

Annual Review of Statistics and Its Application
**Consequences of Asking
 Sensitive Questions in Surveys**

Ting Yan

Westat, Rockville, Maryland 20850, USA; email: tingyan@westat.com

Annu. Rev. Stat. Appl. 2021. 8:109–27

First published as a Review in Advance on
 November 13, 2020

The *Annual Review of Statistics and Its Application* is
 online at statistics.annualreviews.org

<https://doi.org/10.1146/annurev-statistics-040720-033353>

Copyright © 2021 by Annual Reviews.
 All rights reserved

**ANNUAL
 REVIEWS CONNECT**

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

sensitive questions, total survey error, measurement error, unit nonresponse, wave nonresponse, item nonresponse, social desirability

Abstract

I review selected articles from the survey methodology literature on the consequences of asking sensitive questions in censuses and surveys, using a total survey error (TSE) framework. I start with definitions of sensitive questions and move to examination of the impact of including sensitive questions on various sources of survey error—specifically, survey respondents' willingness to participate in a survey (unit nonresponse), their willingness to respond to next rounds of interviews (wave nonresponse), their likelihood to provide an answer to sensitive questions after agreeing to participate in the survey (item nonresponse), and the accuracy of respondents' answers to sensitive questions (measurement error). I also review the simultaneous impact of sensitive questions on multiple sources of error in survey estimates and discuss strategies to mitigate the impact of asking sensitive questions on measurement errors. I conclude with a summary and suggestions for future research.

1. INTRODUCTION

The US decennial Census was in the spotlight for more than a year in advance of launching Census 2020 on March 12, 2020. The Department of Commerce pushed to add a question on citizenship to the census form, evoking heated national debates on the legitimacy and the consequences of asking about citizenship—a very sensitive question. Policy makers, researchers, and laypeople were all concerned about adding the citizenship question, but for different reasons. Although the US Supreme Court's ruling on June 27, 2019, blocked the addition of the question on the 2020 Census, the consequences of asking questions such as this deserve systematic investigation. Indeed, the Census is not the only survey that asks sensitive questions. People face survey requests in their daily lives; the requests come from government agencies, academic researchers, and market researchers. Surveys include sensitive questions of different types and natures, and it is critical for researchers to understand the impact of asking them.

To this end, I review selected articles from the survey methodology literature on the consequences of asking sensitive questions in censuses and surveys, following a total survey error (TSE) framework. The TSE model was first delineated by Groves (1989) and focuses on coverage, sampling, nonresponse, and measurement errors in surveys. Coverage and sampling errors refer to the inability of a survey to mirror the target population the survey tries to study. Unless the inclusion of the sensitive questions changes the definition of the target population and the use of the sampling frame and sampling method, adding sensitive questions to a general population survey is assumed to have a minimal impact on sampling and coverage errors. As a result, this review focuses on the impact of asking sensitive questions on nonresponse and measurement error.

I start with definitions of sensitive questions in Section 2. Section 3 looks into the impact of including sensitive questions on the willingness of respondents to participate in a one-time survey or census—unit nonresponse, in survey methodology's terminology. Section 4 examines the impact of sensitive questions on the willingness of respondents to continue their participation in later rounds of a longitudinal survey (i.e., wave nonresponse), and Section 5 focuses on the impact of sensitive questions on item nonresponse (i.e., the willingness of survey respondents to provide an answer to sensitive questions after agreeing to participate in the survey). The accuracy of respondents' answers to sensitive questions is investigated in Section 6, and Section 7 examines the simultaneous impact of sensitive questions on the multiple sources of errors in a survey. A list of strategies employed by researchers to mitigate the impact of asking sensitive questions on measurement error is described in Section 8; conclusions and suggestions for future research are offered in Section 9.

2. DEFINITIONS OF SENSITIVE QUESTIONS

It is not uncommon for large-scale government-sponsored censuses and surveys to ask about sensitive topics. For example, the National Survey of Family Growth (NSFG) includes questions about Americans' sex life and family planning. The National Survey on Drug Use and Health asks about Americans' use of tobacco, alcohol, and drugs. Questions about citizenship were included on the short form of the decennial Census from 1890 to 1950, the long form of the decennial Census from 1970 to 2000, and then the American Community Survey (ACS) after 2000.

However, what makes a survey question sensitive? The survey literature offers three definitions (Tourangeau et al. 2000, Krumpal 2013, Andreenkova & Javeline 2019). First, questions are sensitive because they are intrusive and out of bounds for ordinary conversations. Questions about sexual behavior and income fall into this category. This type of sensitive question risks offending all respondents, regardless of their answers. The second definition of sensitive questions has to do with threats or risks of disclosing certain information to the survey organization or a third party

involved while answering the sensitive questions. Questions on citizenship are a good example of this type of sensitive question because respondents are concerned about the risk of disclosing their citizenship status on the census form. The third definition involves the concept of social desirability. That is, societal norms prescribe some behaviors and attitudes as socially desirable and others as socially undesirable. For instance, good citizens are expected to vote but not to drive under the influence of alcohol. As a result, questions on voter turnout and driving under the influence of alcohol are prone to social desirability bias—bias toward answering in a socially desirable manner.

Many questions are sensitive for more than one reason. A typical example is questions on criminal justice involvement (Yan & Cantor 2019). The NSFG, for example, asks all male respondents whether they have spent any time in a jail, prison, or juvenile detention facility in the past 12 months. Such questions on criminal justice involvement are sensitive because of the intrusiveness of the topic, negative consequences (even legal concerns) from disclosing the information, and social stigma associated with criminal justice involvement. As a result, sensitive questions are sometimes also called threatening questions, intrusive questions, or private questions. These terms tend to be used interchangeably in the survey literature. Regardless of which term is used to describe and label sensitive questions, such questions subject respondents to emotional and cognitive difficulties (Andreenkova & Javeline 2019).

Sensitivity is not a stable characteristic of survey questions. Instead, it varies by where and how the questions are asked and who answers the questions. Andreenkova & Javeline (2019) show empirically that questions on income and ownership of items indicating household finance, risk behaviors (such as binge drinking and smoking), and health were perceived to be similarly sensitive across the ten countries they examined. By contrast, questions on political behavior and views, family structure, knowledge, and values appeared to be extremely sensitive in some countries but less so in others. Furthermore, Johnson & van de Vijver (2002) conclude that social desirability varies across cultures even within one country; for instance, minority groups in the United States (such as African Americans and Mexican Americans) found questions on alcohol consumption more sensitive when interviewed by an interviewer of a different culture than by an interviewer of their own culture. However, non-Hispanic whites were equally comfortable answering the same questions on alcohol consumption regardless of who interviewed them (Johnson et al. 1997). Kreuter and associates (2008) asked their respondents to rate four survey questions on academic performance on sensitivity. They found that the same items were perceived to be more sensitive when administered by an interviewer over the phone than when administered online or through interactive voice recognition (IVR); interviewers were not involved in the latter two modes of data collection. In addition, respondents with poor academic performance perceived the four items to be more sensitive than those with good academic performance.

To sum up, a survey question is considered sensitive when it is intrusive, incurs negative consequence from disclosure, or invokes social desirability concerns. The level of sensitivity perceived by respondents varies across country, culture, mode, and respondents' true status. The following sections review empirical findings from survey literature on the impact of asking sensitive questions in surveys and censuses.

3. IMPACT ON UNIT NONRESPONSE TO CROSS-SECTIONAL SURVEYS

Cross-sectional surveys invite sampled respondents for one-time participation. Typically, sampled respondents learn about the topic and sponsor of the survey through an advance letter or a brief introduction provided by an interviewer over the phone or at the doorstep. Sampled respondents then have to make a decision on whether or not to participate in the survey based on the limited

amount of information they have. The survey literature provides theoretical frameworks (e.g., Groves & Couper 1998) and empirical evidence (e.g., Groves et al. 2004) on how respondents make a participation decision.

As suggested by the leverage-salience theory (Groves et al. 2000), sample respondents take into consideration a variety of survey attributes when deciding to participate. However, sampled respondents attach different leverage to a survey attribute such as the topic. The more salient a survey attribute is made to a respondent, the more leverage that attribute has on the respondent's participation decision. For instance, if sampled respondents are interested in the topic of a survey, announced in the advance letter, they are more likely to participate than are those who are not interested in that survey topic (see Groves et al. 2004, 2006; Tourangeau et al. 2010).

The survey topic is one of the many survey attributes people consider when making a participatory decision. Potentially, if sampled respondents were made aware that the survey includes sensitive questions, they would be less likely to participate. Using a set of vignettes, Couper and colleagues (2008, 2010) found that survey respondents expressed significantly lower willingness to participate in surveys described as dealing with a sensitive topic (e.g., sex or finance) and surveys with extreme disclosure risks than in surveys with a nonsensitive topic (such as leisure activities) and surveys with fewer disclosure risks. In addition, they found that the return rate of mail surveys decreases as the topic becomes more sensitive; the return rate was 41% when the mail survey was announced to be about leisure activities, 33% for a work related survey, 22% for a survey on sex, and 18% when the survey was about money (Couper et al. 2010).

In addition, for surveys on socially desirable topics such as voting and volunteering, it has been demonstrated empirically that people with the corresponding socially desirable characteristics are more likely to respond than those without. For instance, voters are more likely to participate in surveys about voting and politics than nonvoters (Tourangeau et al. 2010). Similarly, volunteers are more likely to participate in a survey about volunteering than nonvolunteers (Groves et al. 2012).

However, the description of a survey topic tends to be general and brief no matter whether it is communicated in an advance letter or delivered by an interviewer. For instance, the NSFG introduces itself as a survey about family life, marriage and divorce, having and raising children, and health and health care, even though it includes sensitive questions on the number of sexual partners, whether or not a female respondent had an abortion, and whether or not a male respondent was in jail or prison in the past 12 months. As a result, including sensitive questions in a one-time cross-sectional survey does not automatically translate into a low response rate unless respondents are made aware of the presence of the sensitive questions.

It is a somewhat different situation with regard to adding the citizenship question to the US decennial Census. Because of the scope and the political implications of the contentions in 2019, Americans were already aware of the inclusion of the citizenship question on the 2020 Census form, if the US Supreme Court did not rule against its inclusion. Including it could potentially lower response rates. Indeed, Brown and coauthors (2019) estimated that adding a question on citizenship to the 2020 Census could lead to a reduction of self-response rates by eight percentage points for households that may have noncitizens and an overall two percentage point drop in self-response in the 2020 Census.

4. IMPACT ON WAVE NONRESPONSE AND ATTRITION TO LONGITUDINAL SURVEYS

Unlike in cross-sectional surveys, sampled respondents are invited to participate in a longitudinal survey for multiple rounds of interviews. As a result, sampled respondents have a good

understanding of what the survey is about and what kinds of questions are asked after they have completed the first interview. This prior experience with the survey and knowledge of the survey content are expected to affect how respondents behave in later rounds of the interview, a phenomenon called panel conditioning or time-in-sample bias (Warren & Halpern-Manners 2012). Following the same rationale, if sensitive questions are included in a longitudinal survey, respondents would know about it by the end of the first interview. This knowledge might negatively affect their decision to participate in later rounds, if they were offended by the sensitive questions, leading to attrition or wave nonresponse.¹

This speculation is partially borne out in empirical research. A number of investigations have documented higher levels of attrition among heavy alcohol and heavy drug users (see Johnson 2014 for a review). For instance, in a web-based longitudinal study regarding tobacco use, those who reported past 30-day tobacco use had higher odds of not responding to the next wave than those who did not (B. McDonald et al. 2017). In addition, respondents who refused to answer questions on income and finance—a type of sensitive question—in an earlier wave of a longitudinal survey are less likely to respond to the next wave, a trend consistently found in longitudinal surveys conducted in various countries (Loosvedt et al. 2002, Lipps 2007, Uhrig 2008, Watson & Wooden 2009, Régnier-Loilier & Guisse 2012). For example, Loosvedt and associates (2002) found that about 36% of respondents who did not answer the income question refused to participate in the next round of interviews. By contrast, 24% of respondents who provided an answer to the income question did not respond to the next wave, a difference statistically significant at the 0.05 level.

Uhrig (2008) also examined the impact of other types of sensitive questions on wave nonresponse. He found that not providing answers to questions on self-reported health status, mental health problems, traditional family views, and political party increased the likelihood of not responding to the next waves. Furthermore, not providing address information for tracking purposes was also related to wave nonresponse. Watson & Wooden (2009) showed that those who did not return a self-completed questionnaire at the end of an earlier interview were more likely to not participate in later waves (a separate self-completed questionnaire is typically used in face-to-face interviews to collect sensitive information about respondents).

It appears that the presence of sensitive questions offends (at least some) respondents, decreasing their willingness to answer those sensitive questions as well as their likelihood to participate in future rounds of interviews, as predicted and explained by the response continuum model postulated by Yan & Curtin (2010). The model predicts a positive relation between an individual respondent's likelihood to participate in a survey and his or her likelihood to respond to survey items. Its predictions on the joint impact of including sensitive questions on unit and item nonresponse are discussed in Section 7.

5. IMPACT ON ITEM NONRESPONSE

After a potential survey respondent agrees to participate in a survey, he or she may decide not to answer a particular survey item, leading to item nonresponse. Beatty & Herrmann (2002) proposed a theoretical framework to explain item nonresponse. According to their framework, three factors contribute to item nonresponse: (a) respondents' knowledge about the topic, (b) respondents' judgments of the adequacy of what they know relative to the level of exactness or accuracy the question seems to require, and (c) respondents' willingness to report their responses to the survey question. Specifically, respondents are less likely to provide an answer when they truly do

¹I do not make a distinction between attrition and wave nonresponse; instead, I combine both when reviewing the consequences of asking sensitive questions in longitudinal surveys.

not know the answer to the question, when they have a rough estimate but thought the question was asking for an exact estimate, and when they are unwilling to answer the question. The first two factors are cognitive or informational while the third is motivational. With respect to sensitive questions, the motivational factor is presumed to play a bigger role in producing higher item nonresponse because the perceived sensitivity of the items reduces respondents' willingness to provide an answer.

Research has demonstrated empirically that certain types of sensitive questions yield higher item nonresponse. One classic example is income nonresponse (Yan et al. 2010). Questions on personal or household income are considered sensitive because they are widely seen as intrusive. Yan et al. (2010) examined item nonresponse to two types of income questions using 20 years of data from the Surveys of Consumers (SCA). They observed that item nonresponse to the open-ended income question ranged from 11.7% to 26.2% in SCA over the 20-year period, consistent with income nonresponse rates reported in other government-sponsored surveys (see Yan et al. 2010, table 3). Although both cognitive and motivational factors contribute to income nonresponse, Yan and associates produced evidence supporting the bigger impact of the motivational. First, Yan et al. (2010) examined the performance of the unfolding bracket technique in reducing income nonresponse. The unfolding bracket technique asks people to select one of the income brackets offered instead of reporting a dollar amount. It is believed to be effective in reducing cognitive or informational causes of income nonresponse because the use of income brackets conveys to respondents that an exact dollar amount is not needed, lessening the cognitive burden of the request. Yan et al. (2010) found that the use of the unfolding bracket technique produced a reduction of 56% in income nonresponse, resulting in final income nonresponse rates in the range of 5.2% to 15.4%. This is still higher than the 1% to 4% of item nonresponse rates to nonsensitive questions (de Leeuw 1992). Second, Yan et al. (2010) noted that respondents who generated more missing data to other survey items in the same survey (suggesting a low level of motivation), required more call attempts (were difficult to contact), and refused the survey request at least once (showing a higher level of reluctance to participate in the survey) were more likely not to respond to income questions, supporting a relation between low motivation, reluctance, and income nonresponse.

Besides income questions, questions on citizenship (Brown et al. 2019), voting intentions (Krumpal 2013) and number of sexual partners (Tourangeau & Yan 2007) are shown to produce a higher level of missing data than nonsensitive questions such as age (Krumpal 2013, Brown et al. 2019) and education (Krumpal 2013, Tourangeau & Yan 2007). However, item nonresponse to questions on self-involvement with criminal justice is rather low in general population surveys; most of them yield an item nonresponse rate less than 1% (see Yan & Cantor 2019, table 1). By contrast, the question on biological father's involvement with criminal justice had much higher item nonresponse (close to 6%) (see Yan & Cantor 2019, table 1). Given that sensitive questions tend to produce more missing data, item nonresponse is sometimes used by survey researchers to evaluate the perceived sensitivity of survey items (Gnambs & Kaspar 2015, Shoemaker et al. 2002).

6. IMPACT ON MISREPORTING

Another consequence of asking sensitive questions in surveys is manifested as misreporting. After survey respondents have decided to provide an answer to a sensitive question, they tend to misreport to the sensitive question. In other words, their answers do not conform to their true status with respect to that sensitive question, resulting in inaccurate answers, a form of measurement error. Misreporting could be due to respondents not understanding the survey item, not

Table 1 Overreporting of voter turnout in selected general population surveys

Study	Mode	Self-reported voter turnout (%)	Actual voter turnout (%)	Difference (%) ^a	Ratio ^b
Belli et al. (1999)					
Oregon Study	Telephone	72.1	56.0	16.1	1.3
Duff et al. (2007)					
ANES 1994	Face-to-face	55.7	41.1	14.6	1.4
ANES 1996	Face-to-face	71.8	51.8	20.0	1.4
ANES 1998	Face-to-face	52.1	38.4	13.7	1.4
ANES 2002	Face-to-face	64.9	40.2	24.7	1.6
Holbrook & Krosnick (2010)					
Study 1	Telephone	72.0	51.3	20.7	1.4
Study 2	Web	66.1	51.3	14.8	1.3
Study 3	Web	59.5	37.4	22.1	1.6
Study 4	Web	69.9	51.3	18.6	1.4
Kritzing et al. (2012)					
Austria	Web	77.0	78.8	−1.8	1.0
Belgium	Web	87.0	87.7	−0.7	1.0
Denmark	Web	89.0	87.2	1.8	1.0
France	Web	76.0	83.8	−7.8	0.9
Germany	Web	82.0	70.8	11.2	1.2
United Kingdom	Web	79.0	65.1	13.9	1.2
Ireland	Web	76.0	70.1	5.9	1.1
Holbrook & Krosnick (2013)					
Average across states	Face-to-face	Not reported	Not reported	40.3	NA
Zeglovitis & Kritzing (2014)					
Austria	Telephone	85.2	78.8	6.4	1.1
Hanmer et al. (2014)					
United States	Web	Not reported	Not reported	11.8	NA
J.A. McDonald et al. (2017)					
United States	Web	77.0	46.6	30.4	1.7
Kuhn & Vivyan (2018)					
United Kingdom	Web	91.2	66.5	24.7	1.4

^aDifference = self-reported voter turnout – actual voter turnout.

^bRatio = self-reported voter turnout/actual voter turnout.

Abbreviations: ANES, American National Election Studies; NA, not applicable.

memorizing relevant information, failing to integrate recalled information into a judgment or an estimate, or being unable to map their judgment or estimates into one of the response options. However, Tourangeau & Yan (2007) argue that misreporting to sensitive questions largely arises from survey respondents editing their answers prior to reporting them in order to avoid embarrassing themselves. It is a source of response error unique to sensitive questions.

It is well established in the survey literature that survey respondents overreport socially desirable behaviors (such as volunteering, church attendance) and underreport socially undesirable behaviors (e.g., illicit drug use, criminal justice involvement) (see Tourangeau & Yan 2007 for a review). Displayed in **Tables 1** and **2** are evaluations of answers to survey questions on one socially desirable behavior (voter turnout) and one socially undesirable behavior (having had an abortion).

Table 2 Underreporting of having had an abortion in selected general population surveys

Study	Survey	Percentage of actual abortions reported
Fu et al. (1998)	NSFG Cycle 2	45
	NSFG Cycle 3	48
	NSFG Cycle 4	35
	NSFG Cycle 5	45
Jones & Forrest (1992)	NSYW-1976	59
	NSYW-1979	42
	National Longitudinal Surveys of Work Experience of Youth	40
Jagannathan (2001)	Survey on Pregnancy, Childbirth, and Abortion: New Jersey Welfare Clients	29
Jones & Kost (2007)	NSFG 2002	47
Tierney (2019)	National Longitudinal Study of Adolescent to Adult Health	35

Abbreviations: NSFG, National Survey of Family Growth; NSYW, National Surveys of Young Women.

Studies included in the two tables utilize administrative data or external benchmarks to assess the accuracy of survey respondents' answers.

Included in **Table 1** are validation studies that assess survey respondents' answers to a survey item on voter turnout. The survey item used in these studies has two response options (voted versus not voted), and **Table 1** displays the percent of survey respondents who self-reported having voted. Also shown in **Table 1** are either official voter turnout or validated turnout calculated through administrative data (labeled as actual voter turnout in the table), the difference between self-reported and actual voter turnout, and the ratio of self-reported to actual voter turnout. A positive difference and a ratio larger than 1 indicate overreporting of voter turnout; that is, more survey respondents self-reported having voted than did in reality. It is clear from **Table 1** that survey respondents overreported their voter turnout across most surveys, across modes, and across countries. The extent of overreporting ranges from 2% to 50% in these selected studies.

Due to the stigma, abortion is typically underreported by women. Two studies used a sample of women who were known to have had an abortion. Udry and coworkers (1996) found that 19% of these women underreported their abortion history; about half of them denied having had an abortion, and the other half reported fewer abortions. Similarly, Tourangeau et al. (1997a) found that more than a quarter of the women did not report having had an abortion, and only about half reported an abortion during the period in which they were known to have had one (Tourangeau et al. 1997a).

Four studies compared the number of abortions self-reported in various general population surveys to the actual abortions conducted in the United States (the actual numbers are obtained from external surveillance counts). As shown in **Table 2**, abortions were severely underreported; only about 29% to 59% of actual abortions were reported in these surveys. In other words, 40% to 70% of actual abortions were not reported by survey respondents.

The survey literature also indicates that respondents with a socially undesirable characteristic tend to misreport more to questions about that socially undesirable characteristic than those without. For instance, very few people who are shown as US citizens in the administrative data files reported being a noncitizen on the 2016 ACS (less than 0.5%). By contrast, close to 40% of

people who are shown as noncitizens based on administrative records self-reported being a citizen (Brown et al. 2018). Similarly, validation studies confirm that survey respondents underreported arrests (Wyner 1980), delinquent activities that led to police contact (Junger 1989), welfare and unemployment benefit fraud (van der Heijden et al. 2000), and convictions (van der Heijden et al. 2000). But female respondents, older people, people with higher education, people with higher levels of social desirability concerns, and people with more severe offenses underreported criminal justice involvement more often than male respondents, younger respondents, people with lower education, people with lower levels of social desirability concerns, and people with minor offenses, respectively (Wyner 1980, Junger 1989, van der Heijden et al. 2000, Preisendörfer & Wolter 2014, De Jonge 2015). Furthermore, research findings point to increased underreporting to survey questions on substance use as the perceived stigma of the substance being discussed increases (see Johnson 2014 for a review).

7. IMPACT ON MULTIPLE SOURCES OF SURVEY ERROR

Earlier sections review the consequences of asking sensitive questions on individual sources of survey error. However, survey errors are not always independent of each other; instead, they may be additive or cancel each other out. What is more, an effort to reduce one source of error (e.g., unit nonresponse rates) may risk increasing another source of error (e.g., misreporting), leading to an increased TSE. As a result, it is critical to examine multiple sources of survey error at the same time to understand their interplay and to quantify the TSE induced by the inclusion of sensitive items in surveys.

Yan & Curtin (2010) postulate a theoretical link between two sources of nonresponse error: nonresponse to a survey request (i.e., unit nonresponse) and nonresponse to individual survey items (a.k.a. item nonresponse). Their response continuum model predicts a positive relation between an individual respondent's likelihood to participate in a survey and his or her likelihood to respond to survey items. That is, at the individual respondent level, the model posits that respondents who are more likely to not respond to a survey request are also more likely to not answer survey questions. In addition, respondents who have more missing data are also more likely to not respond to the next wave of interviews. When it comes to asking sensitive questions in surveys, it has been shown that respondents who are more reluctant to comply with the survey request are more likely to not answer sensitive questions (Yan et al. 2010), and respondents not responding to sensitive questions are more likely to not respond to the next survey request (e.g., Loosvedt et al. 2002).

Sakshaug and associates (2010) provided empirical support to the response continuum model and quantified the errors arising from unit nonresponse and item nonresponse in a survey of university alumni. Alumni were initially contacted by telephone. After answering a few screening questions, they were randomly assigned to one of three modes for the main interview—telephone, IVR, and web. Regardless of the mode assignment, alumni were told that the survey was sponsored by their university and the survey was about their college experience, interest in alumni activities, and community involvement. The survey included several sensitive questions on academic performance, donation to the university, and alumni membership. **Table 3** displays bias in survey estimates due to unit nonresponse (i.e., bias resulting from not participating in the main interview) and item nonresponse (bias from not responding to the particular sensitive questions).

Three findings are worth noting in **Table 3**. First, biases due to unit and item nonresponse are in the same direction and are additive, supporting the response continuum model. Second, the direction of biases is different for respondents with socially desirable characteristics and respondents without them. Specifically, those with socially undesirable characteristics are less likely

Table 3 Impact of sensitive questions on unit and item nonresponse across all modes

Survey estimates ^a	Unit nonresponse bias ^b	Item nonresponse bias	Total nonresponse bias
Socially undesirable characteristics			
GPA < 2.5	−3.1	−0.1	−3.2
At least one D/F	−1.5	−0.4	−1.9
Dropped a class	−2.2	−0.6	−2.8
Socially desirable characteristics			
GPA > 3.5	2.8	0.6	3.4
Graduated with honors	2.5	0.4	2.9
Ever donated to university	14.7	0.6	15.3
Donated in past year	6.9	0.3	7.2
Alumni member	8.8	0.0	8.8

^aBias estimates are reported in Sakshaug et al. (2010, p. 918, table 2).

^bUnit nonresponse bias is a sum of component bias resulting from noncontact, refusal, and mode switch dropout. Abbreviation: GPA, grade point average.

to participate in the main interview and, once in the survey, are less likely to provide answers to sensitive questions asking about those undesirable characteristics. By contrast, respondents with socially desirable characteristics are more likely to respond to the main interview and are also more likely to respond to sensitive items on those desirable characteristics. Third, the magnitude of unit nonresponse bias is larger than that of item nonresponse bias.

Two studies explored the linkage between nonresponse error and measurement error for surveys with a sensitive topic and estimated the size and direction of each component error. In the same study of university alumni, Sakshaug and associates (2010) compared respondents' answers to the university registrar information for accuracy. Shown in **Table 4** are measurement biases for every survey estimate as well as total nonresponse bias from **Table 3**. It is clear from **Table 4** that survey respondents underreported socially undesirable characteristics (such as having ever dropped a class) and overreported socially desirable characteristics (such as having graduated with

Table 4 Impact of sensitive questions on nonresponse and measurement bias

Survey estimates	Nonresponse bias	Measurement bias	Total bias
Socially undesirable characteristics			
GPA < 2.5 ^a	−3.2	−7.9	−11.1
At least one D/F ^a	−1.9	−15.2	−17.1
Dropped a class ^a	−2.8	−20.2	−23.0
Socially desirable characteristics			
GPA > 3.5 ^a	3.4	1.1	4.5
Graduated with honors ^a	2.9	4.9	7.8
Ever donated to university ^a	15.3	0.8	16.1
Donated in past year ^a	7.2	1.4	8.6
Alumni member ^a	8.8	7.3	16.1
Voted in 2004 ^b	13.2	21.8	35.0
Voted in 2006 ^b	15.9	19.4	35.3

^aBias estimates are taken from Sakshaug et al. (2010, table 2).

^bBias estimates are taken from Tourangeau et al. (2010, table 5).

honors and making donations to the university). In addition, measurement bias and nonresponse bias are in the same direction. That is, respondents with socially undesirable characteristics are less likely to participate in the survey (yielding negative nonresponse bias) and, once persuaded to agree to the survey request, are more likely to underreport their socially undesirable behaviors (leading to negative measurement bias). Similarly, people with socially desirable characteristics are more likely to participate in the survey (resulting in positive nonresponse bias) and are more likely to overreport their socially desirable behaviors as manifested through the positive measurement bias. Furthermore, measurement bias dominates the total bias of estimates on socially undesirable characteristics and contributes about the same amount as nonresponse bias to the total bias of estimates on socially desirable characteristics.

Tourangeau et al. (2010) examined response rates and reporting errors in a sample of Maryland residents who were registered to vote. Among other things, they manipulated the topic of the survey. Half of the sample were told that the survey was about politics, elections, and voting, a sensitive topic for the sample since about half of them did not vote in either the 2004 or the 2006 elections even though they were registered to vote. By contrast, for the other half, the survey was described as a survey on health and lifestyles, a more neutral topic to sampled respondents. Regardless of what survey respondents were told of the topic of the survey, two questions were included asking whether or not they had voted in the 2004 and the 2006 elections.

Tourangeau et al. (2010) found that voters were more likely to respond to the survey of politics, elections, and voting (42%) than nonvoters (24%), yielding positive nonresponse bias for both estimates (the last two rows of **Table 4**). Not surprisingly, respondents overreported their voter turnout for both estimates as well, as signified by the positive measurement biases. Even though the overall measurement bias is slightly larger than nonresponse bias, Tourangeau et al. (2010) found that nonvoters overreported significantly more than voters; overreporting for the 2004 election is 1.5% for voters and 53.1% for nonvoters, whereas overreporting for the 2006 election is 1.6% for voters and 46.5% for nonvoters.

The two studies reviewed in this section point to the positive relation between unit nonresponse error, item nonresponse error, and measurement error as induced by sensitive questions. Respondents with socially undesirable behaviors are less likely to participate in surveys, less likely to provide answers to sensitive questions, and more likely to underreport the undesirable behaviors. By contrast, respondents with desirable behaviors are more likely to participate in surveys, more likely to provide answers to sensitive questions, and more likely to overreport the desirable behaviors. As a result, an effort to bring in more respondents to surveys with sensitive questions could run the risk of reducing nonresponse error at the cost of increasing measurement error, leading to an increase in overall bias (Sakshaug et al. 2010).

8. STRATEGIES TO IMPROVE ANSWERS TO SENSITIVE QUESTIONS

Reviews so far suggest that asking sensitive questions in surveys and censuses could negatively affect response rates of the surveys, increase wave nonresponse to future rounds of interviewing, produce more missing data, and introduce social desirability bias in answers to sensitive questions. Survey researchers and practitioners have been exploring ways to mitigate these negative impacts of sensitive questions on data quality. This section focuses on strategies to improve respondents' answers to sensitive questions for two reasons. First, measurement error poses a greater threat to data quality of surveys with sensitive questions than nonresponse error, as shown in Section 7. Second, there is more empirical research on improving responses to sensitive questions than on reducing nonresponse errors.

Two review articles reviewed and summarized methods to reduce socially desirable responding across a range of outcomes. Tourangeau & Yan (2007) evaluated factors affecting reporting on sensitive topics and noted that five strategies are found to consistently improve reporting on sensitive questions in general. The first strategy is to use a self-administered mode of data collection when sensitive questions are involved. This strategy is also recommended by Yan & Cantor (2019). Self-administered modes of data collection include mail surveys, web surveys, audio computer-assisted self-interviewing surveys, paper and pencil questionnaires, and so on. Unlike interviewer-administered surveys, self-administration eliminates the need to verbally report answers to sensitive questions to an interviewer and removes respondents' concerns about privacy and looking good in front of the interviewer. Furthermore, self-administration also reduces the risk of someone else overhearing the answer. As a result, self-administered modes of data collection are consistently found to have improved reports to sensitive questions (see Tourangeau & Yan 2007 for a review; see also Lind et al. 2013, Preisendörfer & Wolter 2014).

The second strategy recommended by Tourangeau & Yan (2007) is to collect data in private. Although survey interviews are recommended to be conducted alone in a private setting, it is not uncommon that other people are present. Typical bystanders include a respondent's spouse, children, or parents (especially for surveys of youths). The presence of bystanders is found to increase misreporting to sensitive questions, a phenomenon also called bystander effect or third-party presence effect (e.g., Diop et al. 2015). Tourangeau & Yan (2007) conducted meta-analyses to look at the effect of different types of bystanders. The meta-analyses indicate that the presence of a spouse and children does not have a significant overall impact on survey responses. By contrast, parental presence significantly reduces socially undesirable responses; in other words, youths were less likely to report socially undesirable behaviors when parents were around during the interview than when parents were not present, a finding later corroborated by Herrera and associates (2017).

Besides ensuring that no other person is present during the survey interview, other measures to guarantee privacy, confidentiality, and anonymity reduce misreporting to sensitive questions as well. For example, Ong & Weiss (2000) found that, despite being assured of confidentiality, students who had to provide their full name and social security number were less likely to admit cheating, stealing, and masturbation in the past month than students in the anonymity condition. The difference in the percentage answering yes to sensitive questions between the two groups ranges from 26% for stealing, 37% for masturbation, and 49% for reporting cheating (Ong & Weiss 2000, table 1). Yang & Yu (2011) manipulated the cover design of a mail survey to medical doctors and researchers in Taiwan. One-third of the mail questionnaires showed a barcode, one-third had a number code, and one-third had no identifier at all. Unknown to survey respondents, the barcode and the number code were the same for everyone; in other words, the barcode and the number code could not identify the respondent. Still, Yang & Yu (2011) found that questionnaires with an identifier yielded more item nonresponse and fewer socially undesirable answers than questionnaires without an identifier. There was no difference between the two types of identifiers with regard to item nonresponse and number of socially undesirable answers.

Both Tourangeau & Yan (2007) and Yan & Cantor (2019) recommend indirect methods to collect sensitive data. Survey questions directly ask respondents about a sensitive topic such as whether or not they voted in the last US presidential election. Answers to direct questions like this can be linked to respondents' involvement in socially (un)desirable behaviors; an answer of yes means that the respondent voted and an answer of no means that the respondent did not vote. By contrast, indirect methods aim to protect the anonymity of respondents' answers by breaking the link between answers and involvement in the socially (un)desirable behaviors. The randomized response technique (RRT), specifically recommended in Tourangeau & Yan (2007), is one type of indirect method. It relies on a randomizing mechanism (e.g., asking the respondent to flip a coin,

or using the respondent's birth month). The respondent is to answer truthfully to the sensitive question when the randomization results in a certain outcome (e.g., when the coin flip is tails). But when the coin flip is heads, he/she is instructed to either give a predetermined answer (such as always responding no) or answer a nonsensitive question. The randomizing mechanism provides protection for the respondent's true status with respect to the sensitive question, because there is stochastic uncertainty as to which question was answered. As a result, the RRT is found to increase truthful answers to sensitive questions (Lensvelt-Mulders et al. 2005, Rosenfeld et al. 2016), but at the cost of a smaller effective sample size.

The item count technique (ICT) is another type of indirect method in which respondents are randomly assigned to receive a long list of questions including the sensitive question of interest, or the same list without the sensitive question. Respondents are then asked to provide the number of items from the list that apply to them. A respondent's true status with regard to the sensitive question is protected because their answer is a count variable. Rosenfeld and associates (2016) conducted a validation study to assess ICT and RRT and found that both methods produced estimates closer to the truth than when respondents were directly asked about that sensitive question. However, RRT led to a smaller bias and a smaller variance than ICT.

The bogus pipeline technique, as recommended by Tourangeau & Yan (2007), works on a different principle; it aims to convince respondents that researchers have known their answers and will know if they lie. For example, Tourangeau and associates (1997b) told a random half of respondents that their inaccurate answers could be detected by a physiological recording device. Although two meta-analyses failed to find support for the effectiveness of the bogus pipeline technique in improving answers to questions on alcohol and marijuana consumption (Aguinis et al. 1995), Tourangeau and associates (1997b) found that the procedure increased reporting of socially undesirable behaviors (e.g., sexual partners and illicit drug use) and decreased the reporting of exercising (a socially desirable behavior). Hanmer and associates (2014) phrased the introduction to a voting question under the bogus pipeline approach; the introduction told respondents that their answers would be checked against public records kept by election officials. This manipulation reduced the overreporting of voter turnout; the bogus pipeline introduction produced 17% of overreporting compared with 25% for the control condition.

The fifth strategy recommended by Tourangeau & Yan (2007) is to prime respondents to be honest. Two studies experimentally tested the impact of priming on survey responses to sensitive questions. Rasinski and associates (2005) gave undergraduate students a vocabulary task before asking them to complete a seemingly unrelated questionnaire on alcohol assumption. A random half of the respondents received neutral words, whereas the other half received a combination of neutral words and words related to honesty. They found that respondents primed to be honest (through the vocabulary task) reported more sensitive behaviors involving excessive alcohol consumption. They also showed that the priming effect is statistically significant only when questions on alcohol consumption immediately followed the priming task and not when asked later in the survey. Acquisti and associates (2012, study 2c) primed a volunteer sample of *New York Times* readers on privacy through a photo identification task. Participants were asked to identify either phishing emails or endangered fish before completing a web survey. They found that the privacy priming decreased respondents' propensity to admit socially undesirable behaviors.

Survey researchers have also experimented with wording changes made to introductory sentences, question stems, and/or response options (see Tourangeau & Yan 2007, Yan & Cantor 2019). Both review articles suggest forgiving wording when asking sensitive questions. Forgiving wording can be implemented as adding a forgiving introduction, rephrasing the questions in a forgiving tone, and rephrasing or adding response options in a face-saving manner. Forgiving introductions take several forms. The first is the everyone-does-it approach. For example, Holtgraves and

associates (1997, p. 1670) introduced a survey question on vandalism with a sentence: “Almost everyone has probably committed vandalism at one time or another.” The second form of forgiving introduction shows two or more sides of a story using the “some... others...” format. A typical example is provided by Peter & Valkenburg (2011, p. 782): “Some people use erotic or pornographic materials often, while others do this rarely or never.” The third type of forgiving introduction provides a face-saving excuse for not doing something socially desirable or for doing something socially undesirable. Holtgraves and colleagues (1997, p. 1670) preceded an item about global warming with: “You may not have had enough time to learn about the GATT treaty because of a heavy load of school work.” Empirical research found a consistent positive impact of forgiving introductions on answers to knowledge questions (Holtgraves et al. 1997) but no impact on other types of sensitive questions (e.g., Peter & Valkenburg 2011, Persson & Solevid 2014).

Holtgraves and associates (1997) also experimented with face-saving question wording for knowledge questions and socially desirable behavior questions. For example, the face-saving version of a socially desirable knowledge question asks, “Have you had the opportunity to learn the details of the Clinton Health Care Plan?” whereas the control condition’s wording is, “Do you know the details of the Clinton Health Care Plan?” (Holtgraves et al. 1997, p. 1668). Holtgraves and associates (1997) found that face-saving question wording consistently reduced social desirability bias in answers to knowledge questions but not in answers to socially desirable behavior questions.

Two studies examined the impact of face-saving response options on answers to questions on voting and political participation. Persson & Solevid (2014) tested the additions of two face-saving response options (“I have done this before,” “I have not done this but would consider doing it in the future”) to questions on political participation during the past 12 months. They found that the addition of these two response options decreased the reported levels of political participation. Similarly, Zeglovitis & Kritzinger (2014) compared three sets of response options to questions on voting. The first set contains only “yes, I voted” and “no, I did not vote.” The second set includes, in addition to the yes and no options, two face-saving options (“I thought about voting this time but didn’t” and “I usually vote but didn’t this time”). The third set includes the likelihood of turnout in the answer options (“I am sure I did not vote in the federal election in September 2008,” “I am not sure if I voted but I think it is more likely that I did not,” “I am not sure if I voted but I think it is more likely that I did,” and “I am sure that I voted in the federal election in September 2008”). Zeglovitis & Kritzinger (2014) found that the last two sets of response options significantly reduced overreports of voter turnout for a past election.

In an attempt to reduce overreporting of voting behavior, Belli and associates (1999) proposed two alternative methods; the exact wordings of the two alternative methods together with the standard voter turnout question used in the American National Election Studies (ANES) are presented in **Table 5**. The first method, denoted Belli–Long, combines a forgiving introduction, subtle cues to get people to think about the truthfulness of their answers, memory cues about that particular election day (such as how respondents went to vote, the weather that day, and people they were with or met or saw on that day), a new question wording, and two additional face-saving response options. The Belli–Short method uses exactly the same forgiving introduction, question wording, and response options but includes neither the subtle cues on truthfulness of answers nor the memory cues. Both alternatives were found to consistently reduce overreporting of voter turnout as compared with the standard ANES question (Belli et al. 1999, 2006; Duff et al. 2007; Waismel-Manor & Sarid 2011; Kritzinger et al. 2012; Zeglovitis & Kritzinger 2014).

Other strategies related to question wording discussed by Tourangeau & Yan (2007) and Yan & Cantor (2019) include using question wording presupposing the sensitive behavior, using familiar wording, using unfolding brackets and respondent-generated intervals, asking questions about

Table 5 Standard and alternative questions on voter turnout

Standard wording used in American National Election Studies (Belli et al. 1999, p. 106)	In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time. How about you—did you vote in the elections this November?
Belli-Long (Belli et al. 1999, p. 106)	<p>(Forgiving introduction) In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time.</p> <p>(Subtle cues on truthfulness of reporting) We also sometimes find that people who thought that they had voted actually did not vote. Also, people who usually vote may have trouble saying for sure whether they voted in a particular election. (memory cues about election day) In a moment, I'm going to ask you whether you voted on Tuesday, November fifth, which was [time fill] ago. Before you answer, think of a number of different things that will likely come to mind if you actually did vote this past election day; things like whether you walked, drove, or were driven by another person to your polling place [pause], what the weather was like on the way [pause], the time of day that was [pause], and people you went with, saw, or met while there [pause]. After thinking about it, you may realize that you did not vote in this particular election [pause].</p> <p>(New question wording) Now that you've thought about it, which of these statements best describes you?</p> <ol style="list-style-type: none">1. I did not vote in the November fifth election.2. I thought about voting this time but didn't. (face-saving response option)3. I usually vote but didn't this time. (face-saving response option)4. I am sure I voted in the November fifth election.7. (VOLUNTEERED) I VOTED BY ABSENTEE BALLOT.
Belli-Short (Belli et al. 2006, p. 758)	<p>In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time. Which of the following statements best describes you?</p> <ol style="list-style-type: none">1. I did not vote in the November third election.2. I thought about voting this time but didn't.3. I usually vote but didn't this time.4. I am sure I voted in the November third election.7. (VOLUNTEERED) I VOTED BY ABSENTEE BALLOT.

past behavior only or before asking about current behavior, and using a high-frequency response list or the open-ended format. Most of these strategies were experimentally tested in a handful studies, and there is insufficient evidence suggesting that they consistently succeed in reducing misreporting to sensitive questions.

9. CONCLUSIONS AND DISCUSSION

It is sometimes necessary that sensitive questions be included in a survey. Questions are sensitive because they are intrusive, elicit concerns from disclosure, or invoke social desirability concerns. Even though the perceived sensitivity of survey questions differs by respondent characteristics, differs across cultures and regions, and changes over time, sensitive questions subject survey respondents to emotional and cognitive difficulties.

Considerable research has documented the impact of asking such questions on multiple sources of error, and I have summarized them using a TSE framework. Sensitive questions negatively impact a respondent's willingness to participate in a survey when the survey topic explicitly mentions

the nature of the sensitive questions included. For longitudinal surveys, respondents with socially undesirable characteristics (e.g., smokers, substance users) and those who chose not to answer sensitive questions are less likely to respond to later rounds of interviews, contributing to wave nonresponse. Sensitive questions elicit higher item nonresponse rates and produce social desirability bias in the estimates. Furthermore, sensitive questions influence multiple sources of survey error simultaneously and introduce positive relations between survey errors resulting from unit nonresponse, item nonresponse, and measurement error.

Survey researchers and practitioners have employed various strategies to address the measurement error in answers to sensitive questions. Some strategies are found to consistently improve answers to sensitive questions (such as using a self-administered mode of data collection), but there is not enough evidence to support the effectiveness of other strategies (e.g., using familiar wording). Future research is needed to experimentally test those strategies. Meta-analyses are also welcomed to quantitatively summarize the effect of these strategies on reducing measurement error in answers to sensitive questions.

This review has three limitations. First, I do not review the impact of asking sensitive questions on coverage and sampling errors. As pointed out by Johnson (2014), future research should assess the ability of various sampling frames to cover respondents engaged in sensitive behavior of interest and understand how sampling errors may be influenced by other sources of survey error.

Second, this article only reviews strategies to reduce misreporting to sensitive questions. Strategies to reduce nonresponse error induced by sensitive questions are not discussed. I encourage researchers to develop weighting and imputation models capitalizing on the positive relation between unit and item nonresponse to reduce nonresponse bias induced by sensitive questions.

Third, this article reviews a selected number of studies and is not a comprehensive review from a systematic database search. Future researchers are encouraged to conduct a systematic review and to summarize the impact of asking sensitive questions in a meta-analytic approach.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

- Acquisti A, John LK, Lowenstein G. 2012. The impact of relative standards on the propensity to disclose. *J. Mark. Res.* 49:160–74
- Aguinis H, Pierce CA, Quigley BM. 1995. Enhancing the validity of self-reported alcohol and marijuana consumption using a bogus pipeline procedure: a meta-analytic review. *Basic Appl. Soc. Psychol.* 16:515–27
- Andreenkova AV, Javeline D. 2019. Sensitive questions in comparative surveys. In *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, ed. TP Johnson, B Pennell, IAL Stoop, B Dorer, pp. 139–60. New York: Wiley
- Beatty P, Herrmann D. 2002. To answer or not to answer: decision processes related to survey item nonresponse. In *Survey Nonresponse*, ed. RM Groves, DA Dillman, JL Eltinge, RJA Little, pp. 71–85. New York: Wiley
- Belli RF, Moore SE, VanHoewyk J. 2006. An experimental comparison of question forms used to reduce vote overreporting. *Electoral Stud.* 25:751–59
- Belli RF, Traugott MW, Young M, McGonagle KA. 1999. Reducing vote overreporting in surveys: social desirability, memory failure, and source monitoring. *Public Opin. Q.* 63:90–108
- Brown JD, Heggeness ML, Dorinski DM, Warren L, Yi M. 2018. *Understanding the quality of alternative citizenship data sources for the 2020 Census*. Tech. Rep., US Census Bur., Washington, DC

- Brown JD, Heggeness ML, Dorinski DM, Warren L, Yi M. 2019. Predicting the effect of adding a citizenship question to the 2020 Census. *Demography* 56:1173–94
- Couper MP, Singer E, Conrad FG, Groves RM. 2008. Risk of disclosure, perceptions of risk, and concerns about privacy and confidentiality as factors in survey participation. *J. Off. Stat.* 24:255–75
- Couper MP, Singer E, Conrad FG, Groves RM. 2010. Experimental studies of disclosure risk, disclosure harm, topic sensitivity, and survey participation. *J. Off. Stat.* 26:287–300
- De Jonge CPK. 2015. Who lies about electoral gifts? Experimental evidence from Latin America. *Public Opin. Q.* 79:710–39
- de Leeuw ED. 1992. *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam: T.T. Publikaties
- Diop A, Le KT, Traugott M. 2015. Third-party presence effect with propensity score matching. *J. Surv. Stat. Methodol.* 3:193–215
- Duff B, Hanmer MJ, Park W, White IK. 2007. Good excuses: understanding who votes with an improved turnout question. *Public Opin. Q.* 71:67–90
- Fu H, Darroch JE, Henshaw SK, Kolb E. 1998. Measuring the extent of abortion underreporting in the 1995 National Survey of Family Growth. *Fam. Plan. Perspect.* 30:128–38
- Gnambs T, Kaspar K. 2015. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. *Behav. Res. Methods* 47:1237–59
- Groves RM. 1989. *Survey Errors and Survey Costs*. New York: Wiley
- Groves RM, Couper MP. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley
- Groves RM, Couper MP, Presser S, Singer E, Tourangeau R, et al. 2006. Experiments in producing nonresponse bias. *Public Opin. Q.* 70:720–36
- Groves RM, Presser S, Dipko S. 2004. The role of topic interest in survey participation decisions. *Public Opin. Q.* 68:2–31
- Groves RM, Presser S, Tourangeau R, West BT, Couper MP, et al. 2012. Support for the survey sponsor and nonresponse bias. *Public Opin. Q.* 76:512–24
- Groves RM, Singer E, Corning A. 2000. Leverage-salience theory of survey participation: description and an illustration. *Public Opin. Q.* 64:299–308
- Hanmer MJ, Banks AJ, White IK. 2014. Experiments to reduce the over-reporting of voting: a pipeline to the truth. *Political Anal.* 22:130–41
- Herrera AV, Benjet C, Méndez E, Casanova L, Medina-Mora ME. 2017. How mental health interviews conducted alone, in the presence of an adult, a child or both affects adolescents' reporting of psychological symptoms and risky behaviors. *J. Youth Adolesc.* 46:417–28
- Holbrook AL, Krosnick JA. 2010. Measuring voter turnout by using the randomized response technique: evidence calling into question the method's validity. *Public Opin. Q.* 74:328–43
- Holbrook AL, Krosnick JA. 2013. A new question sequence to measure voter turnout in telephone surveys: results of an experiment in the 2006 ANES pilot study. *Public Opin. Q.* 77:106–23
- Holtgraves T, Eck J, Lasky B. 1997. Face management, question wording, and social desirability. *J. Appl. Soc. Psychol.* 27:1650–71
- Jagannathan R. 2001. Relying on surveys to understand abortion behavior: some cautionary evidence. *Am. J. Public Health* 91:1825–31
- Johnson TP. 2014. Sources of error in substance use prevalence surveys. *Int. Sch. Res. Notices* 2014:923290
- Johnson TP, O'Rourke D, Chavez N, Sudman S, Warnecke R, et al. 1997. Social cognition and responses to survey questions among culturally diverse populations. In *Survey Measurement and Process Quality*, ed. LE Lyberg, PP Biemer, M Collins, ED de Leeuw, C Dippo, et al., pp. 87–114. New York: Wiley
- Johnson TP, van de Vijver FJR. 2002. Social desirability in cross-cultural research. In *Cross-Cultural Survey Methods*, ed. JA Harkness, FJR van de Vijver, PPH Mohler, pp. 193–209. New York: Wiley
- Jones E, Forrest JD. 1992. Underreporting of abortion in surveys of U.S. women: 1976 to 1988. *Demography* 29:113–26
- Jones RK, Kost K. 2007. Underreporting of induced and spontaneous abortion in the United States: an analysis of the 2002 National Survey of Family Growth. *Stud. Fam. Plan.* 38:187–97
- Junger M. 1989. Discrepancies between police and self-report data for Dutch racial minorities. *Br. J. Criminol.* 29:273–84

- Kreuter F, Presser S, Tourangeau R. 2008. Social desirability bias in CATI, IVR, and web surveys. *Public Opin. Q.* 72:847–65
- Kritzinger S, Schwarzer S, Zeglovits E. 2012. *Reducing overreporting of voter turnout in seven European countries—results from a survey experiment*. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Orlando, Florida
- Krumpal I. 2013. Determinants of social desirability bias in sensitive surveys: a literature review. *Qual. Quantity* 47:2025–47
- Kuhn PM, Vivyan N. 2018. Reducing turnout misreporting in online surveys. *Public Opin. Q.* 82:300–21
- Lensvelt-Mulders GJLM, Hox JJ, van der Heijden PGM, Maas CJM. 2005. Meta-analysis of randomized response research: thirty-five years of validation. *Sociol. Methods Res.* 33:319–48
- Lind LH, Schober MF, Conrad FG, Reichert H. 2013. Why do survey respondents disclose more when computers ask the questions? *Public Opin. Q.* 77:888–935
- Lipps O. 2007. Attrition in the Swiss Household Panel. *Methoden Daten Anal.* 1:45–68
- Loosveldt G, Pickery J, Billiet J. 2002. Item nonresponse as a predictor of unit nonresponse in a panel survey. *J. Off. Stat.* 18:545–57
- McDonald B, Haardoerfer R, Windle M, Goodman M, Berg C. 2017. Implications of attrition in a longitudinal web-based survey: an examination of college students participating in a tobacco use study. *JMIR Public Health Surveill.* 3(4):e73
- McDonald JA, Scott ZA, Hanmer MJ. 2017. Using self-prophecy to combat vote overreporting on public opinion surveys. *Electoral Stud.* 50:137–41
- Ong A, Weiss DJ. 2000. The impact of anonymity on responses to sensitive questions. *J. Appl. Soc. Psychol.* 30:1691–708
- Persson M, Solevid M. 2014. Measuring political participation—testing social desirability bias in a web-survey experiment. *Int. J. Public Opin. Res.* 26:98–112
- Peter J, Valkenburg PM. 2011. The impact of “forgiving” introductions on the reporting of sensitive behavior in surveys: the role of social desirability response style and developmental status. *Public Opin. Q.* 75:779–87
- Preisendörfer P, Wolter F. 2014. Who is telling the truth? A validation study on determinants of response behavior in surveys. *Public Opin. Q.* 78:126–46
- Rasinski KA, Visser PS, Zagatsky M, Rickett EM. 2005. Using implicit goal priming to improve the quality of self-report data. *J. Exp. Soc. Psychol.* 41:321–27
- Régnier-Loilier A, Guisse N. 2012. *Sample attrition and distortion over the waves of the French Generations and Gender Survey*. Tech. Pap., Inst. Natl. Études Démogr., Aubervilliers, Fr.
- Rosenfeld B, Imai K, Shapiro J. 2016. An empirical validation study of popular survey methodologies for sensitive questions. *Am. J. Political Sci.* 60:783–802
- Sakshaug JW, Yan T, Tourangeau R. 2010. Nonresponse error, measurement error, and mode of data collection: tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opin. Q.* 74:907–33
- Shoemaker PJ, Eichholz M, Skewes EA. 2002. Item nonresponse: distinguishing between don’t know and refuse. *Int. J. Public Opin. Res.* 14:193–201
- Tierney KI. 2019. Abortion underreporting in Add Health: findings and implications. *Popul. Res. Policy Rev.* 38:417–28
- Tourangeau R, Groves RM, Redline CD. 2010. Sensitive topics and reluctant respondents: demonstrating a link between nonresponse bias and measurement error. *Public Opin. Q.* 74:413–32
- Tourangeau R, Rasinski K, Jobe J, Smith TW, Pratt W. 1997a. Sources of error in a survey of sexual behavior. *J. Off. Stat.* 13:341–65
- Tourangeau R, Rips LJ, Rasinski KA. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge Univ. Press
- Tourangeau R, Smith TW, Rasinski KA. 1997b. Motivation to report sensitive behaviors in surveys: evidence from a bogus pipeline experiment. *J. Appl. Soc. Psychol.* 27:209–22
- Tourangeau R, Yan T. 2007. Sensitive questions in surveys. *Psychol. Bull.* 133:859–83
- Udry JR, Gaughan M, Schwingl PJ, van den Berg BJ. 1996. A medical record linkage analysis of abortion underreporting. *Fam. Plan. Perspect.* 28:228–31

- Uhrig N. 2008. *The nature and causes of attrition in the British Household Panel Survey*. ISER Work. Pap. Ser. 2008-05, Inst. Soc. Econ. Res., Univ. Essex, UK
- van der Heijden PGM, van Gils G, Bouts J, Hox JJ. 2000. A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning: eliciting sensitive information in the context of welfare and unemployment benefit. *Sociol. Methods Res.* 28:505–37
- Waismel-Manor I, Sarid J. 2011. Can overreporting in surveys be reduced? Evidence from Israel's municipal elections. *Int. J. Public Opin. Res.* 23:522–29
- Warren JR, Halpern-Manners A. 2012. Panel conditioning in longitudinal social science surveys. *Sociol. Methods Res.* 41:491–534
- Watson N, Wooden M. 2009. Identifying factors affecting longitudinal survey response. In *Methodology of Longitudinal Surveys*, ed. P Lynn, pp. 157–81. New York: Wiley
- Wyner GA. 1980. Response errors in self-reported number of arrests. *Soc. Methods Res.* 9:161–77
- Yan T, Cantor D. 2019. Asking survey questions about criminal justice involvement. *Public Health Rep.* 134:46S–56S
- Yan T, Curtin R. 2010. The relation between unit nonresponse and item nonresponse: a response continuum perspective. *Int. J. Public Opin. Res.* 22:535–51
- Yan T, Curtin R, Jans M. 2010. Trends in income nonresponse over two decades. *J. Off. Stat.* 26:145–64
- Yang M, Yu Y. 2011. Effects of identifiers in mail surveys. *Field Methods* 23:243–65
- Zeglavitis E, Kritzinger S. 2014. New attempts to reduce overreporting of voter turnout and their effects. *Int. J. Public Opin. Res.* 26:224–34