. Consequently, we compared the NIFS survey weights against variants of the weighting methodology to assess their effectiveness given the other constraints (e.g., time and budget) discussed previously.

Precision of estimated vaccination rates.: As discussed previously, table 4 contains the estimated vaccination rates and half-width confidence interval by year and by key population subgroups: age, race/ethnicity, and high-risk medical status. We designed the NIFS to have influenza vaccination coverage estimates with confidence interval half-widths no larger than 5 percentage points. Estimates overall, by age, by high-risk medical condition, and for NH white adults all met this criterion. Results for other race/ethnicity groups were mixed, with the increase in the 2016 sample size having most benefit for the Hispanic and NH black estimates. Confidence interval half-widths for the NH other vaccination estimates remained above 5 percentage points across the three years of the survey. An analysis after the 2014 NIFS suggested that allocating additional sample to the NH other strata could have improved precision in these estimates. However, this change would decrease precision in the overall estimates and those by age group and high-risk medical condition because of, for example, differential weighting. Therefore, we held consistent the relative allocation across the strata from 2014 to 2015.

Additionally, CDC protocols suggest that estimates with a coefficient of variation (CV) —standard error divided by the estimate, also referred to as a relative standard error— greater than 30 percent may be considered unstable and suppressed from publicly available materials. Across the years, all estimates shown in table 4 had a CV no larger than 8.8 percent. This indicates that the precision of the estimates met the NIFS FfP criterion.

Bias in the estimates.: Bias is defined as the difference between the survey estimate and the population value (Lohr 2010; Valliant et al. 2018). Ideally, the CDC is interested in unbiased point estimates, both overall and by subgroup. However, bias was tolerable if the CDC perceived the data to be useful in informing the NIVW communication materials to encourage vaccination within the current season. This made it difficult to assess the FfP design on accuracy since perception was not a fixed construct and could not be quantified. We undertook this evaluation by seeking to inform the CDC of the existing bias to allowing them to determine whether the bias was too large for their needs. Specifically, we approximated estimated bias using two approaches: trends, discussed under the interpretability criterion, and nonresponse bias. Though other threats to accuracy for the NIFS likely exist (e.g., coverage and measurement errors linked to the questionnaire, the respondents, and the mode of data collection), the NIFS design precludes the evaluation of their effects.

Nonresponse bias analyses, conditional on the NIFS sample, were conducted each year of the survey using information known for respondents and nonrespondents. We classify the analyses as conditional because bias associated with recruitment to and maintenance of the KnowledgePanel could not be accounted for in the NIFS nonresponse bias results. The nonresponse bias analyses assess whether any statistically *detectable* difference exists between weighted estimates calculated for the full NIFS sample and the survey respondents, but the results are not definitive. Requiring analytic variables for the full sample means that the analysis is typically limited in scope for general surveys (Groves 2006); however, use of the KnowledgePanel for NIFS afforded a relatively large list of variables.

The KnowledgePanel provided characteristics known to be associated with vaccination status or that are used as important reporting domains for a nonresponse bias analysis: age group, education, race/ethnicity, sex, household composition, housing type, housing ownership status, household income, marital status, metropolitan statistical area status, employment status, internet access, region of residence, and presence of some but not all high-risk medical conditions identified by the CDC. Low socioeconomic status and low education, for example, are classified as barriers to preventive health measures such as influenza vaccination (Kelly, Martin, Kuhn, Cowan, Brayne, et al. 2016). We used the SUDAAN software, Version 11 (RTI 2012) to conduct weighted design-appropriate *t* tests for significant differences at the 0.05 level overall by the four race/ethnicity categories and by the three age groups. A total of 105 tests were conducted with each year of data; a Bonferroni (Korn and Graubard 1990) or some other adjustment for multiple testing was not applied to maintain a conservative evaluation of bias. Across the years, several common statistically significant patterns emerged. Respondents were less prevalent in the lowest and highest income categories, lower education categories, and NH other and NH black race/ethnicity groups. There were no significant levels of bias found in the age groups.

While these analyses suggest some level of bias in the estimates, the CDC decided that they still held face validity to inform vaccination messaging for the influenza season as evidenced in the use of the data in the NIVW press releases and continuation of the survey through at least 2018. Moreover, funds were available for additional sample purchase with the 2016 NIFS. An evaluation of precision and nonresponse bias in the 2014 and 2015 estimates suggested additional sample could reduce MSE for adults 50 years of age and older in categories other than NH white. The 2016 NIFS design yielded 4,305 respondents (table 5). In addition to improving precision, the number of significant differences within the nonresponse bias evaluation was less in 2016 compared with 2015 (7.6 percent versus 11.9 percent).

<u>NIFS weighting methodology.</u>: Upon completion of data collection, the NIFS base weights (inverse inclusion probabilities) were calibrated to CPS March supplement totals for sex, age group, race/ethnicity, education, household income, census region, and an indicator of metropolitan area among adults 18 years of age or older. This adjustment was intended to correct nonresponse and coverage bias (e.g., Spanish-speaking adults) in the data and to align the estimates with the intended target population.

We conducted an evaluation of the weighting approach applied in all three years of the survey using the 2015 data to determine if the current methodology was FfP. New weight calibration models with interaction terms that more closely aligned with the BRFSS and NHIS studies were implemented. Estimates using the revised weights showed small decreases in the vaccination rates and in the nonresponse bias for race/ethnicities other than NH white and slight improvements in the precision for vaccination rates by the age and race/ethnicity groups. Based on the limited improvements found through the 2015 reweighting analysis, a 2016 reweighting analysis was not conducted.

3.2.6 Coherence–estimates are consistent with external sources.—Credibility of survey estimates are heightened when they are consistent with other sources. Consistency found in the common items across the sources may also lend credibility to other estimates that are unique to a single survey. The NIFS questionnaire was designed to collect information on the same constructs included in the NHIS and BRFSS. To this end, we compared NIFS vaccination coverage estimates to those generated from NHIS and BRFSS.

National Health Interview Survey estimates.: Publicly available NHIS data include interview month but not date. We generated the NHIS estimated proportion of adults vaccinated within two periods for NHIS—July through October and July through November—that correspond with the NIFS's item (table 1) and the two-week data collection period. The true comparative NHIS estimates are likely to be somewhere within this range, as discussed later on.

For the "October" NHIS estimates, only interviews conducted in November were retained for the analysis. The survey weights were adjusted to the original twelve-month overall weight sum so that the one-month sample would reflect the original target population. Kaplan-Meier (KM) weighted estimates were produced with the KAPMEIER procedure in SUDAAN (RTI 2012), where the event was equal to a self-reported influenza vaccination sometime between July and October. The same procedure was repeated with the December NHIS interviews to generate "November" estimates using reported influenza vaccination sometime between July and November. With this approach, the KM procedure generated estimates of cumulative coverage rates and confidence intervals at the end of each month. Because NIFS interviews were conducted only in late October through early November, the "October" estimates are likely biased low, while the "November" estimates are likely biased high.

Table 6 compares the 2016 NIFS estimates against the NHIS KM estimates. The NIFS vaccination rates for the subgroups shown are all significantly higher than the 2016 NHIS October estimates. The NIFS vaccination rates were also higher than the November estimates, but significance was attained only for the NH black and hispanic racial groups. As noted previously, the truly comparable estimates from NHIS most likely lies within the range of the overall October and November estimates, such as 29.6 and 38.4. However, this range remains significantly lower for some key subgroups.

Though the rates in table 6 differed, the subgroup with the lowest vaccination rate for each characteristic examined was relatively consistent across NIFS and NHIS, with one

exception: Hispanic adults in the NHIS had the lowest estimated "October" vaccination rate (21.7 percent), a value not meaningfully different from the estimated value for NH black adults (24.3 percent). The vaccination rate was lowest for NH white adults in the 2016 NIFS (39.7 percent); however, the lowest level is not statistically different from the highest level (43.5 percent for Hispanic adults), and the NIFS FfP did not include power to detect differences across subgroups.

<u>Behavioral Risk Factor Surveillance System estimates.:</u> In contrast to NHIS, BRFSS publicly available data include interview date and month. One set of KM estimates was produced with interviews conducted during the same yearly time period as the NIFS. For example, the 2016 BRFSS estimates include respondents interviewed October 29 through November 12, 2016. The event corresponded to a reported vaccination anytime between July 1 and the day prior to the interview date. As before, survey weights for the subset of cases were adjusted to reflect the weighted sums for the full year of data.

Table 7 compares the 2016 NIFS estimates against the BRFSS KM estimates. The majority of NIFS vaccination rates for the subgroups shown are significantly higher than the BRFSS estimates. The exception shown is with the unemployed subgroup, where low precision in estimates from both surveys produces a statistically insignificant difference.

We also note that the BRFSS estimates in table 7 are much lower than the NHIS "November" estimates shown in table 6. Many attributes may explain the differences in these national estimates (see table 1), and both sources could have been labeled a gold standard. Thus, the FfP for the gold standards may assist in choosing the most appropriate survey for a particular comparison.

3.3 Was the NIFS Fit for Purpose?

Surveys are ideally designed with a "fit for purpose" in mind. Ultimately, with the case study discussed here, we must ask the question: was the CDC's intended fit for purpose achieved with the NIFS?

The CDC designed the NIFS to provide early season influenza vaccination rates for civilian, noninstitutionalized adults 18 years of age or older in the United States to develop NIVW materials for communication with health care providers, policy makers, and the public at large. The survey was intended to be less costly than the RFS-NFS, be timelier than either the NHIS or BRFSS, and to include KABBs on influenza and influenza vaccination. However, even though no other source was available to provide the critical timely information to the CDC, was the NIFS fit for its intended purpose?

By all evaluations presented here, NIFS met most of the intended goals. Data were readily obtained from a representative sample of the target population. Influenza vaccination coverage estimates were produced in sufficient time to develop NIVW materials. Precision of the resulting estimates was within an acceptable range for reporting per CDC guidelines and for a survey of this magnitude. Because the sample was drawn from an existing panel, several variables were available to quantify nonresponse bias overall and by subgroup.

Consistency and accuracy were the only FfP criteria warranting further consideration. Consistency across the years of NIFS was confirmed by comparing the estimates overall and by subgroup. Conversely, the lack of consistency with the NHIS and BRFSS may be attributed to conditions that are inherently related to the NIFS's other FfP characteristics: differences in data collection period, nonresponse bias, differences in covered target populations, mode effects, or other threats to accuracy that we were unable to evaluate with this case study.

3.3.1 Data collection period.—Sample members of the NIFS were given at most two weeks to respond, a period that is much shorter than both NHIS and BRFSS. Paradata (West 2011) from the 2014 NIFS, for example, showed that 71.5 percent of NH white sample members responded within the first three days compared with less than 50.0 percent for the other race/ethnicity groups. Owing to the association of race/ethnicity with influenza vaccination and to leverage/saliency theory (Groves, Singer, and Corning 2000), we hypothesize that early responders are more likely to be vaccinated. Though not feasible, a longer NIFS data collection period would have likely captured a higher proportion of unvaccinated adults, thus lowering the estimated coverage rates.

3.3.2 Nonresponse bias.—The patterns of late responders are also consistent with nonresponders. Nonresponse bias analyses for all three years of NIFS indicated that bias remained for certain subgroups historically linked to lower vaccination rates. Higher participation rates among those in the higher socioeconomic groups could have inflated the overall vaccination rates. Note that this existed even after calibrating the weights to the CPS March estimates. Thus, this procedure most likely lowered bias but not fully for all subgroups.

3.3.3 Covered target population.—The conceptual target population for all three surveys align. However, the covered target populations—those accessed via the respective sampling frames—differ (Valliant and Dever 2011). GfK provides equipment and internet access for households without it, thereby expanding the coverage of their estimates and better aligning the three covered populations. However, the NIFS includes only English questionnaires so that non-English speakers were excluded. If this portion of the US population is more likely to remain unvaccinated, such as with elderly Hispanics (Pearson, Zhao, and Ford 2011), then their exclusion might have inflated the vaccination coverage rates for NIFS.

3.3.4 Mode effects.—The NIFS is a self-administered web survey, while BRFSS and NHIS are interviewer-administered computer-assisted telephone interview [CATI] and computer-assisted personal interview [CAPI], respectively. Differing measurement error properties and their effects on estimation have been shown for each of these methods (e.g., Jäckle et al. 2008). Additionally, researchers have demonstrated the interaction of nonresponse and mode on populations of varying demographic characteristics (e.g., Biemer 2001). Leaning on theory but only speculation for the NIFS, interaction of mode and demographic characteristics may explain in part the difference seen in the estimates across the surveys (Keeter, McGeeney, Igielnik, Mercer, and Mathiowetz 2015).

Thus, because the FfP among the three surveys do not fully align, the differences in the estimates should be interpreted with some caution. The characteristics providing possible explanations for the differences are also areas for evaluation in future rounds of the NIFS.

So was the NIFS fit for the purpose of generating early season influenza vaccination rates? The answer is yes. The NIFS fills the gap in knowledge not provided by other sources by producing population estimates with adequate accuracy and face validity for the NIVW. This survey and the estimates are also easily reproducible and each fall within a two-week data collection period.

4. CONCLUSIONS

As shown in our case study, the FfP framework is a powerful resource for researchers in three areas. First, this framework can guide conversations on defining traits of a design to meet their needs. Once this is defined, the researcher may assign their relative importance or even traits that are critical to the success of the project. In our case study, the CDC identified timeliness and cost-effective accessibility as the top two criteria for a new survey of early season influenza vaccination coverage, the National Internet Flu Survey (NIFS).

Second, project teams may use the FfP framework with prioritized criteria to evaluate available data sources for addressing their research questions. Appropriate data may already exist, thus saving what can be a large amount time and money. In another situation, a data source may meet many needs but miss one or more critical traits of the intended survey. In estimating early season influenza vaccination coverage, the CDC found the NHIS and BRFSS useful in providing accurate estimates but only for the prior influenza season.

Third, the FfP framework, established for the researcher's needs, can enable the development of a new survey. Much like a recipe, the specialized framework contains the critical ingredients needed for the desired result—high-quality FfP data obtained from a well-designed survey. The NIFS FfP framework lent itself to the use of an established, probability-based web panel as the sampling frame for timely results, consistency over years of the survey, and other such benefits. This same established framework also may be used to evaluate data, as discussed for the NIFS.

The NIFS provided a real-world case study of an FfP framework in use, not only to design a survey but also to evaluate how well the survey met the intended needs. This case study, however, is not without limitations. For example, bias in the NIFS estimates is itself an estimate without knowledge of the truth; differences in the gold standard estimates from the NHIS and BRFSS also lend to varying measures of bias. Both conditions are a common challenge in the literature (Dever 2019). Additionally, the FfP framework presented here contained six components—timeliness, accessibility, relevance, interpretability, accuracy, and coherence. Owing to its flexibility, researchers could develop their FfP frameworks with additional traits by specializing one or more of the components above or with few traits by eliminating components that are not pertinent. The suggested use, however, would follow our guidance much the same.

Researchers are presented with many paths to obtain data to answer important inquiries of the day. Having too many choices can make the decision process onerous at best; at worst, poor decisions at the design phase without considering all facets may require course corrections during the survey, heavy reliance on adjustments to improve the estimation process, or data that cannot address the intended research hypotheses. The FfP framework can provide an organized approach to limit such corrective measures on the quest for efficient surveys with high-quality results.

References

Abraham K, Maitland A, and Bianchi S (2006), "Nonresponse in the American Time Use Survey: Who is Missing from the Data and How Much Does it Matter?," Public Opinion Quarterly, 70, 676–703.

The American Association for Public Opinion Research (2016), Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys (9th ed), AAPOR [online]. Available at http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf (accessed November 15, 2019).

Biemer PP (2001), "Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing," Journal of Official Statistics, 17, 295–320.

————. (2010), "Total Survey Error: Design, Implementation, and Evaluation," Public Opinion Quarterly, 74, 817–848.

————. (2016), "Chapter 10: Errors and inference," in Big Data and Social Science: A Practical Guide to Methods and Tools, eds. Foster I, Ghani R, Jarmin R, Kreuter F, and Lane J, Boca Raton, FL: CRC Press.

Biemer PP, and Lyberg LE (2003), Introduction to Survey Quality, New Jersey: John Wiley & Sons, Inc.

Bradburn NM, Sudman S, and Wansink B (2004), Asking Questions: The Definitive Guide to Questionnaire Design—for Market Research, Political Polls, and Social and Health Questionnaires (2nd ed.), San Francisco, CA: John Wiley & Sons, Inc.

Brick JM, and Tourangeau R (2017), "Responsive Survey Designs for Reducing Nonresponse Bias," Journal of Official Statistics, 33, 735–752.

Centers for Disease Control and Prevention (2012), "Flu Vaccination Coverage, National Flu Survey" [online]. Available at https://www.cdc.gov/flu/pdf/fluvaxview/national-flu-surveymar2012.pdf (accessed November 15, 2019).

Centers for Disease Control and Prevention (2017), "Behavioral Risk Factor Surveillance System 2016 Summary Data Quality Report" [online], June 29, 2017. Available at https://www.cdc.gov/brfss/annual_data/2016/pdf/2016-sdqr.pdf (accessed November 15, 2019).

Couper MP (2008), Designing Effective Web Surveys, New York, NY: Cambridge.

Deming E (1944), "On Errors in Surveys," American Sociological Review, 9, 359–369.

Dever JA (2019), "Discussion of 'How Errors Cumulate: Two Examples' by Roger Tourangeau," Journal of Survey Statistics and Methodology (in press).

Dever JA, and Valliant R (2014), "Estimation with Non-Probability Surveys and the Question of External Validity," Proceedings of Statistics Canada Symposium 2014, pp. 1–8, Available at http://www.statcan.gc.ca/sites/default/files/media/14288-eng.pdf.

————. (2016), "General Regression Estimation Adjusted for Undercoverage and Estimated Control Totals," Journal of Survey Statistics and Methodology, 4, 289–318.

Dutwin D, and Buskirk TD (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, 213–239.

Groves RM (1989), Survey Errors and Survey Costs, New York: John Wiley & Sons, Inc.

————. (2006), "Nonresponse Rates and Nonresponse Bias in Household Surveys," Public Opinion Quarterly, 70, 646–675.

Groves R, and Peytcheva E (2008), "The Impact of Nonresponse Rates on Nonresponse Bias," Public Opinion Quarterly, 72, 167–189.

Groves RM, and Lyberg L (2010), "Total Survey Error: Past, Present, and Future," Public Opinion Quarterly, 74, 849–879.

Groves RM, Singer E, and Corning A (2000), "Leverage-Saliency Theory of Survey Participation," Public Opinion Quarterly, 64, 299–308. [PubMed: 11114270]

Harter R, Battaglia MP, Buskirk TD, Dillman DA, English N, Fahimi M, Frankel MR, et al. (2016), "Report of the AAPOR Task Force on Address-Based Sampling" [online]. Available at https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx (accessed November 15, 2019).

Jäckle A, Roberts C, and Lynn P (2008), "Assessing the Effect of Data Collection Mode on Measurement." ISER Working Paper Series 2008–08, University of Essex, Institute for Social and Economic Research (ISER), Colchester [online]. Available at https://www.econstor.eu/bitstream/10419/91914/1/2008-08.pdf (accessed November 15, 2019).

Iannacchione VG (2011), "The Changing Role of Address-Based Sampling in Survey Research," Public Opinion Quarterly, 75, 556–575.

Kalton G, and Maligalig DS (1991), "A Comparison of Methods of Weighting Adjustment for Nonresponse," Proceedings of the U.S. Bureau of the Census Annual Research Conference, pp. 409–428.

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 Report, 15 May 2017 [online]. Available at http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/ (accessed November 15, 2019).

Keeter S, McGeeney K, Igielnik R, Mercer A, and Mathiowetz N (2015), "From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls," Pew Research Center Report, 13 May 2015 [online]. Available at https://www.pewresearch.org/methods/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/ (accessed November 15, 2019).

Kelly S, Martin S, Kuhn I, Cowan A, Brayne C, and Lafortune L (2016), "Barriers and Facilitators to the Uptake and Maintenance of Healthy Behaviors by People at Mid-Life: A Rapid Systematic Review," PLoS One, 11, e0145074. [PubMed: 26815199]

Kott PS (2006), "Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors," Survey Methodology, 32, 133–142.

————. (2016), "Calibration Weighting in Survey Sampling," WIREs Computational Statistics, 8, 39–53.

Korn EL, and Graubard BI (1990), "Simultaneous Testing of Regression Coefficients with Complex Survey Data: Use of Bonferroni t Statistics," The American Statistician, 44, 270–276.

Kreuter F, Presser S, and Tourangeau R (2008), "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity," Public Opinion Quarterly, 72, 847–865.

Lohr SL (2010), Sampling: Design and Analysis (2nd ed.), Pacific Grove CA: Duxbury Press.

Lohr SL, and Raghunathan TE (2017), "Combining Survey Data with Other Data Sources," Statistical Science, 32, 293–312.

Lu P-J, Srivastav A, Santibanez TA, Stringer MC, Bostwick M, Dever JA, Kurtz MS, and Williams WW (2017), "Knowledge of Influenza Vaccination Recommendation and Early Vaccination Uptake during the 2015–16 Season among Adults Aged 18 Years – United States," Vaccine, 25, 4746–4354.

Lu PJ, O'Halloran A, Ding H, Srivastav A, and Williams WW (2016), "Uptake of Influenza Vaccination and Missed Opportunities among Adults with High-Risk Conditions, United States, 2013," The American Journal of Medicine, 129, 636.e1–636.e11.

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. Available at https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/ (accessed November 15, 2019).

Meyer BD, Mok WKC, and Sullivan JX (2015), "Household Surveys in Crisis," Journal of Economic Perspectives, 29, 199–226.

National Center for Chronic Disease Prevention and Health Promotion (2013), "The BRFSS (Behavioral Risk Factor Surveillance System) Data User Guide," Atlanta, Georgia, August 15, 2013. NCCDPHP [online]. Available at https://www.cdc.gov/brfss/data_documentation/pdf/UserguideJune2013.pdf (accessed November 15, 2019).

National Center for Health Statistics (2017), "Survey Description, National Health Interview Survey, 2016," Hyattsville, Maryland. NCHS [online]. Available at https://www.cdc.gov/brfss/data_documentation/pdf/UserguideJune2013.pdf (accessed November 15, 2019).

O'Halloran AC, Lu P-J, Williams WW, Bridges CB, and Singleton JA (2016), "Influenza Vaccination Coverage among People with High-Risk Conditions in the U.S," American Journal of Preventive Medicine, 50, e15–e26. [PubMed: 26238603]

Pearson WS, Zhao G, and Ford ES (2011), "An Analysis of Language as a Barrier to Receiving Influenza Vaccinations among an Elderly Hispanic Population in the United States," Advances in Preventive Medicine, 2011, 1.

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

Santos RL (2014), "Presidential Address: Borne of a Renaissance—A Metamorphosis for Our Future," Public Opinion Quarterly, 78, 769–777.

Shook-Sa BE, Currivan DB, McMichael JP, and Iannacchione VG (2013), "Extending the Coverage of Address-Based Sampling Frames: Beyond the USPS Computerized Delivery Sequence File," Public Opinion Quarterly, 77, 994–1005.

Statistics Canada (2017), Statistics Canada's Quality Assurance Framework (3rd ed.). Available at https://www150.statcan.gc.ca/n1/en/pub/12-586-x/12-586-x2017001-eng.pdf?st=ZBVUvHOe (accessed November 15, 2019).

Sterrett D, Malato D, Benz J, Tompson T, and English N (2017), "Assessing Changes in Coverage Bias of Web Surveys in the United States," Public Opinion Quarterly, 81, 338–356.

Tourangeau R (2019), "How Errors Cumulate: Two Examples," Journal of Survey Statistics and Methodology, 10.1093/jssam/smz019.

Tourangeau R, Conrad FG, and Couper MP (2013), The Science of Web Surveys, New York: Oxford University Press.

Tourangeau R, and Yan T (2007), "Sensitive Questions in Surveys," Psychological Bulletin, 133, 859–883. [PubMed: 17723033]

U.S. Census Bureau (2017), "Newsroom, Facts for Features: Hispanic Heritage Month 2017," No. CB17-FF.17. Available at https://www.census.gov/newsroom/facts-for-features/2017/hispanicheritage.html (accessed November 15, 2019).

Valliant R (2018), "Comparing Alternatives for Estimation from Nonprobability Samples," Proceedings from the FCSM Research Conference, 7–9 Mar 2018, Washington, DC.

Valliant R, and Dever JA (2011), "Estimating Propensity Adjustments for Volunteer Web Surveys," Sociological Methods and Research, 40, 105–137.

————. (2018), Survey Weights: A Step-by-Step Guide to Calculation, College Station, TX: StataCorp LLC.

Valliant R, Dever JA, and Kreuter F (2018). Practical Tools for Designing and Weighting Sample Surveys (2nd ed.), New York: Springer.

Valliant R, Hubbard F, Lee S, and Chang C (2014), "Efficient Use of Commercial Lists in U.S. Household Sampling," Journal of Survey Statistics and Methodology, 2, 182–209. [PubMed: 28781972]

Walther BA, and Moore JL (2005), "The Concepts of Bias, Precision and Accuracy, and Their Use in Testing the Performance of Species Richness Estimators, with a Literature Review of Estimator Performance," Ecography, 28, 815–829.

West B (2011), "Paradata in Survey Research," Survey Practice, 4, 1.

Williams D, and Brick JM (2018), "Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort," Journal of Survey Statistics and Methodology, 6, 186–211.

Willis GB, and Lessler JT (1999), "Question Appraisal System BRFSS-QAS: A Guide for Systematically Evaluating Survey Question Wording," Report prepared for CDC/NCCDPHP/ Division of Adult and Community Health Behavioral Surveillance Branch.

Table 1.

Comparison of Characteristics for Different Surveillance Systems on Influenza Vaccination in General US Population

Survey characteristics	2014–2016 NIFS	RFS-NFS	NHIS	BRFSS
Probability sampling frame	Probability-based KnowledgePanel from ABS	Landline and cell (dual-frame) RDD	Area probability	Landline and cell (dual-frame) RDD
Sample design Mode of data collection	Stratified sample Self-administered (CAWI)	Stratified sample Interviewer-administered CATI	Stratified, multistage Interviewer-administered CAPI in the home	Stratified sample Interviewer-administered CATI
Sample weighting	Weighted to national population totals that include both English- and Spanish-speaking adults (discussed below)	Weighted to population counts by geographic area, age, sex, and race/ethnicity	Weighted to population counts by age, sex, and race/ethnicity	Weighted to state-specific population counts by age, sex, and race/ethnicity
Geographic estimates	National-only	National and 20 local areas (RFS); National (NFS)	National-only	State-based surveys combined for national estimates
Data collection schedule	Two weeks only	Two weeks in each of two months	Monthly	Monthly
Most recent weighted completion/response rate [a]	61.1%	31.4% landline and 18.3% cell (CDC 2012)	54.9% [b] (NCHS 2017)	30.7% – 65.0% across states (RR4; CDC 2017)
Interview languages Timeliness of data availability	English only <2 weeks after end of data collection	English, Spanish <2 weeks after end of data collection	English, Spanish <12 months after year-end close	English, Spanish < 3 months after year-end close
Question to ascertain vaccination status	"A flu vaccination can be a shot injected in the arm or a mist sprayed in the nose by a doctor, nurse, pharmacist or other health professional. Since July 1, [year] have you had a flu vaccination?"	"Since July 1st, [year] have you had a flu vaccination? It could have been a shot or a spray or mist in the nose."	"During the past 12 months, have you had a flu vaccination? A flu vaccination is usually given in the fall and protects against influenza for the flu season." If so, "During what month and year did you receive your most recent flu vaccination?" (NCHS 2017)	"During the past 12 months, have you had a flu shot?" If so, which month and year. (CDC 2017)
Questions on KABBs on influenza and influenza prevention	Yes	Yes	No	No

[a] The weighted completion rate for the is a response rate conditioned on the sample selected for the NTFS (AAPOR 2016). An associated unconditional response rate would include rates associated with each recruitment cohort within the KnowledgePanel and is therefore not discussed further. All other rates presented in this row are unconditional response rates with AAPOR formula indicated if known.

[b] Unweighted response rate formula was not documented in NCHS (2017) nor any identified source.

Table 2.

Frame and Sample Counts by Design Stratum and Year: 2014–2016 National Internet Flu Survey

Stratum	Age (in years)	Race/Ethnicity[a]	Frame n[b]	Sample counts[c]		
				2014NIFS	2015 NIFS	2016 NIFS
1	18–49	Hispanic	2,654	1,128	1,008	707
2		NH White	9,567	1,174	1,085	1,045
3		NH Black	1,798	851	792	620
4		NH Other	1,224	665	579	535
5	50–64	Hispanic	1,145	225	209	357
6		NH White	10,086	1,034	894	996
7		NH Black	1,448	327	289	468
8		NH Other	634	198	178	300
9	65	Hispanic	522	80	87	155
10		NH White	10,301	845	798	1,427
11		NH Black	761	124	110	234
12		NH Other	302	120	119	170
Total			**40,442**	**6,771**	**6,148**	**7,014**

[a] The non-Hispanic (NH) other category includes adults who reported their ethnicity as non-Hispanic and their race as American Indian/Alaska Native, Asian, Native Hawaiian, other Pacific Islander, or reported multiple races.

[b] The 2016 NIFS frame counts for eligible KnowledgePanel adults 18 years of age and older who speak English are provided as an example.

[c] The CDC set precision requirements for the NIFS using a 40.0 percent vaccination coverage rate to have a 95 percent confidence interval half-width no larger than 5 percentage points.

Table 3.

Weighted Completion Rates by Subgroup Across Years: 2014–2016 National Internet Flu Survey

Subgroup	2014 Rates (%)		2015 Rates (%)		2016 Rates (%)	
	Unweighted[a]	Weighted[b]	Unweighted[a]	Weighted[b]	Unweighted[a]	Weighted[b]
Overall	49.1	53.1	53.7	57.6	61.4	61.1
Age (in years)						
18–49	39.5	45.6	43.5	49.0	51.4	53.5
50–64	59.3	60.1	65.8	67.3	65.0	67.1
65	64.9	65.7	68.2	69.0	72.2	72.6
Race/ethnicity[c]						
Hispanic	34.5	36.4	37.0	37.6	50.5	48.8
NH white	61.2	60.0	67.3	65.7	70.3	67.5
NH black	38.2	39.3	41.7	43.3	51.8	49.4
NH other	47.3	47.9	51.7	51.3	56.4	55.3

[a] The unweighted rates are calculated as the number of respondents over the number of sample members by row.

[b] The weighted rates are calculated as the weighted number of respondents over the weighted number of sample members by row using the base weight (inverse probability of selection).

[c] The non-Hispanic (NH) other category includes adults who reported their ethnicity as non-Hispanic and their race as American Indian/Alaska Native, Asian, Native Hawaiian, other Pacific Islander, or reported multiple races.

Table 4.

Early Season Influenza Vaccination Rates by Subgroup Across Years: 2014–2016 National Internet Flu Survey

Subgroup	NIFS 95% confidence interval by year		
	2014	**2015**	**2016**
Overall	40.4 ± 1.8	39.9 ± 1.9	40.6 ± 1.7
Age (in years)			
18–9	31.4 ± 2.7	31.8 ± 2.8	34.3 ± 2.7
50–64	43.9 ± 3.2	41.3 ± 3.4	41.7 ± 2.9
65	61.7 ± 3.7	60.4 ± 3.9	56.6 ± 2.7
Race/ethnicity[a]			
Hispanic	36.5 ± 5.2	36.8 ± 5.9	43.5 ± 4.5
NH white	41.3 ± 2.3	41.6 ± 2.3	39.7 ± 2.2
NH black	39.2 ± 4.6	35.5 ± 5.3	40.6 ± 4.4
NH other/multiple races	42.7 ± 6.2	37.9 ± 6.5	43.1 ± 5.5
High-risk medical condition (18–64 years only)[b]			
Yes	43.2 ± 4.1	40.8 ± 4.3	43.5 ± 3.8
No	32.6 ± 2.4	32.7 ± 2.6	34.0 ± 2.4

[a] The non-Hispanic (NH) other category includes adults who reported their ethnicity as non-Hispanic and their race as American Indian/Alaska Native, Asian, Native Hawaiian, other Pacific Islander, or reported multiple races.

[b] High-risk medical conditions include chronic asthma, a lung condition other than asthma, heart disease (other than high blood pressure, heart murmur, or mitral valve prolapse), diabetes, a kidney condition, a liver condition, sickle cell anemia or other anemia, a neurologic or neuromuscular condition, a weakened immune system caused by a chronic illness or by medicines taken for a chronic illness, or obesity. Status was determined from NIFS responses.

Table 5.

Respondent Count by Design Stratum and Year: 2014–2016 National Internet Flu Survey

Stratum	Age (in years)	Race/ethnicity[a]	Respondent counts by year		
			2014 NIFS	**2015 NIFS**	**2016 NIFS**
Total			3,325	3, 301	4,305
1	18–49	Hispanic	328	328	323
2		NH white	651	651	636
3		NH black	254	254	256
4		NH other	275	275	279
5	50–64	Hispanic	118	106	200
6		NH white	647	647	716
7		NH black	169	169	280
8		NH other	124	111	182
9	65	Hispanic	49	49	93
10		NH white	570	570	1,085
11		NH black	74	74	149
12		NH other	66	67	106
	18–49		1,508	1,508	1,494
	50–64		1,058	1,033	1,378
	65		759	760	1,433
		Hispanic	495	483	616
		NH white	1,868	1,868	2,437
		NH black	497	497	685
		NH other	465	453	567
Sample counts			6,771	6,148	7,014

[a]The non-Hispanic (NH) other category includes adults who reported their ethnicity as non-Hispanic and their race as American Indian/Alaska Native, Asian, Native Hawaiian, other Pacific Islander, or reported multiple races.

Table 6.

Comparison of the 2016 National Internet Flu Survey (NIFS) Influenza Vaccination Estimates with the 2016 National Health Interview Survey (NHIS) Kaplan-Meier Influenza Vaccination Estimates

Subgroups[a]	NIFS		NHIS October[b]		NHIS November[c]	
	n	*% (SE)*	*n*	*% (SE)*	*n*	*% (SE)*
Overall	4,305	40.6(0.9)	2,612	29.6(1.1)[*]	2,407	38.4(1.3)
Age (years)						
18–49	1,494	34.3(1.4)	1,275	21.5(1.4)[*]	1,072	27.1(1.7)[*]
50–64	1,378	41.7(1.5)	624	32.0(2.3)[*]	651	43.9(2.5)
65	1,433	56.6(1.4)	713	49.8(2.3)[*]	684	59.9(2.3)
Race/ethnicity						
White, non-Hispanic	2,511	39.7(1.1)	1,814	31.5(1.3)[*]	1,727	43.0(1.6)
Black, non-Hispanic	673	40.6(2.2)	255	24.3(3.4)[*]	232	25.5(3.7)[*]
Hispanic	604	43.5(2.3)	329	21.7(2.7)[*]	275	28.9(3.2)[*]
Other, non-Hispanic	517	43.1(2.8)	214	37.9(4.6)	173	36.3(4.4)
High-risk (HR) medical condition, 18–64						
18–64 year, HR	905	43.5(1.9)	366	35.4(3.1)[*]	382	42.8(3.1)
18–64 year, no HR	1,967	34.0(1.2)	1,530	22.1(1.3)[*]	1,338	30.2(1.6)
Sex						
Male	2,137	39.5(1.3)	1,191	24.6(1.6)[*]	1,090	34.8(1.8)[*]
Female	2,168	41.7(1.2)	1,421	34.2(1.5)[*]	1,317	41.5(1.8)
Employment						
Employed	2,349	37.8(1.2)	1497	25.1(1.4)[*]	1,365	33.7(1.6)[*]
Unemployed	225	32.0(3.8)	83	17.5(5.0)[*]	72	11.5(5.4)[*]
Marital Status						
Married	2,492	42.7(1.1)	1,143	33.6(1.6)[*]	1,134	43.1(1.8)
Never Married	715	35.2(2.2)	630	20.2(2.3)[*]	445	25.2(2.9)[*]
Household Income						
<$35,000	1,040	38.2	893	25.6(1.8)[*]	775	29.1(2.2)[*]
$75,000	1,777	42.4	768	33.7(2.1)[*]	771	43.9(2.3)

NOTE.—%, weighted vaccination coverage estimate; SE, weighted standard error.

[a] A subset of the categories is shown for certain subgroups for brevity.

[b] Influenza vaccination coverage estimate is calculated using the Kaplan-Meier method. For the NHIS November estimates, the calculation includes respondents vaccinated by October 31, 2016 and interviewed during November 2016 (respondents who were vaccinated in November were censored).

[c] Influenza vaccination coverage estimate is calculated using the Kaplan-Meier method. For the NHIS December estimates, the calculation includes respondents vaccinated by November 30, 2016, and interviewed during December 2016.

[*] $p < 0.05$ by t-test (comparing against NIFS).

Table 7.

Comparison of the 2016 National Internet Flu Survey (NIFS) Estimates with the 2016 Behavioral Risk Factor Surveillance System (BRFSS) Kaplan-Meier Estimates

Subgroups[a]	NIFS		BRFSS[b]		NIFS – BRFSS
	n	*% (SE)*	*n*	*% (SE)*	*% (SE)*
Overall	4,305	40.6(0.9)	16,445	26.6(0.8)	14.0(1.2)[*]
Age (years)					
18–49	1,494	34.3(1.4)	5,105	18.4(1.1)	15.9(1.8)[*]
50–64	1,378	41.7(1.5)	4,886	29.7(1.6)	12.0(2.2)[*]
65	1,433	56.6(1.4)	6,454	43.3(1.4)	13.3(2.0)[*]
Race/ethnicity					
White, non-Hispanic	2,511	39.7(1.1)	12,814	28.7(0.9)	11.0(1.4)[*]
Black, non-Hispanic	673	40.6(2.2)	1,209	25.2(3.4)	15.4(4.1)[*]
Hispanic	604	43.5(2.3)	1,159	20.8(2.3)	22.7(3.2)[*]
Other, non-Hispanic	517	43.1(2.8)	980	22.7(2.6)	20.4(3.9)[*]
High-risk (HR) medical condition, 18–64 years only					
18–64 year, HR	905	43.5(1.9)	2,704	30.7(2.2)	12.8(3.0)[*]
18–64 year, no HR	1,967	34.0(1.2)	7,196	19.4(0.9)	14.6(1.5)[*]
Sex					
Male	2,137	39.5(1.3)	6,821	23.7(1.2)	15.8(1.7)[*]
Female	2,168	41.7(1.2)	9,622	29.2(1.0)	12.5(1.6)[*]
Employment					
Employed	2,349	37.8(1.2)	7,653	21.5(0.9)	16.3(1.5)[*]
Unemployed	225	32.0(3.8)	562	21.0(4.4)	11.0(5.8)
Marital Status					
Married	2,492	42.7(1.1)	8627	28.8(1.1)	13.9(1.6)[*]
Never married	715	35.2(2.2)	2,368	20.7(1.8)	14.5(2.8)[*]
Household Income					
<$35,000	1,040	38.2(1.8)	5,296	24.4(1.5)	13.8(2.4)[*]
$75,000	1,777	42.4(1.4)	4,385	27.3(1.3)	15.1(1.9)[*]

NOTE.—% = weighted vaccination coverage estimate; SE, weighted standard error.

[a]A subset of the categories is shown for certain subgroups for brevity.

[b]Influenza vaccination coverage estimate is calculated using the Kaplan-Meier method. For the BRFSS, the calculation includes respondents interviewed October 29 through November 12, 2016, and vaccinated prior to the interview date.

[*] $p < 0.05$ by t-test (comparing against NIFS).