

Determinants of social desirability bias in sensitive surveys: a literature review

Ivar Krumpal

Published online: 19 November 2011
© Springer Science+Business Media B.V. 2011

Abstract Survey questions asking about taboo topics such as sexual activities, illegal behaviour such as social fraud, or unsocial attitudes such as racism, often generate inaccurate survey estimates which are distorted by social desirability bias. Due to self-presentation concerns, survey respondents underreport socially undesirable activities and overreport socially desirable ones. This article reviews theoretical explanations of socially motivated misreporting in sensitive surveys and provides an overview of the empirical evidence on the effectiveness of specific survey methods designed to encourage the respondents to answer more honestly. Besides psychological aspects, like a stable need for social approval and the preference for not getting involved into embarrassing social interactions, aspects of the survey design, the interviewer's characteristics and the survey situation determine the occurrence and the degree of social desirability bias. The review shows that survey designers could generate more valid data by selecting appropriate data collection strategies that reduce respondents' discomfort when answering to a sensitive question.

Keywords Sensitive questions · Social desirability bias · Survey design · Survey Methodology · Measurement error

1 Introduction

An increasing number of survey statisticians and social scientists focus on the investigation of social taboos, illegal behavior and extreme opinions. Different national surveys contain item batteries asking about sensitive information. For example, the German General Social Survey (ALLBUS) 2000 asked interviewees to self-report on following four minor offences: (1) using public transportation without buying a valid ticket, (2) driving a car with more than the permitted level of blood alcohol, (3) taking goods from a department store without paying, and (4) deliberately making false statements on tax forms in order to pay less. Other

I. Krumpal (✉)

Department of Sociology, University of Leipzig, Beethovenstr. 15, 04107 Leipzig, Germany
e-mail: krumpal@sozio.uni-leipzig.de

surveys, like the US National Crime Victimization Survey (NCVS) or the European Crime and Safety Survey (EU ICS), ask questions on sensitive topics like experiences with criminal victimization. Recently, a national study on right-wing extremism was conducted in Germany, collecting data on socially undesirable attitudes like anti-Semitism, xenophobia, and chauvinism (Decker and Brähler 2006). In Switzerland, the Swiss Multicenter Adolescent Survey on Health (SMASH) 2002 asked 16–20 years old youths about their use of illicit drugs, and their drinking and smoking habits. To cite a last example, the US General Social Survey (GSS) monitors the sexual activity of the population and also asks about very sensitive topics like prostitution ('Thinking about the time since your 18th birthday, have you ever had sex with a person you paid or who paid you for sex?') or infidelity ('Have you ever had sex with someone other than your husband or wife while you were married?'). Obtaining valid and reliable data on the basis of such items has proven to be a difficult business and the possibilities of doing so continues to be a lively research activity. Survey methodologists' state-of-the-art knowledge suggests that answers to sensitive questions are often distorted by social desirability bias. The first section of this article reviews the main theoretical explanations regarding the process of self-reporting in sensitive surveys. 'Sensitivity' is a complex theoretical concept whose dimensions are identified and discussed. Next, psychological mechanisms are presented, relating 'sensitivity' to other theoretical constructs and to different aspects of data quality. The review focuses on the behavior of both main actors of a survey interview, the respondent and the interviewer, and discusses how survey response is affected by (a) perceived gains, risks and losses of the respondent and (b) the behavior of the interviewer. The second section reviews empirical findings on the effectiveness of different survey methods (such as randomized response or the unmatched count technique) on the respondent's propensity to misreport in sensitive surveys and outlines future research perspectives.

2 Sensitivity and social desirability: defining the concepts

For the term 'sensitivity' different conceptualizations can be observed in the survey literature (Lee 1993). One approach is post hoc assessment of sensitivity via empirical indicators of survey quality (Lensvelt-Mulders 2008; Tourangeau and Yan 2007). For example, questions that are supposed to be sensitive are often associated with comparatively higher item non-response rates than non-sensitive questions. Table 1 summarizes item nonresponse rates for selected items, taken from the German General Social Survey (ALLBUS). The questions were administered to a national random sample from all German speaking persons who resided in private households in Germany and were 18 years old or older:

Table 1 shows that some items (household net income and voting intention) have consistently more missing data than other items (religious denomination, educational attainment, membership of a trade union, employment status and age). Against the background of the assumption 'the more sensitive the item is the higher item nonresponse will be' it seems apparent that the income question has the highest sensitivity of all items with nonresponse rates ranging from 20.7 to 26.2%. In contrast, questions asking about employment status and age seem to have the lowest sensitivity with proportions of missing data ranging from 0.0 to 0.4% for age and from 0.1 to 0.2% for employment status respectively.

Other empirical approaches ask respondents to assess the sensitivity of survey items on specific rating-scales (Bradburn and Sudman 1979; Coutts and Jann 2011). Recently, Coutts and Jann (2011, p. 184) carried out an online survey study that asked 2,075 respondents from a German access panel to rate several petty offences (keeping to much change, freerid-

Table 1 Rates of item nonresponse (%) for the German general social survey ALLBUS, selected years and items

Topic	ALLBUS 1990 (%)	ALLBUS 2000 (%)	ALLBUS 2006 (%)
Household net-income	26.2	23.5	20.7
Voting intention ^a	14.4	22.7	14.2
Religious denomination	0.4	0.7	0.5
Educational attainment	0.8	0.3	0.2
Membership of a trade union	1.7	0.2	0.4
Employment status	0.1	0.2	0.1
Age	0.4	0.0	0.3

Data was either collected by PAPI—paper and pencil interviewing (ALLBUS 1990 and 2000), CAPI—computer assisted personal interviewing (ALLBUS 2000 and 2006), or CASI—computer assisted self interviewing (ALLBUS 2006)

^a Statistic includes answer category ‘don’t know’

ing, shoplifting, marihuana use and drunk driving) and immoral activities (to cheat on one’s partner). For each item, a total sensitivity score was calculated by adding the proportions of interviewees who stated (1) that the behavior in question is not alright, and (2) that admitting it would be uncomfortable for most. The most sensitive topics turned out to be shoplifting (79%) and infidelity (73%). These were followed by drunk driving (53%) and marijuana use (43%), both with medium sensitivity scores. In contrast, freeriding (22%) and keeping too much change (20%) were considered as topics with lower sensitivity.

Theory-driven approaches try to distinguish different aspects of the theoretical construct ‘sensitivity’. According to Lee and Renzetti a topic labeled ‘sensitive’ is one that “potentially poses for those involved a substantial threat, the emergence of which renders problematic for the researcher and/or the researched, the collection, holding, and/or dissemination of research data” (Lee and Renzetti 1993, p. 5). They argue that research on sensitive topics seems to be linked with risks and costs, such as negative feelings of shame and embarrassment or negative consequences, such as the possibility of sanctions. Finally, they strongly emphasize the social dimension of sensitivity: “In other words, the sensitive character of a piece of research seemingly inheres less in the topic itself and more in the relationship between that topic and the social context within which the research is conducted” (Lee and Renzetti 1993, p. 5).

Another useful specification of the concept ‘sensitivity’ is introduced by Tourangeau and Yan (2007). They distinguish between three distinct aspects of the term ‘sensitivity’:

1. The first dimension is ‘intrusiveness’ and refers to the fact that within a given culture certain questions per se may be perceived as too private or taboo, independent of the respondents’ true status on the variable of interest. Questions asking about the respondents’ sexual preferences, health status or income are often perceived as too intrusive.
2. The second dimension is ‘threat of disclosure’, pertaining to respondents’ concerns about possible risks, costs or negative consequences of truthfully reporting a sensitive behavior should the sensitive answers become known to third persons or institutions beyond the survey setting. Such negative consequences could be: job loss, family upset or even prosecution. Questions asking the respondent to self-report illegal behavior (e.g. employee theft, tax fraud or illegal entry in surveys of immigrants) may fall into this category.
3. The third dimension is ‘social desirability’. This dimension refers to truthfully reporting an attitude or behavior that clearly violates existing social norms and thus is deemed unacceptable by society. To conform to social norms, respondents may present themselves in a

positive light, independent of their actual attitudes and true behaviors respectively. More specifically, ‘social desirability’ refers to the respondents’ tendency to admit to socially desirable traits and behaviors and to deny socially undesirable ones. Finally, socially desirable answers could also be conceptualized as respondents’ temporary social strategies coping with the different situational factors in surveys (e.g. presence of interviewer, topic of question, etc.).

Unlike ‘intrusiveness’, the problem associated with ‘social desirability’ is not the sensitivity of a question but the sensitivity of an answer. Fowler (1995, p. 29) summarizes this issue as follows: “Questions tend to be categorized as ‘sensitive’ if a ‘yes’ answer is likely to be judged by society as undesirable behaviour. However, for those for whom the answer is ‘no’ questions about any particular behaviour are not sensitive.” Whereas answers suggesting deviations from social norms are seen socially undesirable, self-reports suggesting norm-conforming behaviors are considered socially desirable associated with expected gains such as social approval of the interviewer. Given that, respondents tend to underreport socially undesirable behavior and overreport socially desirable behavior. They distort their answers towards the social norm in order to maintain a socially favorable self-presentation (an overview of the literature of social norms can be found in [Rauhut and Krumpal 2008](#)).

Two sub-dimensions of the concept ‘social desirability’ are often distinguished ([Randall and Fernandes 1991](#)): One sub-dimension refers to social desirability as a stable personality characteristic, such as a constant need for social approval and impression management, to cause socially desirable misreporting ([Crowne and Marlowe 1960, 1964](#); [DeMaio 1984](#)). A strong approval motive and an invariant desire to generate a positive image may thus reduce the interviewee’s willingness to disclose self-stigmatizing information. By contrast, the second sub-dimension refers to social desirability as an item characteristic, considering various activities or attitudes to be more or less socially undesirable and thus relates perceived desirability of a behavior to particular items. Thus, effects of social desirability are strongly influenced by characteristics of a specific item ([Groves 1989](#)).

Social desirability refers to making oneself look good in terms of prevailing cultural norms when answering to specific survey questions. However, the general need for social approval and impression management may vary with specific subgroup norms ([Johnson and van de Vijver 2002](#); [Lee and Renzetti 1993](#)). Furthermore, the tendency to give socially desirable responses may vary across cultural orientations like collectivistic cultures that emphasize good relationships with other group members versus individualistic orientations that set value on the pursuit of one’s personal values, attitudes and goals ([Lalwani et al. 2006](#)).

3 Response error and other types of survey errors

Questions asking about sensitive topics are assumed to generate response errors thus having a negative impact on data quality. Before investigating how sensitive questions increase the likelihood of response errors, especially response bias, it is important to define the terms ‘response error’ and ‘response bias’ and distinguish these concepts from other types of survey errors. Useful typologies of survey errors can be seen in [Fox and Tracy \(1986\)](#) and [Groves et al. \(2004\)](#). One fundamental distinction classifies survey errors as either sampling errors or nonsampling errors.

The first class of survey error, sampling error, arises from the fact that only a subset of all potential respondents in the sampling frame is actually measured. The survey methodol-

ogy literature distinguishes two types of sampling errors: sampling variance and sampling bias (Groves et al. 2004, p. 57). Sampling variance arises from the fact that by a random process many different samples each with different subsets of elements could be drawn from the population under investigation. Each possible sample will produce different estimates on the survey statistic. Sampling bias can result from the possibility that certain subgroups in the target population are not represented (or underrepresented) in the sampling frame thus the selection process excluding them systematically. To the extent that excluded members differ from included members on key variables of the survey, the survey statistics will systematically deviate from the true parameters of the target population.

The second class of error, nonsampling error, is much more relevant in sensitive surveys (Fox and Tracy 1986, p. 8). One type of error in this class is nonresponse error, which refers to differences between the values of statistics computed on the basis of the entire sample and statistical estimates based only on the actual respondent subset of the sample. Another type of nonsampling error is response error. This kind of error arises from an observational gap between the true score of a respondent and the actual answer provided (Marquis et al. 1981, pp. 2–8). Response error is derived from the observation process itself, the term ‘response error’ is often used synonymously with ‘measurement error’. Each specific type of nonsampling error can be separated into a random part, which reduces the reliability of measurements and a nonrandom part introducing bias into survey estimates. The systematic part of nonresponse error is often labeled ‘nonresponse bias’ representing systematic differences between respondents and nonrespondents. Technically, nonresponse bias for the sample mean is defined as the product of the nonresponse rate and the difference between the nonrespondent and the respondent mean. The nonresponse rate is “the proportion of eligible sample elements for which data are not collected” (Groves et al. 2004, p. 59). When nonresponse (e.g. refusals to answer a sensitive question) is related to key variables of the survey (e.g. sensitive behavior to be measured) the results may be no longer valid. Additionally, the standard error of the estimates becomes greater as the sample size becomes smaller.

The basic distinction between random and systematic error can also be applied to response error, the best documented source of error in sensitive surveys. The formal notation of response error on an individual level decomposes an answer A_{it} of respondent i on occasion t into three components (Tourangeau et al. 2000, pp. 266–267):

1. The first is the true score T_i , which represents the respondent’s actual status on the variable in question. Respondents’ true scores can sometimes be determined via external data sources like medical or administrative records.
2. The second reflects any general directional tendency across respondents to misreport, more specifically to underreport socially undesirable activities and to overreport socially desirable activities respectively. This component is called bias b . Response bias is “a systematic tendency to respond to a range of questionnaire items on some other basis than the specific item content” (Paulhus 1991, p. 17).
3. The third component, random error e_{it} , is directionless with an expected value assumed to be zero: $E(e_{it}) = 0$. This error component varies between respondents and between occasions within a single respondent:

$$A_{it} = T_i + b + e_{it}$$

If $b \neq 0$, survey measurements of the respondent’s true status are no longer valid. Whereas random error cancels out over repeated measurements, response bias does not. Rather, the systematic difference between observed scores and true scores persists.

There is ample empirical evidence that respondents systematically overreport socially desirable behaviors and attitudes and systematically underreport socially undesirable ones (Barnett 1998; Lee 1993; Tourangeau et al. 2000; Beyer and Krumpal 2010). For example, underreporting is quite common for socially undesirable behaviors like illicit drug use, smoking, alcohol consumption and abortion. Respondents also underreport crime victimization, certain types of income (e.g. welfare) and unpopular attitudes, like racism and anti-Semitism. By contrast, survey studies on socially desirable behavior found empirical evidence for overreporting including activities such as voting, seat belt use, environmentally responsible action (e.g. energy conservation, recycling) and religious participation. Misreporting increases as the questions become more sensitive and decreases as the conditions of data collection become more private (Ong and Weiss 2000).

For sensitive behavior, social desirability bias on an aggregate level not only depends on the extent of sensitivity and privacy but also on the fraction of the population who are engaged in the sensitive behavior in question: "For socially desirable behaviour (...) the potential for overstatement is greater on library card ownership than for voting since only a minority of adults have library cards" (Sudman and Bradburn 1982, p. 56). After specification of the concepts 'sensitivity' and 'response bias' theoretical explanations and empirical studies on respondents' behavior in sensitive surveys will be discussed.

4 Theoretical explanations of social desirability bias in sensitive surveys

Psychological studies on lying indicate that lying in everyday life is a common social interaction process suggesting relatively low cognitive burden (DePaulo et al. 1996, 2003). People lie to avoid negative emotions of shame, embarrassment and losing face in social interactions (Schaeffer 2000). Cognitive psychologists' research suggest misreporting on sensitive questions being a controlled, deliberate and motivated process at least partly under the respondent's voluntary control, rather than an automatic mental process happening completely outside of the respondent's consciousness (Holtgraves et al. 1997; Holtgraves 2004). Respondents are supposed to edit their answers in a socially desirable way, either because of their inclination to impression management or because of their susceptibility to self-deception (Paulhus 2003): In the case of the impression management mechanism, respondents strive for social approval via selecting the answer that is expected to maximize positive valuations and minimize negative reactions by other subjects. In contrast, the concept of self-deception assumes that interviewees want to maintain a positive self-image, to maximize self-worth and to reduce cognitive dissonance resulting from divergence between social norms, self-perception and self-demands on the one hand, and reality on the other hand. According to this perspective, respondents themselves are the main addressees of socially motivated misreporting.

The rational side of answering a sensitive question can be conceptualized within the framework of rational choice theory (RC theory) and subjective expected utility theory (SEU-theory) respectively. Empirical applications of RC theory assume the respondent's likelihood to answer truthfully to be a function of expected risks and losses from answering truthfully (Becker 2006; Becker and Günther 2004). The general assumption of RC theory is that responding to a survey question is a goal-directed, utility-maximizing selection between different response options (Esser 1986; Stocké 2007a,b; Stocké and Hunkler 2007): Interviewees aim to maximize positive feelings of social approval and to avoid dismissive reactions from other individuals. For this purpose, respondents use strategies of impression management such as answering in a socially desirable way. RC theory

postulates three necessary preconditions for social desirability bias (Stocké 2007b): (1) a strong desire for social approval, (2) a nonzero subjective probability of negative sanctions due to a perceived lack of privacy, and (3) respondents believe that the choice of one or another response option matters, i.e. that the other subjects' reactions will be clearly different for response option A compared to response option B. A multiplicative combination of all three factors is assumed to affect response behavior and to determine the strength and the direction of social desirability bias (Stocké 2007b, p. 495): "If only one of these conditions is not given, nothing will affect the prevalence of SD-bias [social desirability bias], and subjects are assumed to report their "true scores"." An analysis of a survey study on racial attitudes indicates empirical evidence for the postulated three-way interaction effect on the respondents' propensity to give a socially desirable answer (Stocké 2007b).

The behavioural model of SEU-theory can be applied to study respondents' perceptions in sensitive surveys, by modelling perceived losses and gains in the interview situation and investigating their impact on the respondent's decision of whether to respond truthfully or not (Rasinski et al. 1994, 1999): If respondents have been engaged in some frowned-upon behavior, one can think of the consideration whether to answer a sensitive question truthfully or not as "making a risky decision with incomplete knowledge about the associated risks and losses" (Rasinski et al. 1999, p. 467). The first factor, 'perceived risks', subsumes the respondent's perceptions of the conditional probabilities of alternative outcomes given each possible response option. The second factor, 'perceived losses and gains', associates each possible outcome with the respondent's valuations of this outcome. Applying the perspective of SEU-theory to the survey context, one can view a respondent's decision whether to admit to a sensitive behavior or not, as consideration of different risks, losses and outcomes associated with that decision.

In case of admission to a socially frowned-upon or illegal behavior perceived losses might be negative feelings of embarrassment during the interview, especially if the interviewer visibly showed disapproval. Furthermore, painful or stressful feelings, like guilt and shame, may arise in consequence of remembering and truthfully reporting embarrassing behavior that conflicts with the respondent's own values (Schaeffer 2000, p. 117). Outside the interview situation, repercussions like informal sanctions, harassment or even prosecution may result from disclosure of sensitive answers to persons or agencies other than the interviewer or the survey research institute. The different risks can be linked to potential subjective costs and losses: "These losses include further intrusions or solicitations, embarrassment, painful memories, and threat to the respondent's self-concept." (Schaeffer 2000, p. 118). Besides risks and losses, respondents might also take into account different perceived gains associated with truthful reporting. Subjective gains could motivate respondents to answer truthfully, especially if the survey is perceived as one with high legitimacy. Positive emotions and personal satisfaction may be generated via consistency with internalized norms (e.g. norms of politeness, cooperation, and norms regarding telling the truth) or via the promotion of public institution and social welfare.

There are several empirical applications of SEU-theory to misreporting on sensitive survey questions (Nathan et al. 1990; Rasinski et al. 1994, 1999; Sirken et al. 1991; Willis et al. 1994). Willis et al. (1994) found a significant relationship between evaluations of risks and losses concerning response disclosure and the decision to answer truthfully to a sensitive question. Rasinski et al. (1994, 1999) did a series of experimental studies using vignettes to investigate the impact of perceived risks, losses and variation in survey design and context on the likelihood of telling the truth in sensitive surveys: Survey design (e.g. interviewer- versus self-administered interviews) and context variables (e.g. husband and children are present

or not) were varied via vignettes. Perceived risks¹ and losses² were measured directly via specific items. For different conditions, experimental subjects rated whether the respondents in the hypothetical survey interviews were likely to tell the truth. [Rasinski et al. \(1994\)](#) found some empirical evidence for the SEU-theory's predictions relating respondents' perceived risks and losses to their tendency of responding truthfully to a sensitive survey question. They subjected 96 male and 96 female subjects to written hypothetical scenarios describing an interview situation. In the women's versions, the hypothetical interviewees were asked about abortion and drunk driving. In the men's versions, the interview topics were number of sex partners prior to marriage and drunk driving. The scenarios varied 3 aspects of the interview situation: a) the data collection mode (interviewer- versus self-administered); b) the interviewer's age (20 years versus 50 years); and c) the presence of family members (present versus absent). Subjects were randomly assigned to the scenarios. On a ten-point scale, subjects rated the likelihood of the hypothetical interviewee admitting to the sensitive behavior, and then judged risks and losses of possible outcomes (such as being embarrassed, receiving understanding or respondent's spouse finding out the sensitive information). The results of the vignette study indicate a lower probability of truthful reporting if family members were present and the interviewer was older. Furthermore, female subjects showed the lowest probability to tell the truth if the questions were interviewer-administered and family members were at home. Finally, [Rasinski et al. \(1994, p. 500\)](#) tested the SEU-theory's core prediction of a statistical association between risk/loss perceptions (independent variables) and the likelihood of admitting to a sensitive behavior (dependent variable).³ They found that subjects' decisions whether to tell the truth or not were statistically correlated with perceived risks, such as the risk of embarrassment over the interviewer's reaction or the risk of disclosure to the respondent's spouse. Fear of embarrassment with respect to the interviewer's reactions suggest the value of data collection modes that respondents perceive as more private, especially self-administered procedures without presence of an interviewer. [Rasinski et al. \(1999\)](#) conducted two experiments to test the SEU-model and to further investigate the effect of privacy on the probability of telling the truth in sensitive surveys. Again, context variables and features of the survey design influenced both, the subjective probabilities of negative outcomes and the level of misreporting.

To summarize, SEU-theory turns out to be a useful tool to parsimoniously conceptualize and measure respondent's perceptions of threat and to explain misreporting in sensitive surveys. [Rasinski et al. \(1994, 1999\)](#) showed that respondents are concerned about different risks and losses and that these concerns varied with specific survey conditions. From a practical viewpoint, survey designers may influence respondents' perceptions of different risks and losses. As a baseline, lack of privacy lowers the respondent's willingness to self-report norm-violating behavior. By carefully adjusting different features of the survey design and context, researchers may create a more private and comfortable interview setting. As a consequence of the improved survey conditions, it is assumed that respondents reduce their subjective probabilities of negative outcomes associated with truthfully responding to

¹ E.g. "likelihood the interviewer would show disapproval if respondent told the truth" or "likelihood spouse would learn truth about respondent" ([Rasinski et al. 1999, p. 470](#)).

² E.g. "degree of embarrassment felt if interviewer showed disapproval" or "degree of negative consequences [like family upset] if spouse learned respondent's answer" ([Rasinski et al. 1999, p. 470](#)).

³ The independent variables representing risk/loss perceptions were formed by multiplying each judgment of the probability of an outcome by the perceived costs of an outcome: (a) Embarrassment = interviewer showing disapproval · degree of embarrassment if interviewer showed disapproval; (b) Relief = interviewer showing understanding · degree of feeling better if interviewer showed understanding; (c) Spouse = spouse finds out · degree of negative consequences if spouse found out.

sensitive questions. Fewer concerns about social or legal costs and reduced feelings of jeopardy in turn are expected to improve the accuracy of respondents' self-reports in sensitive surveys.

5 Methods and context variables affecting social desirability bias in sensitive surveys

The development of techniques for eliciting from people what they would prefer to keep secret can be traced back to the fifteenth and sixteenth centuries. Excluding torture, priests in the Western Church were taught, via written manuals, "to probe the mind of penitents in order to bring forth admissions of sinfulness" (Lee 1993, p. 97). Tentler (1977, p. 94) describes rules of conduct, confessors were instructed to follow, like "not to show amazement; exhibit a contorted face; show revulsion (no matter what enormities are confessed); rebuke the penitent; or exclaim 'Oh, what vile sins!'" Furthermore, Archbishop Borromeo invented the confessional box in 1565 in order to create more anonymity, to reinforce social distance and to reduce the unease of the confessor and the sinner. The term 'confessing' refers to revealing intimate, personally threatening or self-discrediting information to another person. Eliciting sensitive information from the informant is a difficult business for the questioner. It becomes less difficult when trust is likely to develop via the establishment of a confidential, non-condemning and private atmosphere reducing the discomfort of the parties involved.

During the twentieth and twenty-first centuries, modern survey research has conducted numerous methodological experiments to study the impact of changes in situational variables and survey design on social desirability bias in sensitive surveys. Two sorts of studies can be distinguished to assess the accuracy of survey reports and to compare two or more different methods to ask the sensitive question: validation studies or record check designs, and comparative studies without validation data (Groves 1989; Marquis et al. 1986; Tourangeau et al. 2000). Validation studies can be considered as the golden standard for testing the value of a method because they enable the researcher to compare individual survey responses with the true status of each individual. They also allow for a comparison of survey estimates to the true means or proportions at the aggregate level. In practice, true scores are sometimes known from external data sources, such as medical or administrative records. To evaluate the effectiveness of alternative methods in reducing response bias, the proportion of correct answers can be calculated for each experimental condition. The method generating the comparatively highest proportion of correct answers can be considered most effective in improving response accuracy. In contrast, comparative studies are experimental surveys comparing different survey conditions without the possibility of external validation with some objective criterion due to strict data protection rules or the simple lack of such data. Survey researchers interpret the results of comparative studies according to the 'more-is-better' assumption for socially undesirable behavior and the 'less-is-better' assumption for socially desirable behavior respectively. When underreporting is expected, a higher survey estimate (mean or proportion) is assumed to be a more valid estimate for the population parameter. When overreporting is plausibly assumed, a lower estimate is interpreted as more valid.

5.1 The data collection mode

The data collection mode is assumed to be one important factor explaining the level of misreporting in sensitive surveys (Des Jarlais et al. 1999; Holbrook et al. 2003; Metzger et al. 2000; Okamoto et al. 2002; Tourangeau et al. 1997; Tourangeau and Yan 2007; Turner et al. 1998, 2005). The main distinction among the different modes is whether the questions

are interviewer- or self-administered. The most common interviewer-administered modes are: paper-and-pencil personal interviews (PAPI), computer-assisted personal interviews (CAPI) and computer-assisted telephone interviews (CATI). Common self-administered modes are: paper-and-pencil self-administered questionnaires (SAQ), walkman-administered questionnaires (audio-SAQ), computer-assisted self-administered interviews (CASI), audio computer-assisted self-interviewing (ACASI), interactive voice response (IVR, the telephone application of ACASI also referred to as T-ACASI) and web-surveys. Typologies of data collection methods and mixed-mode designs can be found in [De Leeuw et al. \(2008\)](#).

Several experimental field studies have investigated and quantified the effects of self-administration (compared to interviewer-administration) on response accuracy in sensitive surveys. In many cases self-administered survey modes increased levels of reporting of socially stigmatizing medical conditions, such as anxiety, depression and other psychiatric disorders. Furthermore, respondents self-reported more socially undesirable activities, illicit drug use (e.g. marijuana and cocaine), alcohol problems, risky sexual behavior and abortions when the questions were self-administered ([Tourangeau and Yan 2007](#), pp. 863–867). In addition, self-administration also decreased social desirability bias in answers to questions about racial attitudes ([Krysan 1998](#)), sexual activity ([Gribble et al. 1999](#)), same-gender sex ([Villarroel et al. 2006](#)) and sexually transmitted diseases ([Villarroel et al. 2008](#)). Empirical studies show that women tend to underreport the number of their past opposite-sex sexual partners, at the same time men tend to overreport the quantity of their sex partners ([Smith 1992](#)). Women seem to be embarrassed to report too many sexual partners and men seem to have difficulties to report little sexual experience. [Tourangeau and Smith \(1996\)](#) showed that the discrepancy between self-reports of men and women diminished when the questions were self-administered. In the absence of an interviewer the average number of self-reported sexual partners increased for women and decreased for men. In summary, methods of self-administration, minimizing the presence of the interviewer, seem to increase respondents' privacy, to reduce feelings of jeopardy and to decrease subjective probabilities of painful emotions like shame and embarrassment associated with the presence of an interviewer thus generating more honest answers to sensitive questions.

5.2 Interviewer effects

In the presence of an interviewer during the data collection, effects of interviewers' characteristics (e.g. gender and socio-economic status) and assumed interviewers' expectations on social desirability bias can be observed. [Katz \(1942\)](#) found increased reporting of pro-labour attitudes when interviews were conducted by working class interviewers. In two experimental surveys, [Robinson and Rhode \(1946\)](#) found fewer anti-Semitic opinions made in front of interviewers of Jewish appearance or name. [Schuman and Converse \(1971\)](#) found lower reporting of racial attitudes (e.g. identification with black militancy or anti-white sentiment) in black respondents depending upon whether the interviewer was black or white. In a non-experimental validation study comparing self-reported voting with local registration and voting records, [Anderson et al. \(1988\)](#) found black non-voters interviewed by black interviewers to be more likely to incorrectly report that they voted compared to black non-voters who were interviewed by white interviewers. [Fowler and Mangione \(1990, p. 105\)](#) assume such interviewer effects to occur most likely in situations "when the topic of a survey is very directly related to some interviewer characteristics so that potentially a respondent might think that some of the response alternatives would be directly insulting or offensive or embarrassing to an interviewer." In order to avoid conflict in personally insignificant situations the respondent guesses the interviewer's internalized norms and expectations and adjusts his answer

accordingly. In terms of SEU-theory's predictions, heightened subjective costs of a truthful response increase the respondent's tendency to give a socially desirable response.

Besides the respondent's uneasiness facing an interviewer, some researchers also focus on the interviewer's feelings of embarrassment and discomfort associated with asking a specific question (Bradburn and Sudman 1979; Hox and De Leeuw 2002; Schnell and Kreuter 2005). If the interviewer feels uncomfortable about asking a certain question, she may skip the question entirely or deliberately change the wording of the question. Such interviewer effects may seriously distort answers to sensitive questions. Empirical findings suggest a positive association between interviewers' expectations of the study's difficulty and the actual problems they experience in the data collection process: Singer and Kohnke-Aquirre (1979) and Sudman et al. (1977a) gathered empirical evidence indicating lower reports of sensitive behaviors for interviewers expecting a study to be difficult. In the case of asking sensitive questions in face-to-face interviews, researchers often use a method called 'sealed envelope technique' (Barton 1958; De Leeuw 2001; Sudman and Bradburn 1974). In the sensitive questions part of the interview a self-administered questionnaire is handed to the respondent. The interviewer asks the respondent to fill out the separate questionnaire containing the sensitive questions. After that, respondents are requested to put the completed questionnaire into an envelope, seal it and give it to the interviewer. Thus the interviewers do not have to read the embarrassing question and also remain ignorant about the respondents' answers. In spite of the increased privacy, some respondents still refuse to fill out the confidential questionnaire (Becker 2006; Becker and Günther 2004).

5.3 Bystander effects

Another area of research on sensitive topics investigates the impact of bystanders (Aquilino 1997; Aquilino et al. 2000; Hartmann 1995; Reuband 1987, 1992; Smith 1997). Aquilino (1997) formulated a model and derived propositions on third-party effects in the interview situation. According to his considerations, the strength and direction of bystander (e.g. parents, spouse, siblings or children) effects is moderated by the following three factors: (a) the type of question, either asking for subjective (e.g. opinions) versus factual information (e.g. behavior), (b) the bystander's prior knowledge of the information, and (c) the subjective probability of negative consequences as a result of disclosing the sensitive information. If the question is factual and the bystander has no information about the respondent's true status or if the question asks about providing subjective information, the strength of the bystander effect will depend on the subjective probability of negative sanctions as a result of the bystander overhearing the sensitive information. Facing a high subjective probability of negative repercussion, third-party presence will elicit less socially undesirable answers. In contrast, if the survey question is factual and the bystander already knows the true status of the respondent, bystander presence is assumed to not affect self-reports or possibly generate more accurate self-reports. In an experimental study, Aquilino et al. (2000) empirically investigated the impact of the presence of third persons in the interview situation. Their empirical findings were consistent with the predictions of Aquilino (1997).

Several empirical studies focused on the effect of parental presence in surveys of adolescents asking sensitive questions about tobacco, alcohol, drug use, sexual intercourse and violence. They found lower levels of self-reporting of these illegal or socially stigmatized behaviors when the survey questionnaire was completed at home compared to a school setting (Brener et al. 2006; Gfroerer et al. 1997; Kann et al. 2002; Rootman and Smart 1985). Youths tend to deny or downplay drug use when their parents are involved in the survey and

at the same time show the propensity to overreport when the survey is instead conducted at their school and peers are present: “Indeed, one interpretation of consistent prevalence differences across school and home settings over time may be that the school setting fosters exaggeration, while the home setting fosters underreporting” (Fendrich and Johnson 2001, p. 635).

Tourangeau and Yan (2007, pp. 868–869) conducted a meta-analysis to quantify the average effect of bystanders’ presence on sensitive survey reporting. They included 9 non-experimental surveys focusing on either the general population or special populations like adolescents comparing answers given in the presence of a bystander with those given without a bystander’s presence. The empirical studies asked questions on various sensitive topics like drug use, smoking, alcohol consumption, voting behavior and attitudes on cohabitation and separation. Separate mean effects for the presence of the respondent’s spouse and parents respectively were estimated. The results for the presence of a spouse indicated overall lower reporting of sensitive characteristics when the spouse was absent; whereas the estimated mean effect size was not significant. In contrast, the presence of parents generated more socially desirable answers. The findings on the average impact of the parental presence indicate a strong and significant effect. Furthermore, the single effect sizes of the parental presence were consistent in direction and replicable across studies indicating a robust result.

5.4 Question wording and question context

Sensitive surveys often attempt to formulate and to present questions in a neutral way to lower respondents’ concerns about how the admission of a certain behavior will be judged. Researchers often write sensitive questions using unthreatening, euphemistic, familiar and forgiving words or phrases. In his humorous article, “Asking the embarrassing question”, Barton (1958) gives an overview of different possibilities to ask sensitive questions in non-embarrassing ways. Many textbooks on survey research and questionnaire design give general wisdom rules how to write and contextualize sensitive questions (Bradburn and Sudman 1979; Fowler 1995, pp. 28–45; Groves et al. 2004, pp. 230–232; Lee 1993, pp. 75–79; Sudman and Bradburn 1982). Most of these recommendations are based on survey practice. A recent experimental study conducted by Näher and Krumpal (2011) shows inconsistent results. In addition to question wording, appropriateness and question context seems to matter. Fowler (1995) recommends making sure that the sensitive question is appropriate for a certain respondent: “Researchers should be asking people questions only when there is a clear role for the answers in addressing the research questions” (Fowler 1995, p. 34). Other approaches recommend embedding the sensitive question in a series of questions starting with unoffending general questions connected to the topic of interest, and then gradually narrowing the focus to more specific behaviors. Carefully embedding the sensitive question in a carefully constructed context is assumed to lower the respondent’s feelings of jeopardy and to “reduce the focus on a specific behavior question” (Sudman and Bradburn 1982, p. 61).

Other contextual features assumed to affect reporting in sensitive surveys are confidentiality and data protection assurances. Many questionnaire introductions contain such assurances to increase respondents’ trust in data protection and to induce cooperation. Singer et al. (1995) reviewed the experimental literature on the effects of confidentiality assurances. They found lower item nonresponse, higher response rates and higher response accuracy for sensitive items (mostly illegal or socially disapproved behavior, but also income) in studies involving confidentiality assurances, although the average effect size was small. Overall, data protection assurances seem to reduce respondents’ concerns and to improve response quality. However, some positions argue that confidentiality assurances that are too elaborate and sophisticated

might have unintended effects in terms of heightening respondents' suspicions and concerns about who might get access to the data. In three experimental studies on nonsensitive topics, [Singer et al. \(1992\)](#) found higher nonresponse rates for sophisticated data protection assurances compared to shorter assurances or none at all.

5.5 The bogus pipeline procedure

There is some empirical evidence that a data collection strategy called the 'bogus pipeline procedure' increases respondents' motivation to report potentially self-discreditable information more accurately. The term 'bogus pipeline' refers to any methodology, in which respondents believe that an objective procedure (e.g. lie detector, biochemical test) will be used to reveal false self-reports independent of whether such verification actually takes place or not ([Akers et al. 1983](#); [Campanelli et al. 1987](#); [Jones and Sigall 1971](#); [Roese and Jamieson 1993](#)). In terms of SEU-theory, the rationale of bogus pipeline is to increase the respondent's subjective costs of misreporting. Being unmasked as a liar involved in some socially undesirable activities is assumed to generate more embarrassment than simply admitting to a frowned-upon behavior like using drugs or stealing. For the measurement of sensitive attitudes [Roese and Jamieson \(1993, p. 364\)](#) argue: "Hence, the BPL [bogus pipeline] was predicated on the motivational assumption that a desire to avoid appearing to be a liar or to be self-unaware would supersede the typically assumed tendency to exaggerate possession of favourable traits (...)." The empirical evidence in many cases supports the assumed effectiveness of bogus pipeline, although some studies show unexpected results: In the context of an experimental school study in Michigan, [Campanelli et al. \(1987\)](#) randomly assigned 291 students to either a bogus pipeline (173 students) or control condition (118 students). In addition to the questionnaire completion, saliva samples were collected from each student in the bogus pipeline condition. Furthermore, the same students were announced that the experimental results would enable a check on how accurate their alcohol consumption self-reports were. Contrary to their expectations, [Campanelli et al. \(1987\)](#) found no significant bogus pipeline effect; students in the bogus pipeline condition did not self-report more alcohol use and misuse compared to students in the control condition. [Aguinis et al. \(1993\)](#) conducted a meta-analysis to test whether the use of a bogus pipeline methodology generated more valid self-reports, compared to self-report measures alone, for the assessment of cigarette smoking behavior. The results of the quantitative literature review indicate that, overall, a larger fraction of respondents admitted to smoke frequently, compared to respondents interviewed in the control condition. [Aguinis et al. \(1995\)](#) conducted two meta-analyses of experimental studies on the effectiveness of the bogus pipeline procedure in improving the veracity of adolescents' marijuana and alcohol self-reports. They found no evidence for the bogus pipeline methodology to generate more self-disclosure of marijuana and alcohol consumption compared to the control condition. They explained their unexpected findings partly with the post hoc assumption that adolescents might have perceived the behaviors in question as not socially undesirable. In a meta-analysis of opinion studies, [Roese and Jamieson \(1993\)](#) investigated the impact of bogus pipeline use on self-reporting socially undesirable attitudes like racism and sexism. The research synthesis of 31 experimental studies indicates that the bogus pipeline procedure, overall, reduced socially desirable responding.

5.6 The randomized response technique

The randomized response technique (RRT) guarantees the respondent to maintain privacy via the possibility of randomizing his answer ([Warner 1965](#)). The technique has been refined

and several variants of the original method have been developed and implemented.⁴ All of these variants rely on the principle that the respondent uses a randomizing device, which is ideally under her control, to select which of two (or more) questions she will answer. Only the respondent knows the outcome of the randomizing device (e.g. cards, coins, dice) and whether they answered to the sensitive question or a surrogate. Since the interviewer is unaware of the outcome of the random experiment, a given answer does not reveal anything definite about the respondent's true status. Given the assumption that respondents understand the RRT scheme and comply with the RRT procedure, more accurate self-reports to sensitive questions are expected compared to direct questioning.

In the following, Warner's original scheme is described to illustrate the rationale of RRT schemes in general: The respondent is confronted with two statements, the socially undesirable one (e.g., 'I sometimes smoke marijuana') and its negation (e.g., 'I never smoke marijuana'). The interviewer asks 'Do you agree with the following statement?' Using a randomizer, the respondent determines which of the two statements he will answer. For example, the respondent may be given a box of 9 coloured marbles, 5 yellow and 4 blue marbles, and told to take one marble out of the box and to respond to the first statement if a yellow marble is selected, but to respond to the second statement if a blue marble is selected. Without revealing the outcome of the random experiment to the interviewer, the respondent answers with either a 'yes' or a 'no' according to his marijuana smoking habits.

In Warner's design, the prevalence of the socially undesirable behavior can be estimated on the basis of elementary probability theory: The expected value ϕ of observing a 'yes' answer can be modelled as $\phi = p\pi + (1 - p)(1 - \pi)$, where π is the unknown proportion of marijuana smokers in the population, and $p(p \neq 0.5)$ is the probability that the statement 'I sometimes smoke marijuana' is selected. Since the observed sample proportion of 'yes' answers is an estimate of ϕ , and the selection probability p is given by design, the population prevalence of the socially undesirable behavior π can be estimated. Such probabilistic link between the observed answer and the respondent's true status is also at the heart of alternative RRT schemes (overviews of proportion and variance estimators for different RRT schemes can be found in [Fox and Tracy 1986](#)). Furthermore, the relationship between explanatory variables and the socially undesirable characteristic is of interest. Adapted logistic regression models allow for the analysis of the relationship between a response variable measured by the RRT and background variables ([Maddala 1983](#); [Scheers and Dayton 1988](#)).

Several experimental studies comparing results generated by the RRT with alternative data collection methods indicate that the RRT provides more valid estimates of stigmatized, illegal or socially undesirable behavior like drug use (RRT versus direct questioning; [Goodstadt and Gruson 1975](#)), child abuse (RRT versus self-administered interview using the sealed-envelope technique versus classical mail survey; [Zdep and Rhodes 1976](#)), premature sign-offs on audits (RRT versus self-administered questionnaire; [Buchman and Tracy 1982](#); [Reckers et al. 1997](#)), academic cheating behavior among students (RRT versus self-administered questionnaire; [Scheers and Dayton 1987](#)) and abortion (RRT versus face-to-face interview versus audio computer-assisted self-interview versus self-administered questionnaire; [Lara et al.](#)

⁴ Development of RRT schemes in chronological order: (1) Warner's original method: statement and negation of the statement procedure ([Warner 1965](#)); (2) the unrelated question technique with unknown population prevalence of the innocuous attribute ([Greenberg et al. 1969, 1971](#); [Horvitz et al 1967](#)); (3) the forced response technique ([Boruch 1971](#)); (4) Moor's procedure ([Moors 1971](#)); (5) Kuk's 'two packs of cards' technique ([Kuk 1990](#)); (6) two-stage RRT formats ([Mangat 1994](#); [Mangat and Singh 1990](#)).

Also, another body of statistical literature shows considerable progress in the optimization and enhancement (e.g. improving efficiency) of RRT estimators: [Bellhouse \(1980\)](#), [Bourke and Moran \(1988\)](#), [Dowling and Shachtman \(1975\)](#), [Folsom et al. \(1973\)](#), [Liu and Chow \(1976\)](#), [Loynes \(1976\)](#), [O'Hagan \(1987\)](#), [Pollock and Bek \(1976\)](#), [Sen \(1974\)](#) and [Tamhane \(1981\)](#).

2004). In addition, validation studies comparing responses generated by different methods to individual 'true scores' determined from administrative or medical records, confirm the effectiveness of the RRT in reducing response bias in socially sensitive self-reports like self-reported arrests (RRT versus direct questioning; Tracy and Fox 1981) and welfare benefit fraud (RRT versus computer-assisted self-interview versus face-to-face direct questioning; Van der Heijden et al. 2000). Lensvelt-Mulders et al. (2005) conducted a meta-analysis of 6 validation studies and 32 experimental studies without validation data comparing the RRT with other interview methods such as computer-assisted self-interviews or face-to-face direct questioning. Overall, their results indicate that self-reports on sensitive issues are more accurate and more socially undesirable answers are elicited when RRT is employed.

However, other studies have found no superiority of the RRT and standard direct questioning sometimes elicited more socially undesirable answers than did the RRT (for an overview see Holbrook and Krosnick 2010a). Furthermore, some studies yielded evidence suggesting difficulties in making the RRT practical (McAuliffe et al. 1991; Stem and Steinhorst 1984; Weissman et al. 1986). RRT questions impose a higher cognitive burden on the question-and-answer process compared to more conventional data collection methods (Lensvelt-Mulders and Boeije 2007). Some respondents assume a trick and therefore may be confused and distrustful (Boeije and Lensvelt-Mulders 2002; Landsheer et al. 1999; Wiseman et al. 1976). Empirical evidence indicates that a substantial proportion of respondents do not comply with the RRT instructions. They give self-protective 'no'-answers regardless of the outcome of the random device. New developments in the analysis of RRT data account for such self-protective response behavior (Cruyff et al. 2007; Ostapczuk et al. 2009; Coutts and Jann 2011). Alternatively, the crosswise model was proposed to overcome some of the drawbacks of the RRT (Yu et al. 2008; Jann et al. 2011; Coutts et al. 2011).

5.7 The unmatched count technique

The unmatched count technique (UCT) was developed as an alternative to the RRT to protect the respondent's privacy in sensitive surveys. Several other labels, which all refer to the same method, can be found in the literature: 'Block total response' (Raghavarao and Federer 1979; Smith et al. 1974), 'item count technique' (Chaudhuri and Christofides 2007; Droitcour et al. 1991; Tsuchiya 2005), 'unmatched count technique' (Dalton et al. 1994; Coutts and Jann 2011) or 'unmatched block count' (Dalton et al. 1997).

Two subsamples of respondents are generated via randomization. One group of respondents is asked to answer to a short list (SL) of items including only a set of innocuous items (j). The other group of respondents is requested to respond to a long list (LL) of items including the same set of innocuous items plus the sensitive item ($j + 1$). For example, to estimate the prevalence of bilking the following list of items could be used:

1. Have you been to a restaurant, café or bar during the last year? (SL and LL)
2. Have you had dinner in a three-star restaurant during the last four years? (SL and LL)
3. Have you ever left a restaurant or café without paying the bill on purpose? (sensitive item, LL only)
4. Have you run a restaurant by yourself during the last five years? (SL and LL)

Without telling the interviewer which specific items were answered 'yes', respondents in both groups count the number of 'yes'-answers and report solely the sum of items with 'yes'-answers. Since only the number of items in which the respondent was involved is reported, it is not possible to infer whether the respondent engaged in the stigmatized behavior unless 'yes'-answers were given to all or none of the items in the list. The procedure is expected to heighten

the respondents' sense of privacy thus eliciting more socially undesirable answers compared to standard direct questioning. An unbiased estimate of the population's proportion involved in the norm-violating behavior can be obtained by calculating the difference between the two subsample means $\hat{\pi} = \bar{x}_{j+1} - \bar{x}_j$, where \bar{x}_{j+1} is the observed sample mean of 'yes'-responses in the group answering to the long list (LL) and \bar{x}_j is the observed sample mean of 'yes'-responses in the subsample answering to the short list (SL). A double-lists variant of the UCT produces a more efficient estimator compared to the basic procedure (overviews of proportion and variance estimators for different UCT schemes can be found in [Biemer et al. 2005](#); [Droitcour et al. 1991](#); [Tsuchiya et al. 2007](#)).

In several empirical studies, the UCT was applied to estimate the proportion of persons involved in socially undesirable activities (for an overview see [Holbrook and Krosnick 2010b](#)). Some studies found significantly higher proportions of stigmatizing self-reports in the UCT condition compared to direct questioning for behaviors like employee theft ([Wimbush and Dalton 1997](#)), sexual risk behavior after drinking among college students ([LaBrie and Earleywine 2000](#)), hate crime victimization among college students ([Rayburn et al. 2003](#)), eating disordered behavior and attitudes ([Anderson et al. 2007](#)), and shoplifting ([Tsuchiya et al. 2007](#)). However, studies on cocaine use prevalence found significantly lower survey estimates for the UCT compared to standard direct questioning ([Biemer and Brown 2005](#); [Biemer et al. 2005](#)). The reason for failure of the UCT to yield more accurate estimates of cocaine use could have resulted from measurement errors due to problems in cognitive processing: "Deciding which items apply and keeping a running tally of these as the list is read has proven difficult for some respondents." ([Biemer et al. 2005](#), p. 150).

5.8 The nominative technique

The 'nominative technique' (NT) is a variant of the multiplicity methods developed by Sirken ([Sirken 1970, 1975](#); [Sirken et al. 1975](#)). The NT requires the interviewee to serve as an informant by reporting about threatening or illegal behaviors of other persons (e.g. relatives or friends) which are 'nominated' by the interviewee. The anonymity of the persons whose behavior is reported is guaranteed because the interviewer is ignorant about whom the incriminating information is being provided ([Lee 1993](#)). One attempt to estimate the prevalence of socially undesirable behavior ('being intoxicated' and 'smoking marijuana') via NT can be seen in [Sudman et al. \(1977b\)](#). Furthermore, the National Surveys on Drug Abuse 1977, 1979, and 1982 (conducted in the U.S.) used a variant of the NT to estimate the lifetime prevalence of heroin use. As a result, the NT yielded higher estimates of lifetime heroin prevalence compared to the corresponding self-report estimates. The NT version asked the respondent to report, first, how many of his or her close friends have ever used heroin, and second, how many other close friends of each reported heroin user also knew that he or she used heroin. The second question estimates the number of persons which could also report that heroin user and thus allows for the calculation of weights that correct for multiple reports/counts of a particular heroin user ([Droitcour 1985](#), p. 108). Each reported heroin user is weighed inversely to his probability of being reported: The reporting weight is defined as the inverse of the total number of subjects T in the population who are eligible to report a specific heroin user i : $\frac{1}{1+T_{ij}}$.

For each respondent j , the reporting weight must be attached to each report of a heroin-using friend i . In the case of a complete census of the population, weighted individual reports of heroin users could simply be added for all users in the population, so that the total sum of weighted counts of heroin users would equal the total number of heroin users in

the population. In order to obtain the population's proportion of heroin users, the total sum of weighted counts has to be divided by the total population size. The same logic could be applied to the calculation of survey estimates from sample data.

One advantage of the NT is smaller sampling error of estimates compared to estimates based on self-reports. Reported clusters of x 'nominated' friends provide an effective sample size between one and x times the individual sample size. The intra-cluster correlation ρ among reports i within respondents j will determine the decrease in sampling variance of survey estimates based on NT. A decreasing ρ , measuring the degree of homogeneity of reports within respondents, increases the effective sample size and decreases the sampling variance of survey estimates based on clusters (Kish 1965). However, sampling error is only one dimension of total response error. Response accuracy is questionable because many respondents don't know the number of persons who also know about the heroin use of the respondent's close friends: "In fact, the nominative approach might tend to produce over-estimates, because of the potential for undercounts of the numbers of others who 'know' (Droitcour 1985, p. 116). Future survey studies should conduct additional validity tests of the NT by comparing sensitive behaviors with non-sensitive behaviors. The assumption 'higher validity in the NT condition' would be supported if the NT estimates yielded higher prevalence rates for sensitive behaviors and prevalence estimates similar to results based on self-reporting for non-sensitive behaviors.

6 Summary and perspectives

The review of the theoretical and empirical literature on the determinants of social desirability bias identified several factors assumed to affect the respondent's propensity to self-report accurately on sensitive topics. On the one hand, the concept 'sensitivity' refers to certain behaviors that are taboo, illegal or socially sanctioned. On the other hand, the term also encompasses unsocial attitudes and opinions. A respondent's confession of being involved in activities that clearly violate social norms could either cause embarrassment in the interview situation or result in legal or social sanctions in the case that the sensitive information would become public. Anticipating such risks and threats, survey respondents often choose to misreport on sensitive topics or not to answer at all. The respondent's need for social approval, self-presentation concerns and impression management strategies yield socially desirable responses on the individual level and a predictable bias in survey estimates on the aggregate level. In the case of socially undesirable activities, sample proportions will underestimate the true prevalence and frequency of the frowned upon activities in question. In the case of socially desirable behavior, an overestimation of the true level will occur. The degree of social desirability bias in sensitive surveys depends on the perceived items' sensitivity, the degree of privacy in the interview situation, the fraction of the population who have behaved in the socially undesirable manner, and on specific aspects of the survey design.

Improving the validity of answers to sensitive questions has proved a difficult task. Social desirability bias could be reduced by appropriately tailoring the survey design. The literature review of the recent research indicates that cognitive psychologists, social scientists and survey statisticians have made some progress in reducing measurement errors due to deliberate misreporting on sensitive topics, principally by increasing the anonymity of the question-and-answer process (e.g. via the randomized response technique or self-administered interviews), by decreasing the respondent's concerns in admitting to some taboo (e.g. via confidentiality assurances or clever wording and framing of the sensitive item), by increasing the respondent's subjective probability of being caught as a liar (e.g. via the bogus pipeline procedure),

by heightening the respondent's subjective benefit of telling the truth (e.g. via emphasizing the importance and scientific character of the survey), or by manipulating the survey situation (such as reducing the presence of the interviewer and bystanders).

Future research on sensitive topics could investigate the interaction effect between psychological variables and design aspects on survey response in more detail. Such research would deepen our understanding of the cognitive mechanisms underlying the link between survey design and survey response. For example, the literature review shows that the effectiveness of the bogus pipeline procedure is still controversial. Psychological variables such as the perceived level of the items' sensitivity or trust in the institution conducting the survey could moderate the effectiveness of these methods. Therefore, prospective survey studies are encouraged to advance the approaches of Bradburn and Sudman (1979) and Rasinski et al. (1994, 1999) and to include rating-scales measuring the respondents' and the interviewers' perceived sensitivity of the overall survey situation and the perceived sensitivity of specific items. It would be possible to identify specific subgroups which are characterized by different degrees of sensitivity perceptions and to study the complex interactions between these perceptions and the survey design.

Finally, a stronger foundation of the research on sensitive topics in a general theory of human action could further illuminate the social mechanisms operating in the interview situation. The survey interview involves social interactions between several actors (respondents, interviewers, bystanders, and data collection institutions). The impact of varying degrees of anonymity in interactive social situations could be analyzed by means of rational choice theory (e.g. Levitt and List 2007). A clearer understanding of the social interactions in the data collection process and the effects of these interactions on data quality could provide applied researchers with a substantiated basis for designing better data collection instruments and conducting high quality surveys.

References

- Aguinis, H., Pierce, C.A., Quigley, B.M.: Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette-smoking—a meta-analytic review. *J. Appl. Soc. Psychol.* **23**, 352–373 (1993)
- Aguinis, H., Pierce, C.A., Quigley, B.M.: Enhancing the validity of self-reported alcohol and marijuana consumption using a bogus pipeline procedure—a metaanalytic review. *Basic Appl. Soc. Psychol.* **16**, 515–527 (1995)
- Akers, R.L., Massey, J., Clarke, W., Lauer, R.M.: Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. *Soc. Forc.* **62**, 234–251 (1983)
- Anderson, B.A., Silver, B.D., Abramson, P.R.: The effects of race of the interviewer on measures of electoral-participation by blacks in SRC national elections studies. *Publ. Opin. Q.* **52**, 53–83 (1988)
- Anderson, D.A., Simmons, A.M., Milnes, S.M., Earleywine, M.: Effect of response format on endorsement of eating disordered attitudes and behaviors. *Int. J. Eat. Disord.* **40**, 90–93 (2007)
- Aquilino, W.S.: Privacy effects on self-reported drug use: interactions with survey mode and respondent characteristics. In: Harrison, L., Hughes, A. (eds.) *The Validity of Self-Reported Drug Use: Improving the Accuracy Of Survey Estimates*. National Institute on Drug Abuse Monograph 167, NIH, DHHS, Washington (1997)
- Aquilino, W.S., Wright, D.L., Supple, A.J.: Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Subst. Use Misuse* **35**, 845–867 (2000)
- Barnett, J.: Sensitive questions and response effects: an evaluation. *J. Manag. Psychol.* **13**, 63–76 (1998)
- Barton, A.H.: Asking the embarrassing question. *Publ. Opin. Q.* **22**, 67–68 (1958)
- Becker, R.: Selective response to questions on delinquency. *Qual. Quant.* **40**, 483–498 (2006)
- Becker, R., Günther, R.: Selektives Antwortverhalten bei Fragen zum delinquenten Handeln—Eine empirische Studie über die Wirksamkeit der, sealed envelope technique“ bei selbstberichteter Delinquenz mit Daten des ALLBUS 2000. *ZUMA-Nachrichten* **54**, 39–59 (2004)

- Bellhouse, D.R.: Linear models for randomized response design. *J. Am. Stat. Assoc.* **75**, 1001–1004 (1980)
- Beyer, H., Krumpal, I.: “Aber es gibt keine Antisemiten mehr”: Eine experimentelle Studie zur Kommunikationslatenz antisemitischer Einstellungen. *Kölner Z. für Soziol. und Sozialpsychol.* **62**, 681–705 (2010)
- Biemer, P., Brown, G.: Model-based estimation of drug use prevalence using item count data. *J. Off. Stat.* **21**, 287–308 (2005)
- Biemer, P., Jordan, B.K., Hubbard, M.L., Wright, D.: A test of the item count methodology for estimating cocaine use prevalence. In: Kennet, J., Gfroerer, J. (eds.) *Evaluating and Improving Methods Used in the National Survey on Drug use and Health, Substance Abuse and Mental Health Service Administration, Office of Applied Studies, Rockville* (2005)
- Boeije, H., Lensvelt-Mulders, G.J.L.M.: Honest by chance: a qualitative interview study to clarify respondents’ (non-)compliance with computer-assisted randomized response. *Bull. Methodol. Sociol.* **75**, 24–39 (2002)
- Boruch, R.F.: Assuring confidentiality of responses in social research: a systematic analysis. *A. Psychol.* **26**, 413–430 (1971)
- Bourke, P.D., Moran, M.A.: Estimating proportions from randomized response using the EM algorithm. *J. Am. Stat. Assoc.* **83**, 964–968 (1988)
- Bradburn, N.M., Sudman, S.: *Improving Interview Method and Questionnaire Design*. Jossey-Bass, San Francisco (1979)
- Brener, N.D., Eaton, D.K., Kann, L., Grunbaum, J.A., Gross, L.A., Kyle, T.M., Ross, J.G.: The association of survey setting and mode with self-reported health risk behaviors among high school students. *Publ. Opin. Q.* **70**, 354–374 (2006)
- Buchman, T.A., Tracy, J.A.: Obtaining responses to sensitive questions: conventional questionnaire versus randomized response technique. *J. Account. Res.* **20**, 263–271 (1982)
- Campanelli, P.C., Dielman, T.E., Shope, J.T.: Validity of adolescents self-reports of alcohol-use and misuse using a bogus pipeline procedure. *Adolescence* **22**, 7–22 (1987)
- Chaudhuri, A., Christofides, T.C.: Item Count Technique in estimating the proportion of people with a sensitive feature. *J. Stat. Planning Infer.* **137**, 589–593 (2007)
- Coutts, E., Jann, B.: Sensitive questions in online surveys: experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociol. Methods Res.* **40**, 169–193 (2011)
- Coutts, E., Jann, B., Krumpal, I., Näher, A.-F.: Plagiarism in student papers: prevalence estimates using special techniques for sensitive questions. *J. Econ. Stat.* **231**, 749–760 (2011)
- Crowne, D., Marlowe, D.: A new scale of social desirability independent of psychopathology. *J. Consult. Psychol.* **24**, 349–354 (1960)
- Crowne, D., Marlowe, D.: *The Approval Motive*. John Wiley, New York (1964)
- Cruyff, M.J.L.F., van den Hout, A., van der Heijden, P.G.M., Bockenholt, U.: Log-linear randomized-response models taking self-protective response behavior into account. *Sociol. Methods Res.* **36**, 266–282 (2007)
- Dalton, D.R., Wimbush, J.C., Daily, C.M.: Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior. *Pers. Psychol.* **47**, 817–828 (1994)
- Dalton, D.R., Daily, C.M., Wimbush, J.C.: Collecting “sensitive” data in business ethics research: a case for the unmatched count technique (UCT). *J. Bus. Ethics* **16**, 1049–1057 (1997)
- De Leeuw, E.D.: Reducing missing data in surveys: an overview of methods. *Qual. Quant.* **35**, 147–160 (2001)
- De Leeuw, E.D., Hox, J.J., Dillman, D.A.: Mixed mode surveys: when and why? In: De Leeuw, E.D., Hox, J.J., Dillman, D.A. (eds.) *The International Handbook of Survey Methodology*, Erlbaum/Taylor & Francis, New York (2008)
- Decker, O., Brähler, E.: *Vom Rand zur Mitte: Rechtsextreme Einstellungen und ihre Einflussfaktoren in Deutschland*. Friedrich-Ebert-Stiftung, Forum Berlin (2006)
- DeMaio, T.J.: Social desirability and survey measurement: a review. In: Turner, C.F., Martin, E. (eds.) *Surveying Subjective Phenomena*, pp. 257–281. Russel Sage, New York (1984)
- DePaulo, B.M., Kirkendol, S.E., Kashy, D.A., Wyer, M.M., Epstein, J.A.: Lying in everyday life. *J. Pers. Soc. Psychol.* **70**, 979–995 (1996)
- DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. *Psychol. Bull.* **129**, 74–118 (2003)
- Des Jarlais, D.C., Paone, D., Milliken, J., Turner, C.F., Miller, H., Gribble, J., Shi, Q.H., Hagan, H., Friedman, S.R.: Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: a quasi-randomised trial. *Lancet* **353**, 1657–1661 (1999)
- Dowling, T.A., Shachtman, R.H.: On the relative efficiency of randomized response technique. *J. Am. Stat. Assoc.* **70**, 84–87 (1975)
- Droitcour, J.: The nominative technique: a new method of estimating heroin prevalence. In: Rouse, B.A., Kozel, N.J., Richards, L.G. (eds.) *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity*, Fishers Lane, pp. 104–124. National Institute on Drug Abuse (NIDA), Rockville (1985)

- Droitcour, J., Caspar, R.A., Hubbard, M.L., Parsely, T.L., Visscher, W., Ezzati, T.M.: The item count technique as a method of indirect questioning: a review of its development and a case study application. In: Biemer, P., Groves, R.M., Lyberg, L., Mathiowetz, N., Sudman, S. (eds.) *Measurement Errors in Surveys*, pp. 85–210. Wiley, New York (1991)
- Esser, H.: Können Befragte lügen?—Zum Konzept des “wahren Wertes” im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *Kölner Zeitschrift für Soziologie Und Sozialpsychologie* **38**, 314–336 (1986)
- Fendrich, M., Johnson, T.P.: Examining prevalence differences in three national surveys of youth: impact of consent procedures, mode, and editing rules. *J. Drug Issues* **31**, 615–642 (2001)
- Folsom, R.E., Greenberg, B.G., Horvitz, D.G., Abernathy, J.R.: The two alternate questions randomized response model for human surveys. *J. Am. Stat. Assoc.* **68**, 525–530 (1973)
- Fowler, F.J. Jr.: *Improving Survey Questions: Design and Evaluation*. Sage, Thousand Oaks (1995)
- Fowler, F.J. Jr., Mangione, T.W.: *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Sage, Newbury Park (1990)
- Fox, J.A., Tracy, P.E.: *Randomized Response: A Method for Sensitive Surveys*. Sage, Beverly Hills (1986)
- Gfroerer, J., Wright, D., Kopstein, A.: Prevalence of youth substance use: the impact of methodological differences between two national surveys. *Drug Alcohol Depend.* **47**, 19–30 (1997)
- Goodstadt, M.S., Gruson, V.: The randomized response technique: a test of drug use. *J. Am. Stat. Assoc.* **70**, 814–818 (1975)
- Greenberg, B.G., Abul-El, A.-L.A., Simmons, W.R., Horvitz, D.G.: The unrelated question randomized response model for human surveys. *J. Am. Stat. Assoc.* **64**, 520–539 (1969)
- Greenberg, B.G., Kuebler, R.R. Jr., Abernathy, J.R., Horvitz, D.G.: Application of the randomized response technique in obtaining quantitative data. *J. Am. Stat. Assoc.* **66**, 243–250 (1971)
- Gribble, J.N., Miller, H.G., Rogers, S.M., Turner, C.F.: Interview mode and measurement of sexual behaviors: methodological issues. *J. Sex Res.* **36**, 16–24 (1999)
- Groves, R.M.: *Survey Errors and Survey Costs*. Wiley, New York (1989)
- Groves, R.M., Fowler, F.J. Jr., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R.: *Survey Methodology*. Wiley, Hoboken (2004)
- Hartmann, P.: Response behavior in interview settings of limited privacy. *Int. J. Publ. Opin. Res.* **7**, 383–390 (1995)
- Holbrook, A.L., Krosnick, J.A.: Measuring voter turnout by using the randomized response technique: evidence calling into question the method’s validity. *Publ. Opin. Q.* **74**, 328–343 (2010a)
- Holbrook, A.L., Krosnick, J.A.: Social desirability bias in voter turnout reports: tests using the item count technique. *Publ. Opin. Q.* **74**, 37–67 (2010b)
- Holbrook, A.L., Green, M.C., Krosnick, J.A.: Telephone versus face-to-face interviewing of national probability samples with long questionnaires - Comparisons of respondent satisficing and social desirability response bias. *Publ. Opin. Q.* **67**, 79–125 (2003)
- Holtgraves, T.: Social desirability and self-reports: testing models of socially desirable responding. *Pers. Soc. Psychol. Bull.* **30**, 161–172 (2004)
- Holtgraves, T., Eck, J., Lasky, B.: Face management, question wording, and social desirability. *J. Appl. Soc. Psychol.* **27**, 1650–1671 (1997)
- Horvitz, D.G., Shah, B.V., Simmons, W.R.: The unrelated question randomized response model. In: *Proceedings of the Social Statistics Section*, pp. 65–72. ASA (1967)
- Hox, J.J., De Leeuw, E.D.: The influence of interviewers’ attitude and behavior on household survey nonresponse: an international comparison. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (eds.) *Survey Nonresponse*, pp. 103–120. Wiley, New York (2002)
- Jann, B., Jerke, J., Krumpal, I.: Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism. *Publ. Opin. Q.* (2011). doi:[10.1093/poq/nfr036](https://doi.org/10.1093/poq/nfr036)
- Johnson, T., Vijver, F.J. van de: Social desirability in cross-cultural research. In: Harness, J., Vijver, F.J. van de, Mohler, P. (eds.) *Cross-Cultural Survey Methods*, pp. 193–202. Wiley, New York (2002)
- Jones, E.E., Sigall, H.: Bogus pipeline—new paradigm for measuring affect and attitude. *Psychol. Bull.* **76**, 349–354 (1971)
- Kann, L., Brener, N.D., Warren, C.W., Collins, J.L., Giovino, G.A.: An assessment of the effect of data collection setting on the prevalence of health risk behaviors among adolescents. *J. Adolesc. Health* **31**, 327–335 (2002)
- Katz, D.: Do interviewers bias poll results? *Publ. Opin. Q.* **6**, 248–268 (1942)
- Kish, L.: *Survey Sampling*. Wiley, New York (1965)
- Krysan, M.: Privacy and the expression of white racial attitudes—a comparison across three contexts. *Publ. Opin. Q.* **62**, 506–544 (1998)
- Kuk, A.Y.C.: Asking sensitive questions indirectly. *Biometrika* **77**, 436–438 (1990)

- LaBrie, J.W., Earleywine, M.: Sexual risk behaviors and alcohol: higher base rates revealed using the unmatched-count technique. *J. Sex Res.* **37**, 321–326 (2000)
- Lalwani, A.K., Shavitt, S., Johnson, T.: What is the relation between cultural orientation and socially desirable responding?. *J. Pers. Soc. Psychol.* **90**, 165–178 (2006)
- Landsheer, J.A., Van der Heijden, P.G.M., Van Gils, G.: Trust and understanding, two psychological aspects of randomized response. *Qual. Quant.* **33**, 1–12 (1999)
- Lara, D., Strickler, J., Olavarrieta, C.D., Ellertson, C.: Measuring induced abortion in Mexico. A comparison of four methodologies. *Soc. Methods Res.* **32**, 529–558 (2004)
- Lee, R.M.: *Doing Research on Sensitive Topics*. Sage, London (1993)
- Lee, R.M., Renzetti, C.M.: The Problems of Researching Sensitive Topics: An Overview and Introduction. In: Renzetti, C.M., Lee, R.M. (eds.) *Researching Sensitive Topics*, Sage, London (1993)
- Lensvelt-Mulders, G.J.L.M., Boeije, H.R.: Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating. *Comput. Hum. Behav.* **23**, 591–608 (2007)
- Lensvelt-Mulders, G.J.L.M.: Surveying sensitive topics. In: De Leeuw, E.D., Hox, J.J., Dillman, D.A. (eds.) *The international Handbook of Survey Methodology*, Erlbaum/Taylor & Francis, New York/London (2008)
- Lensvelt-Mulders, G.J.L.M., Hox, J.J., Heijden, P.G.M. van der, Mass, C.J.M.: Meta-analysis of randomized response research. thirty-five years of validation. *Sociol. Methods Res.* **33**, 319–348 (2005)
- Levitt, S.D., List, J.A.: What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* **21**, 153–174 (2007)
- Liu, P.T., Chow, L.P.: The efficiency of the multiple trial randomized response technique. *Biometrics* **32**, 607–618 (1976)
- Loynes, R.M.: Asymptotically optimal randomized response procedures. *J. Am. Stat. Assoc.* **71**, 924–928 (1976)
- Maddala, G.S.: *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York (1983)
- Mangat, N.S.: An improved randomized response strategy. *J. R. Stat. Soc B (Methodol.)* **56**, 93–95 (1994)
- Mangat, N.S., Singh, R.S.: An alternative randomized response procedure. *Biometrika* **77**, 439–442 (1990)
- Marquis, K.H., Duan, N., Marquis, M.S., Polich, J.M.: *Response Errors in Sensitive Topics Surveys*. The Rand Corporation, CA (1981)
- Marquis, K.H., Marquis, M.S., Polich, J.M.: Response bias and reliability in sensitive topic surveys. *J. Am. Stat. Assoc.* **81**, 381–389 (1986)
- McAuliffe, W.E., Breer, P., Ahmadifar, N.W., Spino, C.: Assessment of drug abuser treatment needs in Rhode Island. *Am. J. Publ. Health* **81**, 365–371 (1991)
- Metzger, D.S., Koblin, B., Turner, C., Navaline, H., Valenti, F., Holte, S., Gross, M., Sheon, A., Miller, H., Cooley, P., Seage, G.R.: Randomized controlled trial of audio computer-assisted self-interviewing: utility and acceptability in longitudinal studies. *Am. J. Epidemiol.* **152**, 99–106 (2000)
- Moors, J.J.A.: Optimization of the unrelated question randomized response model. *J. Am. Stat. Assoc.* **66**, 627–629 (1971)
- Nathan, G., Sirken, M., Willis, G.B., Esposito, J.: Laboratory experiments on the cognitive aspects of sensitive questions. In: *International Conference on Measurement Error in Surveys*. Tuscon, Arizona (1990)
- Näher, A.-F., Krumpal, I.: Asking sensitive questions: the impact of forgiving wording and question context on social desirability bias. *Qual. Quant.* (2011). doi:[10.1007/s11135-011-9469-2](https://doi.org/10.1007/s11135-011-9469-2)
- O'Hagan, A.: Bayes linear estimators for randomized response models. *J. Am. Stat. Assoc.* **82**, 580–585 (1987)
- Okamoto, K., Ohsuka, K., Shiraishi, T., Hukazawa, E., Wakasugi, S., Furuta, K.: Comparability of epidemiological information between self- and interviewer-administered questionnaires. *J. Clin. Epidemiol.* **55**, 505–511 (2002)
- Ong, A.D., Weiss, D.J.: The impact of anonymity on responses to sensitive questions. *J. Appl. Soc. Psychol.* **30**, 1691–1708 (2000)
- Ostapczuk, M., Musch, J., Moshagen, M.: A randomized-response investigation of the education effect in attitudes towards foreigners. *Eur. J. Soc. Psychol.* **39**, 920–931 (2009)
- Paulhus, D.L.: Measurement and control of response bias. In: *Measures of personality and social psychological attitudes*, vol. 1. San Diego, CA: Academic Press (1991)
- Paulhus, D.L.: Self-presentation measurement. In: Fernandez-Ballesteros, R. (ed.) *Encyclopedia of Psychological Assessment*, pp. 858–860. Sage, Thousand Oaks (2003)
- Pollock, K.H., Bek, Y.: A comparison of three randomized response models for quantitative data. *J. Am. Stat. Assoc.* **71**, 884–886 (1976)
- Raghavarao, D., Federer, W.T.: Block total response as an alternative to the randomized response method in surveys. *J. R. Stat. Soc. B Methodol.* **41**, 40–45 (1979)

- Randall, D.M., Fernandes, M.F.: The social desirability response bias in ethics research. *J. Bus. Ethics* **10**, 805–817 (1991)
- Rasinski, K.A., Baldwin, A.K., Willis, G.B., Jobe, J.B.: Risk and Loss Perceptions Associated with Survey Reporting of Sensitive Topics. pp. 497–502. National Opinion Research Center (NORC), Chicago (1994)
- Rasinski, K.A., Willis, G.B., Baldwin, A.K., Yeh, W.C., Lee, L.: Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Appl. Cogn. Psychol.* **13**, 465–484 (1999)
- Rauhut, H., Krumpal, I.: Die Durchsetzung sozialer Normen in low-cost und high-cost situationen. *Z. für Soziol.* **37**, 380–402 (2008)
- Rayburn, N.R., Earleywine, M., Davison, G.C.: Base rates of hate crime victimization among college students. *J. Interpers. Violence* **18**, 1209–1221 (2003)
- Reckers, P.M.J., Wheeler, S.W., Wong-On-Wing, B.: A comparative examination of auditor premature sign-offs using the direct and the randomized response methods. *Audit. J. Pract. Theory* **16**, 69–78 (1997)
- Reuband, K.H.: Unerwünschte Dritte beim Interview: Erscheinungsformen und Folgen. *Zeitschrift Für Soziologie* **16**, 303–308 (1987)
- Reuband, K.H.: On 3rd persons in the interview situation and their impact on responses. *Int. J. Publ. Opin. Res.* **4**, 269–274 (1992)
- Robinson, D., Rhode, S.: 2 experiments with an anti-semitism poll. *J. Abnorm. Soc. Psychol.* **41**, 136–144 (1946)
- Roese, N.J., Jamieson, D.W.: 20 years of bogus pipeline research—a critical-review and metaanalysis. *Psychol. Bull.* **114**, 363–375 (1993)
- Rootman, I., Smart, R.G.: A comparison of alcohol, tobacco and drug-use as determined from household and school surveys. *Drug Alcohol Depend.* **16**, 89–94 (1985)
- Schaeffer, N.C.: Asking questions about threatening topics: a selective overview. In: Stone, A., Turkkan, J., Bachrach, C., Cain, V., Jobe, J., Kurtzman, H. (eds.) *The Science of Self-Report: Implications for Research and Practice*, pp. 105–121. Erlbaum, Mahwah (2000)
- Scheers, N.J., Dayton, C.M.: Improved estimation of academic cheating behavior using the randomized-response technique. *Res. Higher Educ.* **26**, 61–69 (1987)
- Scheers, N.J., Dayton, C.M.: Covariate randomized response models. *J. Am. Stat. Assoc.* **83**, 969–974 (1988)
- Schnell, R., Kreuter, F.: Separating interviewer and sampling-point effects. *J. Off. Stat.* **21**, 389–410 (2005)
- Schuman, H., Converse, J.M.: Effects of black and white interviewers on black responses in 1968. *Publ. Opin. Q.* **35**, 44–68 (1971)
- Sen, P.K.: On unbiased estimation for randomized response models. *J. Am. Stat. Assoc.* **69**, 997–1001 (1974)
- Singer, E., Kohnke-Aquirre, L.: Interviewer expectation effects—replication and extension. *Publ. Opin. Q.* **43**, 245–260 (1979)
- Singer, E., Hippler, H.J., Schwarz, N.: Confidentiality assurances in surveys—reassurance or threat. *Int. J. Publ. Opin. Res.* **4**, 256–268 (1992)
- Singer, E., Vonthurn, D.R., Miller, E.R.: Confidentiality assurances and response—a quantitative review of the experimental literature. *Publ. Opin. Q.* **59**, 66–77 (1995)
- Sirken, M.: Household surveys with multiplicity. *J. Am. Stat. Assoc.* **65**, 257–266 (1970)
- Sirken, M.: Network surveys of rare and sensitive conditions. In: *Advances in Health Survey Research Methods*. National Center on Health Statistics Research Proceedings Series, pp. 31–32 (1975)
- Sirken, M., Indefurth, G.P., Burnham, C.E., Danchik, K.M.: Household sample surveys of diabetes: design effects of counting rules. In: *Proceedings of the American Statistical Association*, pp. 659–663. Social Statistics Section (1975)
- Sirken, M., Willis, G.B., Nathan, G.: Cognitive aspects of answering sensitive survey questions. *Bull. Int. Stat. Inst.* **48**, 628–629 (1991)
- Smith, T.W.: Discrepancies between men and women in reporting number of sexual partners - a summary from 4 countries. *Soc. Biol.* **39**, 203–211 (1992)
- Smith, T.W.: The impact of the presence of others on a respondent's answers to questions. *Int. J. Publ. Opin. Res.* **9**, 33–47 (1997)
- Smith, L.L., Federer, W.T., Raghavarao, D.: A comparison of three techniques for eliciting truthful answers to sensitive questions. In: *Proceedings of the American Statistical Association*, pp. 447–452. Social Statistics Section (1974)
- Stem, D.E., Steinhorst, R.K.: Telephone interview and mail questionnaire applications of the randomized response model. *J. Am. Stat. Assoc.* **79**, 555–564 (1984)
- Stocké, V.: Determinants and consequences of survey respondents' social desirability beliefs about racial attitudes. *Methodology* **3**, 125–138 (2007a)
- Stocké, V.: The interdependence of determinants for the strength and direction of social desirability bias in racial attitude surveys. *J. Off. Stat.* **23**, 493–514 (2007b)

- Stocké, V., Hunkler, C.: Measures of desirability beliefs and their validity as indicators for socially desirable responding. *Field Methods* **19**, 313–336 (2007)
- Sudman, S., Bradburn, N.M.: *Response Effects in Surveys: A Review and Synthesis*. Aldine, Chicago (1974)
- Sudman, S., Bradburn, N.M.: *Asking Questions: A Practical Guide to Questionnaire Design*. Jossey-Bass, San Francisco (1982)
- Sudman, S., Bradburn, N.M., Blair, E., Stocking, C.: Modest expectations—effects of interviewers prior expectations on responses. *Sociol. Methods Res.* **6**, 171–182 (1977a)
- Sudman, S., Blair, E., Bradburn, N., Stocking, C.: Estimates of threatening behavior based on reports of friends. *Publ. Opin. Q.* **41**, 261–264 (1977b)
- Tamhane, A.C.: Randomized response techniques for multiple sensitive attributes. *J. Am. Stat. Assoc.* **76**, 916–923 (1981)
- Tentler, T.N.: *Sin and Confession on the Eve of the Reformation*. Princeton University Press, Princeton (1977)
- Tourangeau, R., Smith, T.W.: Asking sensitive questions—the impact of data collection mode, question format, and question context. *Publ. Opin. Q.* **60**, 275–304 (1996)
- Tourangeau, R., Rasinski, K.A., Jobe, J., Smith, T.W., Pratt, W.F.: Sources of error in a survey on sexual behavior. *J. Off. Stat.* **13**, 341–365 (1997)
- Tourangeau, R., Rips, L.J., Rasinski, K.A.: *The Psychology of Survey Response*. Cambridge University Press, Cambridge (2000)
- Tourangeau, R., Yan, T.: Sensitive questions in surveys. *Psychol. Bull.* **133**, 859–883 (2007)
- Tracy, P.S., Fox, J.A.: The validity of randomized response for sensitive measurements. *Am. Sociol. Rev.* **46**, 187–200 (1981)
- Tsuchiya, T.: Domain estimators for the item count technique. *Survey Methodol.* **31**, 41–51 (2005)
- Tsuchiya, T., Hirai, Y., Ono, S.: A study of the properties of the item count technique. *Publ. Opin. Q.* **71**, 253–272 (2007)
- Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., Sonenstein, F.L.: Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science* **280**, 867–873 (1998)
- Turner, C.F., Villarroel, M.A., Rogers, S.M., Eggleston, E., Ganapathi, L., Roman, A.M., Al-Tayyib, A.: Reducing bias in telephone survey estimates of the prevalence of drug use: a randomized trial of telephone audio-CASI. *Addiction* **100**, 1432–1444 (2005)
- Van der Heijden, P.G.M., Gils, G. van, Bouts, J., Hox, J.J.: 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**, 505–537 (2000)
- Villarroel, M.A., Turner, C.F., Eggleston, E., Al-Tayyib, A., Rogers, S.M., Roman, A.M., Cooley, P.C., Gordke, H.: Same-gender sex in the United States—impact of T-ACASI on prevalence estimates. *Publ. Opin. Q.* **70**, 166–196 (2006)
- Villarroel, M.A., Turner, C.F., Rogers, S.M., Roman, A.M., Cooley, P.C., Steinberg, A.B., Eggleston, E., Chromy, J.R.: T-ACASI reduces bias in STD measurements: the national STD and behavior measurement experiment. *Sex. Transm. Dis.* **35**, 499–506 (2008)
- Warner, S.L.: Randomized response: a survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* **60**, 63–69 (1965)
- Weissman, A.N., Steer, R.A., Lipton, D.S.: Estimating illicit drug use through telephone interviews and the randomized response technique. *Drug Alcohol Depend.* **18**, 225–233 (1986)
- Willis, G.B., Sirken, M., Nathan, G.: The cognitive aspects of responses to sensitive survey questions. In: Working Paper Series 9. Hyattsville, MD: National Center for Health Statistics, Cognitive Methods Staff (1994)
- Wimbush, J.C., Dalton, D.R.: Base rate for employee theft: convergence of multiple methods. *J. Appl. Psychol.* **82**, 756–763 (1997)
- Wiseman, F., Moriarty, M., Schafer, M.: Estimating public-opinion with randomized response model. *Publ. Opin. Q.* **39**, 507–513 (1976)
- Yu, J.W., Tian, G.L., Tang, M.L.: Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika* **67**, 251–263 (2008)
- Zdep, S.M., Rhodes, I.N.: Making the randomized response technique work. *Publ. Opin. Q.* **40**, 513–537 (1976)