HHS Public Access
Author manuscript
Field methods. Author manuscript; available in PMC 2022 August 02.

Published in final edited form as:
Field methods. 2020 February 18; 32(2): 159–179. doi:10.1177/1525822x20901802.

The Effectiveness of Incentives on Completion Rates, Data Quality, and Nonresponse Bias in a Probability-based Internet Panel Survey

Marshica Stanley[1], **Jessica Roycroft**[1], **Ashley Amaya**[2], **Jill A. Dever**[2], **Anup Srivastav**[3]

[1]RTI International, Durham, NC, USA

[2]RTI International, Washington, DC, USA

[3]Immunization Services Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention and Leidos Inc., Atlanta, GA, USA

Abstract

Previous research has shown that increasing the size of incentives can increase response rates for probability-based, cross-sectional surveys. However, the effects of incentives on web panels have not been extensively studied. We sought to answer the question: What is the effect of larger, postpaid incentives on (1) response, (2) data quality, and (3) nonresponse bias for individuals in a web panel? We analyzed data from the 2015 and 2016 National Internet Flu Survey, a survey that uses the GfK KnowledgePanel® as its sampling frame. We compare panel members who received a postpaid, standard 1,000-point (the equivalent of US$1) incentive in 2015 to panelists who received a larger, 5,000-point (the equivalent of US$5) incentive in 2016. We found that larger incentives were associated with increased interview completion rates with minimal impact on data quality or bias.

Web panels are widely used as a source of survey samples (Baker et al. 2010; Blom et al. 2015; Callegaro et al. 2014:chap. 1). Candidate panel members are recruited through various means such as an address-based probability sample and screened to assess eligibility (English et al. 2018). This screening information, updated on a periodic basis, affords tailored sampling of willing participants from the panel for specific surveys. Vendors use a variety of methods to maintain the viability of the panel such as replacing inactive panelists with new members and limiting the number of surveys within an agreed-on period such as one month (Watson et al. 2018).

Incentives are generally beneficial for improving participation and lowering errors in household surveys (Hsu et al. 2017). Vendors typically provide a standard, nominal incentive for each survey completed not only to maintain panel engagement but also with the goal of obtaining an adequate number of respondents for specified analytic objectives (Callegaro

Corresponding Author: Marshica Stanley, RTI International, 3040 E. Cornwallis Rd., Durham, NC 27709, USA. mstanley@rti.org.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

et al. 2014; Unangst et al. 2019). Researchers have conducted numerous experiments to determine the cost-effective incentive level for their surveys, many finding benefits for a small increase above base levels (Hsu et al. 2017). However, to our knowledge, no evaluation has been made for web panels to determine whether an increase to the standard incentive could improve quality of a survey conducted on a sample of panelists.

In this article, we examine the effects of increasing the standard postpaid incentive for a nationally representative web panel on participation, data quality, and nonresponse bias. A random sample of adult panelists was selected for the National Internet Flu Survey (NIFS). We present results from key survey metrics and population estimates in this evaluation.

Background

Literature largely shows that incentives (both prepaid and postpaid) are effective in increasing response rates for probability-based, cross-sectional surveys (Brick et al. 2007; Halpern et al. 2002; Hsu et al. 2017; Mercer et al. 2015; Singer et al. 1999; Trussell and Lavrakas 2004). Although not definitive, increasing response rates can reduce the likelihood of nonresponse bias in the population estimates (Brick and Tourangeau 2017). Additionally, increasing the size of the incentive further increases response rates for probability-based, cross-sectional surveys (Fox et al. 1988; Singer et al. 1999). Despite the plethora of literature on survey incentives for cross-sectional surveys, the effect of incentives on web panels has not been extensively studied.

There are several reasons that the benefits of incentives[1] when used in web panels may be different than other modes and sampling frames. First, Callegaro et al. (2014:chap. 2) note that most of the literature on survey incentives involved respondents with no prior or limited experience with surveys. Web panelists have experience with surveys, which may build trust and feelings of reciprocity. For panelists with extensive survey experience, Callegaro et al. (2014:chap. 2) surmise that the current impact of postpaid incentives may differ from what the historic postpaid incentive literature suggests.

Second, the potential relative increase on response is often smaller for web panels than other surveys. Because panelists have previously consented to participate in the panel, research suggests that the completion rate on any given survey can be as high as 77.0%, a far departure from the 9% response rates seen on some other studies such as telephone surveys (Baker et al. 2010; Keeter et al. 2017; Vonk et al. 2006). As a result, the pool of individuals who may be converted from refusal to respondent with the use of incentives is relatively small, and the relative effect of the incentive may be diminished. For example, let us assume an incentive can convert 20% of individuals who would otherwise become nonrespondents. If the completion rate is 40%, an incentive could increase it to 52% ($= 100 \times [0.40 + (1 - 0.40) \times 0.20]$), resulting in a relative gain of 30% ($= 100 \times 0.52/0.40 - 100$). However, if the completion rate starts at 60%, an incentive could only increase it to 68%, for a relative gain of 13.3%.

[1] Through the rest of this article, references to incentives should be interpreted as postpaid incentives. Prepaid incentives for web panels are logistically challenging, can significantly increase costs, and are rarely employed on web panels (Callegaro et al. 2014:chap. 2).

Third, web respondents may be unaware of the increased incentive, rendering them ineffective. While some panelists may receive as few as one survey invitation per year, others receive a new request every day. In a comparison study of 20 U.S. web panels, Miller (2008) found that 33.0% of respondents reported taking 10 or more online surveys in the previous 30 days. Moreover, the method of invitation within a given panel is relatively consistent (Callegaro and DiSogra 2008). With few exceptions, panelists receive an e-mail from the same source with standardized instructions and a link to the survey. Panelists may be given a similar time frame to complete each survey and a standard incentive for completion. Given the frequency of invitation and the rarity with which information within the invitation changes, participants may try to save time and effort by skimming the invitation. In the rare event that new information is included in the survey invitation (e.g., a unique incentive offer), panelists who skim may miss this information if it is not explicitly called out with different font, size, or color (Dillman et al. 2014). Moreover, social exchange theory (Dillman et al. 2014) suggests that individuals put in the same amount of effort (i.e., whether to participate) as they receive back (i.e., incentive). If panelists are unaware of an incentive that is larger than the typical amount, they will assume the standard incentive amount and decide how much effort to contribute, negating the potential effect of the larger incentive.

Even if larger incentives for web panelists do increase completion rates, their effect on data quality and nonresponse bias is debatable. Although most research suggests that incentives have no effect on data quality (Shettle and Mooney 1999; Singer and Kulka 2002; Singer et al. 1998; Singer and Ye 2013; Toepoel 2012), at least one study suggested possible data deterioration (Barge and Gehlbach 2012). The risk exists that incentives may reduce data quality by attracting respondents who normally would not have participated in the survey and whose primary goal is to collect the incentive with as little effort as possible (Goritz 2006). Research suggests that such respondents speed through the survey without giving questions much thought (Goritz 2006). Barge and Gehlbach (2012) found that 81.0% of incentivized respondents engaged in at least one form of satisficing behavior (i.e., speeding, item nonresponse, or straightlining), and 41.0% engaged in at least two forms. Satisficing increases measurement error as people are not providing accurate or thoughtful responses. Jäckle and Lynn (2008) found item nonresponse increased by 10–17% in waves of a panel survey in which incentives were used, though the differences were not statistically significant. To the extent that item nonresponse is not missing at random, it may introduce nonresponse bias that is difficult to enumerate and quantify its impact on the overall survey quality.

In this article, we examine the effect of two postpaid incentive amounts on (1) response, (2) data quality, and (3) nonresponse bias of key outcomes collected in the NIFS. We hypothesized that the response among web panel members receiving the *larger* 5,000-point incentive would be similar to those receiving the *standard* 1,000-point incentive. Given the current literature, we did not formulate hypotheses on the effect of incentives on data quality and nonresponse bias and instead only provide an evaluation.

Survey Data

The NIFS sample was drawn from the GfK (2012) KnowledgePanel®, a probability-based Internet panel designed to be representative of the noninstitutionalized U.S. population 18 years and older. The sample was stratified by age-group (18–49 years, 50–64 years, and 65 years or older) and race/ethnicity (Hispanic, non-Hispanic white, non-Hispanic black, and non-Hispanic other), creating 12 strata (Table 1). The samples were selected using a single-stage stratified design meant to oversample minorities and with selection probabilities inversely proportional to the KnowledgePanel® survey weight, a base weight adjusted for nonresponse (Lu et al. 2017).

Data collection occurred October 29–November 11, 2015, and October 27–November 9, 2016, for the 2015 and 2016 NIFS, respectively.[2] In 2015, sampled individuals were invited by e-mail to participate in the NIFS via a unique web link, and all were promised the standard 1,000-postpaid points (the equivalent of US$1) if they completed the survey.[3] For the 2016 NIFS, we identified six strata with lower than average completion rates in years 2014 and 2015 and offered them a larger postpaid incentive of 5,000 points (the equivalent of US$5). Those offered the 5,000-point incentives included: Hispanic 18–49 years, non-Hispanic black 18–49 years, non-Hispanic other/multiple races 18–49 years, Hispanic 50–64 years, non-Hispanic black 50–64 years, and Hispanic 65 years and older. The remaining six strata received the standard 1,000-point postpaid incentive.

Key operational aspects of the NIFS remained constant across years such as the length of the questionnaire, survey topic, information provided in the introduction, and data collection field period. In both 2015 and 2016, respondents were informed that the interview would take about 10 minutes to complete, and reminder e-mails were sent approximately halfway through data collection to all nonrespondents. In 2016, additional reminders were sent to strata with lagging completion rates to meet prespecified targets regardless of their incentive level. A total of 3,301 and 4,305 interviews were completed in 2015 and 2016, respectively (Table 1).

Method

Data for the 2015 and 2016 NIFS were evaluated in three ways: survey participation (response), quality of the respondent data, and bias. We discuss analytic methods in this section. All analyses were conducted in SAS version 7.1 and SUDAAN version 11 with base weights (inverse probability of selection) to account for different sampling rates across strata and years.

Response

We evaluated response using two metrics—completed interview (yes/no) and average number of days from survey request to completion of the questionnaire among respondents[4]

[2]. On review, only 160 individuals across the two years responded twice. Given the small proportion, they were included in the analysis and assumed to be independent observations.
[3]. GfK offers survey-specific incentives when requested by clients; they were unable to quantify the frequency of this request or the average amount of the increased incentive across surveys.

—for strata associated with the incentive increase. The "average number of days" analysis provides insight into whether receiving a larger incentive results in individuals completing the survey in a timelier manner, thus reducing the need for additional reminders. We used a two-sided *t*-test to compare the 2016 metrics across the six strata receiving a 5,000-point incentive to the 2015 metrics across the corresponding strata that received the standard 1,000-point incentive.

The above analyses estimate the net effect of the incentive on the six affected strata. However, we did not randomly assign individuals or strata to the larger incentive group for a true experiment. To assess whether different age/race groups responded differently to the change in incentive and to isolate the effect of the nonexperimental stratum assignment, we also regressed the two metrics on incentive amount, stratum, and their interaction. We used a logit model for the completion rate and linear models to predict average days to complete.

Moreover, any observed significant differences may have been associated with changes over time and not differences in incentive amounts. To isolate the effect of incentive from time, difference-in-difference (DID) models were run on the full 2015 and 2016 combined sample (i.e., 12 strata). In these models, each dependent variable was regressed on year, incentive group (six strata with incentive change across years vs. six strata with no change), and their interaction; a significant interaction indicates that the larger incentive had a unique effect.[5]

Data Quality

Data quality of the respondent data was assessed using three metrics: the proportion of completed interviews with any item nonresponse, the average length of responses to an open-ended question, and the average time in minutes to complete the questionnaire. Completes with item nonresponse were those in which the respondent left one or more of the administered questions blank, refused, or answered as "don't know."[6] Callegaro et al. (2014:chap. 11) note that item-specific nonresponse is a more conventional indicator of low quality, while straightlining or random answer selections are more sophisticated and identifiable using response time.

Length of response was measured as the number of characters in the open-ended question response for the type of doctor's office or place visited (see Appendix for exact question wording). Respondents selected "other specialist or medical place" and a text response limited to 60 characters.

Finally, minutes to complete were calculated as the number of seconds between the start of the interview and the time he or she completed the questionnaire. Values higher than 30,000

4. We also calculated the average days as the difference between the invitation and the day the survey was started but found no substantive changes in our conclusion.

5. Ideally, the models would also control for other covariates of response such as device used to complete the survey. Unfortunately, these additional variables were not included for several reasons. First, some covariates were available only for respondents and thus not relevant for analyses of completion rates. Second, sample sizes were limited, so we did not have the degrees of freedom required to include additional covariates. Finally, additional covariates were not required as the difference-in-difference models controlled for year and any characteristic aligned with year (e.g., number of reminders sent). As a result, we were not able to evaluate additional differences distinct to each year.

6. An alternative measure of item nonresponse would have been to use a count variable—number of missing items per respondent. However, given the low prevalence of missing data (0.53%) and the highly skewed distribution, we opted not to use this approach.

seconds (8.3 hours) were excluded from the analysis. These extraordinarily high values were the result of individuals starting the interview, logging off, and logging back in at a later point in time.[7]

Mimicking the analyses outlined above for response, we used a two-sided *t*-test for comparison of the 2015 and 2016 quality metrics and regression models to assess whether incentives affected data quality similarly across age and race/ethnicity groups. A logit model was used to regress the item nonresponse indicator onto incentive, stratum, and the interaction; linear models were used to regress the length of open-ended response and, separately, the time to complete onto the same independent variables. The bivariate analyses and models included all respondents in the six strata for which the incentive changed across years. Also, similar to the response analyses, DID models were used to isolate the effect of the larger incentive from other changes across years.

Nonresponse Bias

Next, we evaluated nonresponse bias using 17 key items available for the entire NIFS sample.[8] We also used information collected in the NIFS questionnaire: *influenza vaccination* (Yes or intending to/No), *prevalence of a doctor visit in the past year*, and *vaccination in the previous year* (Yes/No). For each item, we limited analyses to respondents in the six strata where the incentive changed between years. To test for incentive effects, we first conducted two-sided *t*-tests to compare full-sample estimates against the respondent-based estimates.

Next, we evaluated differences in demographic metrics against the 2015 American Community Survey (ACS) "gold standard" values with χ^2 tests for the categorical items and two-sided *t*-tests otherwise. Because large demographic shifts in the population do not occur annually, we did not account for change over time for these three comparisons. Unfortunately, a gold standard was not available for the health-related metrics. Instead, we used logit models to regress each health-related measure onto incentive amount, device used to complete the survey, strata, income, sex, and education.

Results

Response

The first section of Table 2 displays the results of the bivariate analyses to assess the effect of different incentives on the response metrics. While the larger incentive yielded a significantly higher completion rate (48.7% vs. 39.7%, $p < .0001$), the 5,000-point incentive was not associated with a change in the number of days to complete the questionnaire (3.7 vs. 3.9 days, $p = .075$).

7. We also ran the analysis including these outliers. While they slightly increased the model variance, their inclusion did not substantively change the results.

8. Variables evaluated were age category, census region, current employment, education, head of household, home own/rent indicator, household income, household size, housing type, marital status, metropolitan statistical area indicator, race/ethnicity, sex, and presence of child within four age groups.

Both response metrics were also regressed on the incentive amount, stratum, and their interaction to identify differential results by stratum (models not shown). The regression model for interview completion failed to identify a significant interaction between incentive amount and strata; the stratum main effect remained significant as with the bivariate results. Thus, individuals in *all* affected strata were significantly more likely to complete the survey when offered the higher incentive.

The model for the number of days to complete the interview identified significant interactions (Figure 1). Having been offered a higher incentive was significantly associated with an increased average number of days among Hispanics and non-Hispanic blacks both aged 18–49 years ($p < .0001$ and $p = .003$, respectively), a decreased average among Hispanics aged 50–64 years ($p = .003$), and Hispanics aged 65 years or older ($p < .0001$). These findings may suggest a significant interaction with age; individuals 18–49 years on average completed the survey in fewer days when offered a larger incentive, while the 5,000-point incentive had the reverse effect on older individuals.

The models that control for strata suggest that the nonrandom selection of strata did not affect completion but may have affected the number of days required to complete the questionnaire. The completion rate changes may be applicable to any population (not just the racial/ethnic and age categories selected to receive the larger incentive), but an experiment is needed to test the generalizability of the days-to-completion findings.

Finally, DID models used to control for change over time showed that the larger incentive accounted for a 7-percentage point increase in the completion rate (Table 3). The same percentage point increase was found when stratum was also controlled (results not shown). Unlike the first two analyses, the DID model suggested that the larger incentive significantly increased the average number of days to complete the questionnaire: Individuals who received the larger incentive took, on average, two-thirds of a day longer to complete the questionnaire. Given the inconsistent findings across the bivariate and multivariate analyses, we ideally would have investigated the interaction effect between strata and incentive. Unfortunately, the DID models did not allow for this type of analysis because of insufficient power and variability in the data. Additional research is required to further test the effect of larger incentives on days to completion.

Returning to our research question—do incentives affect response?—we found that the larger incentive had a consistent and positive impact on the completion rate, increasing it by 7-percentage points once other factors were controlled. While the incentive also increased the average days to complete, the statistical significance of this result varied by analysis and by strata. Regardless, the magnitude of the incentive effect on time was relatively small.

Data Quality

The second half of Table 2 displays the results of the bivariate analyses of proxy indicators for data quality. Among the data quality metrics examined, only item nonresponse was marginally significantly affected by the higher incentive (46.7% vs. 51.2%, $p = .05$); significance was not found with the two multivariate analyses. The length of the open-ended response for the question evaluated failed to reach significance in all three analyses. Only the DID analyses suggested that the larger incentive significantly increased the number of minutes to complete the survey: Respondents who received the larger incentive took on average 3.11 minutes longer to complete the survey. We hypothesize that this increase in time may be a function of social exchange theory or the norm of reciprocity discussed previously (Dillman et al. 2014).

Nonresponse Bias

In the last set of analyses, we evaluated the effect of a larger incentive on nonresponse bias. First, the comparison of full-sample and respondent-based estimates produced 83 test statistics—approximately 10.8% and 12.0% were significant ($p < .05$) for the 2015 and 2016 NIFS data, respectively, with over 86% of the tests in agreement (details not shown). This suggests that detectable levels of nonresponse bias were similar, and there was no sizable influence from the increased incentives.

Table 4 displays estimated outcomes from the ACS and the 2015 and 2016 NIFS, along with three sets of bivariate tests. The first set evaluates differences between the two incentive groups; they are provided only for comparative purposes and are not reflective of bias in one group or the other. The remaining sets (last two columns of Table 4) compare the NIFS estimates against the gold standard ACS values. Compared to the ACS, the 2015 and 2016 NIFS surveys recruited roughly equal proportion of males and females but interviewed more individuals with at least a bachelor's degree; the 2015 NIFS had a slightly higher proportion of those without a high school diploma ($p = .003$, 2015 NIFS; $p = .014$, 2016 NIFS). However, individuals from the 2016 survey had higher incomes on average than individuals from the 2015 survey ($p = .003$), suggesting that the incentive level may have influenced the distribution of the respondents on this metric.

Regarding health outcomes, individuals who were offered the larger incentive were more likely to report having received or planning to receive the influenza vaccination in the 2016 NIFS (53.2% vs. 47.4%, $p = .005$). No significant differences were observed on the other two comparisons. To control for true change over time, we regressed each health outcome on a variety of covariates (models not shown). The results were consistent with the bivariate comparisons found in Table 4. While we may be tempted to conclude that the incentive affects bias of the estimated proportion of adults who were or planned to be vaccinated in the influenza season, we neither can claim generalizability of the findings because of the nonexperimental nature of the study nor determine whether the larger incentive increases or decreases the bias without a comparative gold standard.

Summary and Conclusion

We initially hypothesized that larger incentives would be ineffective at improving response in a web panel due to historically high completion rates, the frequency with which panel members receive survey requests, and the homogeneity of those requests. However, evidence from the NIFS suggests that, like traditional cross-sectional surveys, larger incentives do increase completion rates with minimal effect on time to completion. Specifically, using the DID models, we found that the 5,000-point incentive increased the completion rate by 7-percentage points over the 1,000-point incentive. This is similar to the effect of observed in other cross-sectional, probability-based surveys of the general population (Fox et al. 1988; Singer et al. 1999) and suggests that previous literature regarding the direct correlation between postpaid incentive amount and response may be applicable to web panels.

We also investigated the effect of a larger incentive on data quality and nonresponse bias, observing minimal effects on both. In terms of data quality, respondents receiving the larger incentive took an average of 3.11 minutes more to complete the survey than their 1,000-point counterpart. One variable used to analyze nonresponse bias was significantly affected by the larger incentive—individuals receiving the larger incentive reported higher levels of income, on average, compared to the lower incentive group, making an already skewed income distribution worse. Additionally, differences were observed on the current year's influenza vaccination rate; however, the analyses available could not deduce whether the larger incentive increased or decreased bias. We did not observe any meaningful incentive effect on the indicators of data quality and nonresponse bias. As with the research on response, these findings were relatively consistent with previous literature conducted in other modes and using other sampling frames—while larger incentives may increase the risk for lower data quality and higher nonresponse bias, it appears that the realization of such a risk and/or the magnitude of the effects are small (see, e.g., Singer and Ye 2013).

Limiting consideration to response, data quality, and nonresponse bias and given the goals of the NIFS, we concluded that the larger incentive was superior. This statement and supporting results, however, may not be applicable to all surveys. For the NIFS, the gain in the completion rate was more important than what we considered to be small negative consequences. Another researcher with another set of priorities may draw a different conclusion. We recommend that researchers use these findings to quantify the effect of larger incentives and consider those effects in the context of their unique goals.

Researchers should also consider the applicability of these findings to their own research. Web panels vary significantly (Unangst et al. 2019). Some are probability based, and some are not, and all have different guidelines on the frequency of invitations. To the extent that these features alter the effectiveness of larger incentives, our findings may have limited applicability.

Additionally, we did not have a true experimental design, potentially limiting the generalizability of these findings. All 2015 sampled members were offered a 1,000-point incentive, while in 2016 a nonrandom set of individuals in six affected strata were offered a 5,000-point incentive. The lack of random assignment coupled with the fact that all individuals selected for the larger incentive were racial/ethnic minorities, limited our ability to isolate the effect of incentives from true change or other differences across surveys. The analyses performed sought to minimize this limitation by controlling for strata in the analysis and by using DID models. The lack of large interaction effects by strata suggested that the incentives would have similar effects on all subdomains, minimizing this limitation. The DID models provided similar evidence that the 7-percentage point effect on the completion rate was also robust. However, further research using an experiment should be replicated to confirm (or dispel) our findings.

Finally, this research cannot speak to the overarching relationship between larger incentives and response, data quality, and bias. We only tested two incentive amounts. If the relationship between amount and effect is linear, then researchers may use algebra to identify the effect of any incentive amount. But, if they are not linear (e.g., if a larger incentive has diminishing returns), then additional research using varying amounts will be required.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Appendix

2016 NIFS Survey Instrument

Selected questions

Q1. 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, 2016, have you had a flu vaccination?

- Yes
- No
- Don't know

[IF Q1 = 2 or 3 or refused]

Q6. How likely are you to get a flu vaccination before the end of June 2017?

- Very likely
- Likely
- Unlikely
- Very unlikely

Q14. Since July 1, 2016, have you visited a doctor or other health professional about your own health at a doctor's office, hospital, clinic, or some other place?

- Yes
- No
- Don't know

Q14a. What type of doctor**'**s office or place did you visit? Check all that apply

- Primary care doctor, family doctor, general practitioner, or internal medicine doctor
- OB/GYN
- Urgent care center

- Emergency room
- Inpatient in hospital
- Other specialist or medical place. (SPECIFY): ______[text—allow 60 characters]______
- Don't Know

Q15. During the last flu season did you get a flu vaccination between July 1, 2015, and June 30, 2016?

- Yes
- No
- Don't know

References

Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, and Dillman D, et al. 2010. Research synthesis: AAPOR report on online panels. Public Opinion Quarterly 74:711–81.

Barge S, and Gehlbach H. 2012. Using the theory of satisficing to evaluate the quality of survey data. Research in Higher Education 53:182–200.

Blom AG, Gathmann C, and Krieger U. 2015. Setting up an online panel representative of the general population. Field Methods 27:391–408.

Brick JM, Brick PD, Dipko S, Presser S, Tucker C, and Yuan Y. 2007. Cell phone survey feasibility in the U.S.: Sampling and calling cell numbers versus landline numbers. Public Opinion Quarterly 71:23–39.

Brick JM, and Tourangeau R. 2017. Responsive survey designs for reducing nonresponse bias. Journal of Official Statistics 33:735–52.

Callegaro M, Baker R, Bethlehem J, Goritz AS, Krosnick JA, and Lavraka P, eds. 2014. Online panel research: A data quality perspective. London: John Wiley.

Callegaro M, and Disogra C. 2008. Computing response metrics for online panels. Public Opinion Quarterly 72:1008–32.

Dillman DA, Smyth JD, and Christian JM. 2014. Internet, mail, and mixed-mode surveys: The tailored design method, 4th ed. Hoboken, NJ: John Wiley.

English N, Kennel T, Buskirk T, and Harter R. 2018. The construction, maintenance, and enhancement of address-based sampling frames. Journal of Survey Statistics and Methodology 7:66–92.

Fox RJ, Crask MR, and Jonghoon K. 1988. Mail survey response rate: A meta-analysis of selected techniques for inducing response. Public Opinion Quarterly 52:467.

GfK. 2012. KnowledgePanel® Design Summary [Internet]. http://www.knowledgenetworks.com/ganp/docs/KnowledgePanel(R)-Design-Summary.pdf (accessed May 24, 2017).

Goritz AS 2006. Incentives in web studies: Methodological issues and a review. International Journal of Information Security 1:58–70.

Halpern SD, Ubel PA, Berlin JA, and Asch DA. 2002. Randomized trial of $5 versus $10 monetary incentives, envelope size, and candy to increase physician response rates to mailed questionnaires. Medical Care 40:834–39. [PubMed: 12218773]

Hsu JW, Schmeiser MD, Haggerty C, and Nelson S. 2017. The effect of large monetary incentives on survey completion. Public Opinion Quarterly 81: 736–47.

Jäckle A, and Lynn P. 2008. Respondent incentives in a multi-mode panel survey: Cumulative effects on nonresponse and bias. Survey Methodology 34:105–17.

Keeter S, Hatley N, Kennedy C, and Lau A. 2017. What low response rates mean for telephone surveys. http://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys (accessed May 7, 2019).

Lu Peng-Jun, Srivastav A, Santibanez TA, Stringer CM, 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 35:4346–54. [PubMed: 28676381]

Mercer A, Caporaso A, Cantor D, and Townsend R. 2015. How much gets you how much? Monetary incentives and response rates in household surveys. Public Opinion Quarterly 79:105–29.

Miller J 2008. Burke panel quality R and D. Cincinnati, OH: Burke.

Shettle C, and Mooney G. 1999. Monetary incentives in U.S. government surveys. Journal of Official Statistics 15:231–50.

Singer E, and Kulka RA. 2002. Paying respondents for survey participation. In Studies of welfare populations: Data collection and research issues, eds. Ploeg MV, Moffitt RA, and Citro CF, 105–28. Washington, DC: National Academy Press.

Singer E, van Hoewyk J, Gebler N, Raghunathan T, and McGonagle K. 1999. The effect of incentives on response rates in interviewer-mediated surveys. Journal of Official Statistics 15:217–30.

Singer E, van Hoewyk J, and Maher MP. 1998. Does the payment of incentives create expectation effects? Public Opinion Quarterly 62:152.

Singer E, and Ye C. 2013. The use and effects of incentives in surveys. The Annals of the American Academy of Political and Social Science 645:112–41.

Toepoel V 2012. Effects of incentives in surveys. In Handbook of survey methodology for the social sciences, eds. Gideon L, 209–23. New York: Springer.

Trussell N, and Lavrakas Paul J.. 2004. The influence of incremental increases in token cash incentives on mail survey response: Is there an optimal amount? Public Opinion Quarterly 68:349–67.

Unangst J, Amaya A, Sanders H, Howard J, Ferrell A, Karon S, and Dever JA. 2019. A process for decomposing total survey error in probability and nonprobability surveys: A case study comparing health statistics in U.S. Internet panels. Journal of Survey Statistics and Methodology. doi: 10.1093/jssam/smz040.

Vonk T, van Osenbruggen R, and Williams P. 2006. The effects of panel recruitment and management on research results: A study across 19 online panels. Proceedings of the ESOMAR Panel Research, ESOMAR, Amsterdam, the Netherlands.

Watson N, Leissou E, Guyer H, and Wooden M. 2018. Best practices for panel maintenance and retention. In Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC), eds. Johnson TP, Pennell B-E, Stoop IAL, and Dorer B, 583–96. Hoboken, NJ: John Wiley.

Author Manuscript
Author Manuscript
Author Manuscript

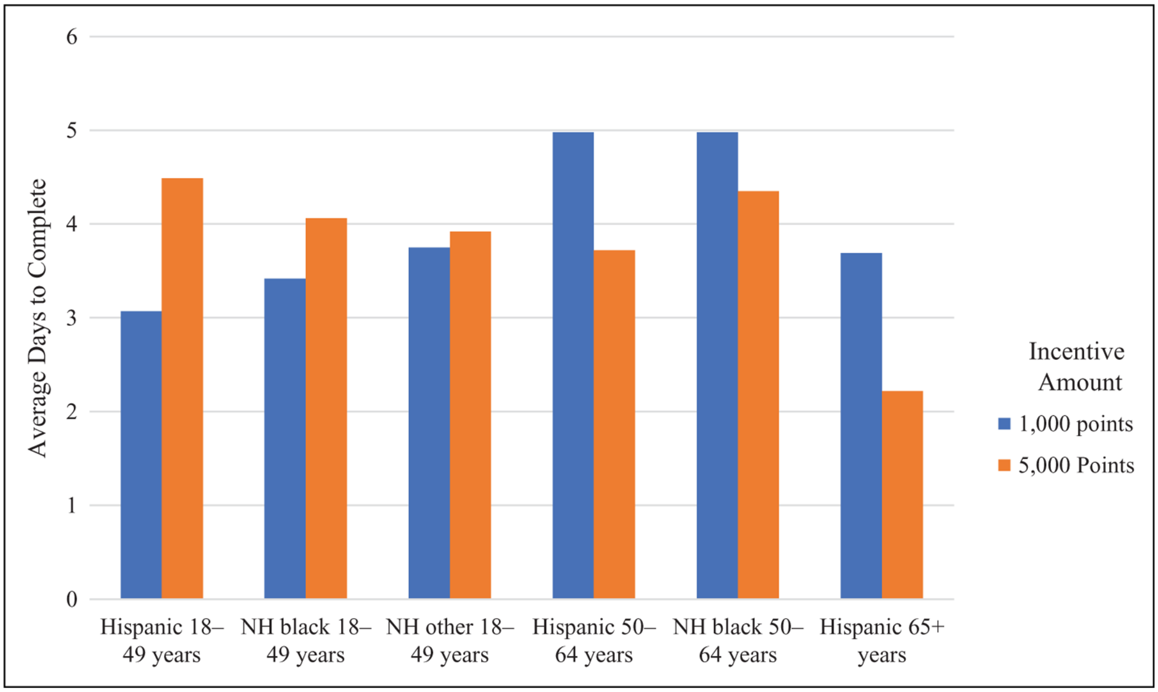

Figure 1.
Average days to complete by strata and incentive received, 2015 and 2016 National Internet Flu Survey.

Table 1.

Completion Rates by Design Stratum and Year, 2015 and 2016 National Internet Flu Survey (NIFS).

		2015		2016	
Age (Years)	**Race/Ethnicity**[a]	**Sampled Individuals (*n*)**	**Completion Rate (%)**[b]	**Sampled Individuals (*n*)**	**Completion Rate (%)**[b]
18–49	Hispanic[a]	1,008	31.8	707	45.0
	NH white	1,085	59.3	1,045	60.3
	NH black[a]	792	31.5	620	41.0
	NH other[a]	579	46.7	535	51.9
50–64	Hispanic[a]	209	50.2	357	55.6
	NH white	894	72.2	996	71.6
	NH black[a]	289	58.3	468	59.4
	NH other	178	61.4	300	60.4
≥65	Hispanic[a]	87	55.2	155	59.7
	NH white	798	71.6	1,427	75.9
	NH black	110	67.1	234	63.6
	NH other	119	56.6	170	62.0
	Total	6,148	57.6	7,014	61.1

Note: NH = non-Hispanic.

[a]In 2016, NIFS, sample members in six strata received a 5,000-point incentive, while the remaining strata received 1,000 points. Sample members in all strata received 1,000 points in 2015 NIFS.

[b]Completion rate is defined as the base-weighted number of respondents divided by the base-weighted number in the sample.

Table 2.

Comparison of Response and Data Quality Metrics by Incentive Received, 2015 and 2016 National Internet Flu Survey (NIFS).

	2015 NIFS (1,000-point Incentive)		2016 NIFS (5,000-point Incentive)		
	N	**Value**	***N***	**Value**	***t*-Statistic (*p* Value)**
Response metrics					
Completion rate	2,964	39.7%	2,842	48.7%	6.70 (<.0001)
Average days to complete	1,181	3.7	1,431	3.9	1.78 (.075)
Data quality metrics					
Any item nonresponse rate	1,181	46.7%	1,431	51.2%	1.96 (.050)
Average length of responses to open-ended question	89	13.1	98	11.4	1.38 (.168)
Average number of minutes to complete the questionnaire	1,139	12.3	1,368	14.9	1.60 (.110)

Note: N = sample count.

Table 3.

Difference-in-difference Models to Isolate the Effect of Incentive from Year—Response and Data Quality Outcomes, 2015 and 2016 National Internet Flu Survey. Standard

	Response Metrics		Data Quality Metrics		
	Completion	**Average Days to Complete**	**Presence of Any Item Nonresponse**	**Average Length Response**	**Average Minutes to Complete**
	Coefficient (*SE*)	**Coefficient (*SE*)**	**Coefficient (*SE*)**	**Coefficient (*SE*)**	**Coefficient (*SE*)**
N	13,162	7,606	7,606	826	7,366
Intercept	0.65*** (0.01)	3.86*** (0.06)	0.44*** (0.01)	14.65*** (0.57)	11.25*** (0.52)
Group (ref. = strata receiving 1,000-point incentive both years)	−0.26*** (0.01)	−0.11 (0.12)	0.10*** (0.02)	−1.42 (1.42)	1.18 (1.14)
Year (ref. = 2015)	0.02*** (0.01)	−0.47*** (0.08)	0.01 (0.01)	−1.21 (0.76)	−0.52 (0.72)
Year × Group	0.07*** (0.02)	0.67*** (0.16)	−0.05 (0.03)	−0.69 (1.91)	3.11* (1.50)

Note: N = sample count; SE = error.

* $p = .05$.

** $p < .01$.

*** $p < .001$.

Table 4.

Distribution of Key Survey Outcomes by Incentive Received, 2015 and 2016 National Internet Flu Survey and 2015 American Community Survey.

	ACS	2015 NIFS (1,000-point Incentive)	2016 NIFS (5,000-point Incentive)	1,000- versus 5,000-point Incentive		ACS versus 1,000-point Incentive		ACS versus 5,000-point Incentive	
Survey Outcome	Percent	Value (%)	Value (%)	% Difference	Test Statistic (*p* Value)	% Difference	Test Statistic (*p* Value)	% Difference	Test Statistic (*p* Value)
Sociodemographic indicators									
Sex (female)	51.0	49.5	51.1	−1.6	0.81 (.421)	1.5	1.01 (.316)	−0.1	0.01 (.921)
Age (in years)									
18–9	76.1	66.7	68.8	−2.1	1.17 (.556)	9.4	2,421 (<.001)	7.3	10,197 (<.001)
50–64	19.0	26.5	24.8	1.7		−7.5		−5.8	
65+	4.9	6.9	6.4	0.5		−2.0		−1.5	
Race/ethnicity									
Hispanic	49.9	44.9	47.8	−2.9	3.30 (.348)	5.0	65,125 (<.001)	2.1	51,451 (<.001)
NH white	0.0	5.1	4.0	1.1		−5.1		−4.0	
NH black	33.1	32.1	31.0	1.1		1.0		2.1	
NH other	17.0	17.9	17.3	0.6		−0.9		−0.3	
Household income									
<$35,000	28.8	40.2	34.7	5.5	13.88 (.003)	−11.4	76.4 (<.001)	−5.9	23.7 (<.001)
$35,000-$49,999	14.3	11.6	13.1	−1.5		2.7		1.2	
$50,000-$74,999	19.5	18.9	17.0	1.9		0.6		2.5	
≥$75,000	37.4	29.3	35.2	−5.9		8.1		2.2	
Education									
Less than high school	21.5	17.8	14.8	3.0	4.02 (.259)	3.7	25.1 (<.001)	6.7	53.2 (<.001)
High school	27.8	29.3	30.3	−1.0		−1.5		−2.5	
Some college	30.6	27.8	29.2	−1.4		2.8		1.4	
Bachelor's degree or higher	20.1	25.1	25.7	−0.6		−5.0		−5.6	
Health outcomes									
Vaccinated or planned to get vaccinated this year	N/A	47.4	53.2	−5.8	2.84 (.005)	N/A	N/A	N/A	N/A
Vaccinated last year	N/A	44.6	44.7	−0.1	0.04 (.970)	N/A	N/A	N/A	N/A

Survey Outcome	ACS	2015 NIFS (1,000-point Incentive)	2016 NIFS (5,000-point Incentive)	1,000- versus 5,000-point Incentive		ACS versus 1,000-point Incentive		ACS versus 5,000-point Incentive	
	Percent	Value (%)	Value (%)	% Difference	Test Statistic (*p* Value)	% Difference	Test Statistic (*p* Value)	% Difference	Test Statistic (*p* Value)
Visited a doctor or health professional in the past year	N/A	58.5	61.3	−2.8	1.38 (.167)	N/A	N/A	N/A	N/A

Note: 1,181 and 1,431 respondent records were analyzed from the 2015 NIFS and 2016 NIFS, respectively. ACS = 2015 American Community Survey; N/A = not available; NIFS = National Internet Flu Survey; NH = non-Hispanic.