

Interviewer effects in public health surveys

R. E. Davis^{1*}, M. P. Couper², N. K. Janz³, C. H. Caldwell³ and K. Resnicow³

Abstract

Interviewer effects can have a substantial impact on survey data and may be particularly operant in public health surveys, where respondents are likely to be queried about racial attitudes, sensitive behaviors and other topics prone to socially desirable responding. This paper defines interviewer effects, argues for the importance of measuring and controlling for interviewer effects in health surveys, provides advice about how to interpret research on interviewer effects and summarizes research to date on race, ethnicity and gender effects. Interviewer effects appear to be most likely to occur when survey items query attitudes about sociodemographic characteristics or respondents' engagement in sensitive behaviors such as substance use. However, there is surprisingly little evidence to indicate whether sociodemographic interviewer–respondent matching improves survey response rates or data validity, and the use of a matched design introduces possible measurement bias across studies. Additional research is needed to elucidate many issues, including the influence of interviewers'

sociodemographic characteristics on health-related topics, the role of within-group interviewer variability on survey data and the simultaneous impact of multiple interviewer characteristics. The findings of such research would provide much-needed guidance to public health professionals on whether or not to match interviewers and respondents on key sociodemographic characteristics.

Introduction

Due in part to a growing emphasis on cultural sensitivity in public health research and practice, public health professionals often debate whether to match survey interviewers to the anticipated sociodemographic characteristics of respondents. As a consequence, there is a need for guidance about whether and how interviewer characteristics influence health survey data.

Herbert Hyman argued that:

...the demonstration of error marks an advanced stage of a science. All scientific inquiry is subject to error, and it is far better to be aware of this, to study the sources in an attempt to reduce it, and to estimate the magnitude of such errors in our findings, than to be ignorant of the errors concealed in the data. [1, p. 4]

To these ends, this paper reviews the state of knowledge of interviewer effects to assist public health professionals in collecting, analysing and interpreting survey data. Studies were only included if they were conducted with adults in the United

¹Department of Environmental Health Sciences, University of Michigan School of Public Health, #6610B SPH Tower, Ann Arbor, MI 48109-2029, USA, ²Program in Survey Methodology, University of Michigan, 4040 Institute for Social Research Ann Arbor, MI 48104, USA and

³Department of Health Behavior and Health Education, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109-2029, USA

*Correspondence to: R. E. Davis.

E-mail: reda@umich.edu

States; however, this paper equips the reader to evaluate research in other populations and settings. And, although interviewer-related measurement error can result from role-dependent effects, such as differences in the way interviewers administer a survey, read questions, probe or provide off-script instructions to respondents [2, 3], this review focuses on how respondents react to the role-independent, perceived interviewer sociodemographic characteristics of race, ethnicity and gender [4]. This paper defines interviewer effects, describes the potential influence of interviewer effects on survey data, outlines aspects to consider in evaluating interviewer effects research, reviews existing research on race/ethnicity and gender effects and highlights implications for public health research and practice.

What is interviewer error?

Error is the difference between the observed value of a phenomenon, which is measured using survey data, and the true value of that phenomenon, which is often impossible to measure [3]. According to the total survey error model put forth by Groves [3], the total error associated with a particular survey statistic is known as the mean square error. The mean square error represents the sum of both biases and variable errors affecting that survey statistic. Both bias and variable errors describe what would be expected to occur over hypothetical, repeated administrations of the same questionnaire with the same sampling design, recruitment protocol and data collection procedures, regardless of whether or not the survey is actually re-administered. A specific administration of the survey is regarded as a sampling of what would happen over these repeated survey administrations. Whereas bias describes errors that are expected to occur on every administration of the same survey design, variable errors describe errors that are expected to vary across survey administrations. Both bias and variable errors may derive from sources of error relating to coverage, nonresponse, sampling, interviewer, respondent, instrument and mode [3].

Interviewer error refers to variance in survey estimates that arises from the fact that data collected by either a specific individual interviewer or a specific set of interviewers may be different than data collected by another individual or set of interviewers administering the same questionnaire to a sample from the same population of respondents. Interviewer errors are usually variable errors related to the particular interviewers selected; however, bias may result if the source of error is rooted in an aspect of the survey design [3], such as the use of only female interviewers or identical but flawed interviewer training procedures across repeated administrations of the survey. The term ‘interviewer effects’ refers to measurement error attributable to a specific interviewer characteristic such as race or gender [5]. As with interviewer error, interviewer effects incorporate both bias and variable errors but typically refer to variable errors.

Why should public health professionals care?

Public health professionals have an interest in evaluating interviewer effects because interviewer characteristics may interact with those characteristics most often used to target populations for public health research and practice. Public health programs frequently define their target populations by race, ethnicity, gender, age and other sociodemographic characteristics. As a consequence, public health surveys frequently include topics related to these same sociodemographic qualities. Examples include the measurement of racial identity, culturally related health attitudes, sexual behaviors and aging. If interviewer characteristics affect survey estimates, then interviewer effects may therefore have a particularly strong role in public health data collection.

The presence of an interviewer can affect how a respondent forms an answer to a survey question and whether and how a respondent edits his answer before communicating it [6–9]. Williams [10] proposed that survey respondents edit their answers to increase potential rewards and decrease negative

ramifications. Public health surveys may be particularly vulnerable to response editing, as a respondent's answers to a questionnaire may determine whether he will gain access to health-related medical and educational services, expose himself to social harm through the revealing of stigmatized behaviors or health conditions or invite an unwelcome medical diagnosis. In such situations, respondents may not only seek to provide a positive self-image but also to scrutinize their interviewer for clues as to what responses might result in the best outcomes from their participation in the survey. At least one study indicates that African Americans, women, the aged and persons with lower income and education believe that the projection of a positive self-image is an important factor in securing quality medical care [11]. Interviewers are generally discouraged from conveying personal information, but characteristics such as gender, age and race are hard to conceal. Even when visual cues are inaccessible, such cues as an interviewer's name, speech style and voice qualities may enable respondents to form opinions about an interviewer's characteristics and beliefs.

Interviewer effects can have significant consequences for health data. These consequences arise because a single interviewer usually administers multiple surveys, resulting in the clustering of respondent data by interviewers. For example, if a survey employed one brunette interviewer and four blonde interviewers to administer a face-to-face survey to 100 respondents about attitudes toward brunettes, the 20 respondents interviewed by the brunette interviewer may voice more positive attitudes toward brunettes. The differences between the answers that the 20 respondents reported and their true attitudes about brunettes is interviewer error. Because there was a group of respondents who adjusted their responses in the same direction, the errors associated with the 20 respondents are correlated. As a result, the mean for items querying attitudes toward brunettes from all 100 respondents will be artificially high, the errors associated with these means will be interdependent and the variance, standard deviations and standard errors of the means will be inflated by interviewer error

[3]. Interviewer error can weaken the stability of survey statistics, increase or decrease the magnitude of estimates and influence the relationships observed among variables [12].

The Kish intraclass correlation coefficient, ρ_{int}^* (rho), is often used to measure interviewer error associated with a survey statistic. Values for ρ_{int}^* range from 0 to 1. Calculations of ρ_{int}^* across studies suggest average ρ_{int}^* values of 0.031 [3] and 0.009 [13] for face-to-face and telephone surveys, respectively. However, much larger values have been reported [14]. Both interviewer effects and interviewer workloads influence ρ_{int}^* [13]. For example, a ρ_{int}^* of 0.009 and an average interviewer workload of 75 results in an estimated 167% increase in the variance of a sample mean. Holding interviewer workload to 25 and raising ρ_{int}^* to 0.03 results in a comparable increase of 172%. Thus, even small interviewer effects can have a substantial influence on survey statistics.

Interviewer effects can be controlled using multi-level statistical models capable of accounting for the clustering of respondents by interviewers. Such models define interviewers as a separate level of data apart from respondent data and account for the nesting of respondent data within interviewers. By structuring the data into levels, the amount of variance derived from differences across interviewers can be measured. If this between-interviewer variance is substantial, it can be further evaluated to determine the most significant sources of variance. However, if the between-interviewer variance is acceptably mild, it can be ignored, and simpler, single-level statistical models can be used. Several scholars of interviewer effects have documented how their study findings would have been different if they had not accounted for the clustering of respondents by interviewers [5, 15–19]. In general, accounting for clustering appears to dampen the significance of direct effects and interactions among variables because unadjusted models underestimate variance as a consequence of ignoring between-interviewer variance. Studies that do not account for interviewer clustering may, therefore, report stronger interviewer effects than those that would have been found using statistical techniques

capable of adjusting for clustering effects [5, 15, 16], potentially resulting in Type 1 errors.

Interpreting research on interviewer effects

Public health professionals should be mindful of several factors when interpreting studies of interviewer effects. For one, ideal studies utilize large numbers of interviewers and respondents [20]. Just as more respondents lower sampling variance, more interviewers dilute the influence of individual interviewers on study findings [21]. However, some studies involve too few interviewers to isolate a single sociodemographic characteristic as a source of error [e.g. 22–25].

Interpenetrated survey designs randomly assign interviewers to respondents and avoid confounding interviewer and respondent characteristics [20]. However, practical factors often prohibit the use of both interpenetrated designs and large numbers of interviewers [7]. Thus, much of the literature on interviewer effects consists of either telephone surveys with small numbers of interviewers or face-to-face surveys without interpenetrated designs [7].

Rigorous studies also control for the clustering of respondents by interviewers. The availability of statistical software capable of running multi-level regression models is relatively recent. Before the advent of such software in the 1980s, many researchers used analysis of variance to measure interviewer effects. Although there are some limitations to this approach [13, 14, 26–28], studies using analyses of variance can generally be interpreted with confidence if the interviewer was treated as the unit of analysis.

It is important to isolate the characteristic under study [20]. For example, older interviewers may be more experienced; thus, analyses of interviewer age effects that do not control for interviewing experience may be confounded by interviewer experience. Extraneous interviewer effects can be controlled through study design, such as hiring interviewers with similar qualities, or statistical techniques during data analysis.

Studies of interviewer effects may further be affected by survey mode. Face-to-face survey respondents have access to a large range of auditory and visual cues, while telephone survey respondents only have access to auditory cues. Thus, studies of face-to-face surveys may be more likely to yield significant effects.

Surveys reflect prevailing social relations from particular moments in time. As social relations evolve, interviewer effects are likely to covary.

Finally, many interviewer effects studies conduct item-by-item analyses but do not publish the number of tests conducted, making it impossible for the reader to evaluate the odds of finding the reported results by chance alone [5].

Interviewer race and ethnicity effects

Many health professionals believe that respondents will feel more comfortable and be more honest with interviewers of their same race and ethnicity, particularly for surveys of racial or ethnic minorities. However, these seemingly homophilous pairings may induce salient, within-group racial or ethnic attitude differences between interviewers and respondents. If within-group differences yield interviewer effects, the systematic matching of interviewers and respondents by race and ethnicity may result in bias [3]. The following sections summarize research to date on race and ethnicity effects.

Interviewer race and ethnicity—face-to-face surveys

Four studies in the late 1950s and early 1960s suggest that race effects may emerge with Black and White respondents and be invoked by the mere presence of a survey proctor (Table I) [29]. Race effects may be particularly strong when interviewers are low on objectivity [10], when interviewers and respondents are discordant on multiple sociodemographic characteristics [10, 30, 31] and for race-related questions [10, 29, 31]. These effects are in deference to interviewer race, meaning that respondents tend to respond in ways that would be expected to be perceived more positively by someone of the interviewer's race [10, 29,

Table I. *Interviewer race and ethnicity effects studies^a*

Authors	Mode	Number of interviewers	Racial/ethnic composition of interviewers	Number of respondents	Race or ethnicity of respondents	Random ^b assignment of interviewers	Effort ^b to account for interviewer clustering	Queried racial or ethnic attitudes
Lenski and Leggett [30]	FTF	NR	White	624	Black and White	NR	No	No
Summers and Hammonds [29]	FTF	NR	Black and White	NR	White	NR	No	Yes
Williams [31]	FTF	22	Black and White	840	Black	Yes	No	Yes
Williams [10]	FTF	21	Black and White	452	Black	Yes	Yes	Yes
Carr [38]	FTF	6	Black and White	151	Black	NR	No	No
Schuman and Converse [8]	FTF	42	Black and White	619	Black	Yes	No	Yes
Hatchett and Schuman [36]	FTF	16	Black and White	106	White	Yes	Yes	Yes
Welch <i>et al.</i> [37]	FTF	6	Mexican American and Anglo	178	Mexican American	NR	No	No
Freitag and Barry [35]	FTF	54	Black and White	724	White	NR	No	No
Campbell [33]	FTF	12	Black and White	944	Black and White	No	No	Yes
Schaeffer [34]	FTF	NR	Black and White	Varies	Black and White	No	Yes	Yes
Cotter <i>et al.</i> [42]	Phone	12	Black and White	542	Black and White	Yes	No	Yes
Singer <i>et al.</i> [46]	Phone	35	Black and White	1014	NR	No	Yes	No
Anderson <i>et al.</i> [32]	FTF	NR	Black and White	1389	Black	No	No	No
Reese <i>et al.</i> [44]	Phone	15	Hispanic and Anglo	1004	Hispanic and Anglo (White and Black)	Yes	No	Yes
Davis [12]	Phone	76	Black and White	1150	Black	Yes	No	Yes
Davis [45]	Phone	76	Black and White	1150	Black	Yes	No	Yes
Finkel <i>et al.</i> [43]	Phone	NR	Black and White	252	White	Yes	No	Yes
Johnson and Parsons [39]	FTF	14	Black and White	481	Black and White	No	Yes	No
Fendrich <i>et al.</i> [16]	FTF	127	Black, White, Hispanic and Other	3978	NR	No	Yes	No
Wolford <i>et al.</i> [47]	Phone	NR	NR	1206	Black	NR	No	Yes

Table 1. Continued

Authors	Mode	Number of interviewers	Racial/ethnic composition of interviewers	Number of respondents	Race or ethnicity of respondents	Random ^b assignment of interviewers	Effort ^b to account for interviewer clustering	Queried racial or ethnic attitudes
Zimmerman <i>et al.</i> [41]	FTF	15	Black and White	591	Black	NR	No	No
Dailey and Claus [15]	FTF	22	Black and White	8276	Black and White	NR	Yes	No
Webster [40]	FTF	79	40 Hispanic; 39 Anglo	NR	Hispanic and Anglo	No	Yes	Yes
Livert <i>et al.</i> [48]	Phone	343	Black and White	12 872	Black, White and Hispanic	Yes	Yes	No
Davis and Silver [6]	Phone	NR	Black, White, Black and White and other	855	Black and White	Yes	No	No

^aNR = not reported; FTF = face-to-face.

^bThe space limitations of publications often prohibit sufficient description of study methodologies. Any misrepresentation of these studies is regretted.

31]. However, the lack of adjustment for clustering limits the conclusions that can be drawn from these studies.

Studies from the late 1960s to early 1980s indicate that race and ethnicity effects occur in data from African American [8, 32–34], White [33–36] and Mexican American [37] respondents. Many survey items from this era appear not to be affected by interviewer race or ethnicity [32–34, 37, 38], including demographic and health status questions [8, 34]. When effects were present, they primarily occurred for race-related items, and respondents deferred to their interviewer's race [8, 32–34, 36]. Race effects occurred in interviewer- and self-administered surveys when live interviewers were present [33]. However, only two studies [8, 36] appear to have randomly assigned interviewers, and only two [34, 36] studies adjusted for clustering. Thus, many of these studies may have overestimated interviewer effects.

Two 1990s studies suggest that Black and White respondents may be more likely to report substance use to White or Hispanic than Black interviewers [16, 39]. Black and White respondents may also be more likely to report physical abuse to White than to Black interviewers; however, reporting of sexual abuse may not be affected by interviewer race [15]. One study indicates that interviewer ethnicity influences racially and ethnically topical items, ethnic matching may improve response rates and interviewer ethnicity may interact with characteristics such as gender among Anglo and Hispanic respondents [40]. Most non-racial items in these surveys were not influenced by interviewer race or ethnicity [39–41]. Thus, it appears that interviewer race does not affect all survey topics equally. However, only one of these surveys [40] contained explicitly racial or ethnic items. Although none of these studies appears to have randomly assigned interviewers, the fact that most adjusted for clustering improves the validity of these findings.

Interviewer race and ethnicity—telephone surveys

The first set of telephone studies indicates that interviewer race and ethnicity can affect data from

White [42–44], Black [12, 42, 45] and Hispanic [44] respondents despite the absence of visual cues. Race and ethnicity effects were most likely to emerge when items overtly queried racial or ethnic attitudes [12, 42–45]. In these instances, respondents deferred to the interviewer's race. However, perceived race may influence responses more than actual interviewer race [12]. Interviewer race did not affect non-response in a survey without racial content [46]. All but one [46] of these 1980s studies used random interviewer assignment. However, only one study controlled for clustering; thus, these studies may overestimate statistically significant findings.

Findings from three studies conducted in the 1990s–2000s suggest that perceived race can be a stronger predictor of race effects than actual race [6], that racial items are most affected [47], that non-racial items tend not to be affected [6, 48] and that African American respondents will defer to the race of the interviewer [47]. One study suggests that African American respondents may fare poorer on knowledge tests when they believe that their interviewer is White [6]. Another study found that: White respondents were generally unaffected by interviewer race or ethnicity; Black respondents reported more alcohol and marijuana use, approval of alcohol and marijuana use and poorer overall health status to Black interviewers; and Hispanic respondents reported greater perceived harm from alcohol use, marijuana use and approval of marijuana use to Black interviewers [48]. Of these most recent telephone studies, only the latter study utilized random assignment and controlled for the clustering of respondents.

Summary of interviewer race and ethnicity studies

Research to date indicates that interviewer race and ethnicity can influence nonresponse as well as the validity and reliability of survey data. Significant race and ethnicity effects appear to be the exception rather than the rule. But, when effects occur, they tend to occur for overtly racial or ethnic attitude items and not to occur for items where a connection to race or ethnicity is missing or merely implied. Interviewer race and ethnicity appears not to affect

responses to demographic items. However, few studies report analyses of such items, and more research is warranted. Race and ethnicity effects occur in face-to-face, telephone and even self-administered surveys. When effects occur, respondents report in ways that might be perceived more positively by persons of the interviewer's race or ethnicity. Reporting of health status has had mixed findings, but surveys of physical abuse and substance use indicate that significant race and ethnicity effects may be operant. More race and ethnicity effects research is needed with health surveys.

The literature suggests that race and ethnicity effects occur in surveys of Black, White and Hispanic respondents; thus, it is likely that interviewer race and ethnicity influence survey data from other racial and ethnic groups. No studies have explored the effects of racial and ethnic attitude variability within racially and ethnically matched interviewer–respondent dyads. Research is needed on whether respondents' views about their own race and ethnicity affect their interactions with both interviewers of varying races and ethnicities and interviewers with strong versus weak ties to their own racial and ethnic groups.

A few studies suggest that perceived interviewer race and ethnicity may better predict interviewer effects for telephone-administered surveys. Respondents' abilities to judge the race of telephone interviewers vary widely, with estimates of correct race identification ranging from 14% to 82% [12, 47, D. C. Wilson, 2007, unpublished results]. Because of this variability, Wilson (2007, unpublished results) argues that differences between actual and perceived interviewer race may undermine the validity of findings based on actual interviewer race and that future research should measure perceived rather than actual race.

The extant literature on interviewer race and ethnicity effects fails to conclude whether survey respondents feel more comfortable with, trust, prefer or provide more accurate data to interviewers of their same race and ethnicity. It is not known whether survey data obtained by racially and ethnically matched interviewers are more accurate because most studies analyse attitude items for which there

are no accessible validity checks. Whether a respondent provides more accurate data to a matched interviewer is also likely determined by the respondent's own racial and ethnic identity orientation. Additional research on the effect of matching respondents and interviewers by race and ethnicity is needed.

Interviewer gender effects

Gender may be one of the most identifiable interviewer characteristics, and it is likely that respondents invoke gender-based stereotypes when editing their responses, particularly when surveys query issues related to gender norms.

One of the earliest gender effects studies concluded that female interviewers were more

likely than males to rate respondents as frank and honest (Table II) [49]. A slightly later study [50] found no evidence for interviewer gender effects. However, studies from the 1950s to 1960s are too sparse to draw conclusions with confidence.

The 1970s spawned several investigations of interviewer gender. Landis [25] reported that female respondents expressed more feminist responses to a male interviewer than to a female interviewer. However, since only two interviewers were utilized, these findings may be more attributable to individual characteristics other than gender. Other studies reported a lack of consistent, interpretable interviewer gender effects [13, 51, 52]. Freeman and Butler [14] associated higher ρ_{int}^* values with male interviewers, but the number of male interviewers in their study was too small to explore

Table II. Interviewer gender effects studies^a

Authors	Mode	Number of interviewers	Number of respondents	Gender of respondents	Random ^b assignment of interviewers	Effort ^b to account for interviewer clustering	Queried gender attitudes
Benney <i>et al.</i> [49]	FTF	NR	4708	Male and female	NR	No	No
Colombotos <i>et al.</i> [50]	FTF	31	1479	Male and female	No	No	No
Landis [25]	FTF	2	90	Female	Yes	No	Yes
Delamater [51]	FTF	NR	238	Male and female	No	No	No
Johnson and Delamater [52]	FTF	24	1361	Male and female	Yes	No	No
Freeman and Butler [14]	FTF	33	2600	Male and female	Yes	Yes	No
Groves and Magilavy [13]	Phone	30	954	Male and female	Yes	Yes	No
Grimes and Hansen [55]	Phone	NR	240	Male and female	NR	No	Yes
Groves and Fultz [20]	Phone	120	7300	Male and female	Yes	Yes	No
Hutchinson and Wegge [53]	Phone	26	795	Male and female	NR	No	No
Lueptow <i>et al.</i> [56]	Phone	NR	432	Male and female	NR	No	Yes
Johnson and Parsons [39]	FTF	14	481	Male and female	No	Yes	No
Kane and Macaulay [18]	Phone	29	1749	Male and female	No	Yes	Yes
Catania <i>et al.</i> [58]	Phone	40	2030	Male and female	Yes	Yes	No
Fendrich <i>et al.</i> [16]	FTF	127	3978	Male	No	Yes	No
Huddy <i>et al.</i> [59]	Phone	52	658	Male and female	Yes	Yes	Yes
Dailey and Claus [15]	FTF	22	8276	Male and female	NR	Yes	No
Livert <i>et al.</i> [48]	Phone	343	12 872	Male and female	Yes	Yes	No
Wilson <i>et al.</i> [19]	FTF	66	1146	Male and female	No	Yes	No
Pollner [57]	FTF	112	3131	Male and female	NR	No	No
Tourangeau <i>et al.</i> [60]	Web	2 (photos)	3047	Male and female	Yes	NA	Yes

^aNR = not reported; FTF = face-to-face; IVR = interactive voice recognition.

^bThe space limitations of publications often prohibit sufficient description of study methodologies. Any misrepresentation of these studies is regretted.

interviewer–respondent gender interactions. Overall, the 1970s studies provide little evidence for trends in gender effects.

Whereas prior gender studies examined face-to-face surveys, 1980s gender effects studies were all conducted by telephone. These 1980s studies provide little support for gender effects on survey findings [20, 53, 54] or nonresponse [20]. When they did occur, gender effects were most likely when respondents were answering gender attitude items [55, 56], questions about the economy [20] or political attitude items that may have invoked a gender-based stereotype threat [20, 53]. Effects were associated with both male [20, 53, 56] and female [20, 55, 56] respondents, both of whom provided more socially progressive responses or responses that deferred to the interviewer's gender. Only the Groves and Fultz [20] study accounted for the clustering of respondents.

All but 1990s study [57] controlled for respondent clustering. Although most studies found some evidence of gender effects, these were generally limited to a few items per survey. Where effects emerged, it is difficult to see consistent patterns across studies. Gender effects were associated with factual [15, 57], behavioral [16, 19, 39, 58] and attitudinal [18, 59] questions; direct effects [15, 18, 39, 57, 59] and interactions with respondent gender [39, 58], and face-to-face [15, 16, 19, 39, 57] and telephone [18, 58, 59] surveys. All the 1990s surveys queried gender-related or sensitive topics such as substance use or mental health. In the two explicitly gender-focused surveys, both male and female respondents deferred to their interviewer's gender [18, 59]. In contrast, interviewer gender sometimes interacted with respondent gender when sexual behaviors were surveyed [19, 58]. Although more research is needed, it appears that both male and female respondents may provide higher reports of some sexual behaviors to same-gender interviewers [19, 58]. Female interviewers tended to obtain higher reports of psychiatric symptoms [57] and sexual abuse [15]. For face-to-face surveys, it may be that respondents feel more comfortable disclosing information that might be viewed as victimizing to female interviewers

[15, 57], while disclosure of behaviors that might be viewed by some as behaviors of choice, such as illicit drug use, may be higher to male interviewers [16, 39].

Recent experiments suggest that the display of a male versus female photo may affect certain survey responses in Web-based surveys [60]. However, the context of when and why these effects emerge requires further investigation with studies using multiple gendered personae.

Given the wide range of study designs represented in the literature, it is difficult to draw conclusions about gender effects. It is unclear whether data collected by male or female interviewers are more accurate and if or when interviewers and respondents should be matched by gender. Research is needed to determine whether attitude or sensitive behavior questions are more affected by interviewer gender than other types of questions, whether different sensitive behaviors are differentially affected by interviewer gender and whether some topics invoke direct effects while others are conditional upon the gender make-up of an interviewer–respondent dyad. New investigation is required to explore whether and how the sexual orientations of interviewers and respondents contribute to interviewer effects. Public health professionals frequently engage in the measurement of gender-related and sensitive behaviors, as well as behaviors, attitudes and beliefs that may be affected by variability in gender-related self-identity. Further research on gender effects is warranted.

Other interviewer characteristics

Further research is warranted on additional interviewer characteristics. For example, interpretation of the current interviewer age effects literature is obscured by the use of differing age cut-offs to define interviewer age groups for analysis, making it difficult to compare results across studies. Many studies are also limited by the dichotomization of interviewer age into only two age groups [14, 16, 19, 39, 49], thereby aggregating interviewers of vastly different generations. Future age effects

studies should assess effects on age-related survey items and explore the use of perceived age versus actual interviewer age.

More experienced interviewers have been associated with fewer field coding errors [63], more valid data [64], more numerous open-ended item responses [64], higher consent rates for obtaining medical records [65], higher reporting of psychological symptoms [65, 66], lower survey response rates [46] and lower survey administration times [67]. However, findings from interviewer experience studies are difficult to summarize because different definitions of ‘high’ and ‘low’ experience categories are used across studies and the ranges of experience levels within these categories are often quite broad. Further, the very nature of what constitutes experience is in question: does market research or polling work count as interviewing experience or must prior experience be strictly classified as research survey work? Is face-to-face interviewing experience equivalent to telephone interviewing experience? Is experience confounded with other interviewer characteristics, such as enjoyment and interest in interviewing work [65, 66]?

Investigations of the effects of interviewers’ social status on survey responses are equally inconclusive [14, 31, 64, 68–70]. Social status is an elusive characteristic to define and an even more evasive characteristic to measure. However, constructed indices of social distance, or the degree of sociodemographic similarity between interviewers and respondents, can be useful, as they have the potential to advance theory by exploring why interviewer effects occur.

Future directions

Although interviewer effects appear not to affect most survey items, research to date indicates that interviewer effects such as race, ethnicity and gender can nonetheless occur in all interviewer-administered survey modes and can significantly alter survey findings. If these effects are replicated across surveys due to the repeated application of a flawed survey design, such as the consistent use of an all-Black or all-female interviewing staff

when interviewer race or gender effects are present, then bias will result. This bias may affect findings across studies and the general knowledge about a health issue.

More is unknown about interviewer effects than is known. Existing research suggests that interviewer race and ethnicity effects are most likely to occur when survey items query race- or ethnicity-related attitudes and that interviewer gender effects may be more likely to occur for questions about gender attitudes. In general, the direction of effects is in deference to the interviewer’s race, ethnicity or gender.

Future research should also consider additional, more unexplored interviewer characteristics. For instance, interviewer income level may be apparent face-to-face surveys, but this issue has not been explored in isolation of social class. Research is also needed to examine whether interviewers’ racial, ethnic or gender identities affect survey data. Telephone survey researchers should compare perceived with actual interviewer characteristics. Interactions among sociodemographic characteristics should be examined to explore the effects of multiple interviewer qualities, as characteristics such as race and gender are likely to interact in situations in which these characteristics are pertinent. At least one study has explored the influence of interviewers’ expectations about survey findings and found no evidence of strong effects [71]. However, given the popularity of trying to motivate interviewers by educating them about the purpose of a health survey, further investigation of interviewer-led expectations is warranted for additional modes, populations and survey topics [71]. It is also necessary to determine which interviewer characteristics are most influential in specific types of survey situations. Future interviewer effects research should also consider additional types of outcomes. Interviewer characteristics may not only affect non-response and the direction and intensity of survey responses but also whether particular interviewer–respondent combinations yield changes in survey response styles such as acquiescence, extreme responding or satisficing [72]. Additional research is needed to investigate all types of interviewer

effects in health topics, including the exploration of interviewer gender effects among non-racial sensitive topics such as sexual behavior, substance use, sexual abuse and mental health. Researchers should also be cognizant of geographical context. For example, race relations are likely to be different in different regions of the United States, and gender attitudes may be different in urban versus rural areas.

Given the current emphasis on increasing cultural sensitivity in public health research and practice, public health professionals need more empirical guidance on whether to match interviewers and respondents on varied sociodemographic characteristics. Research on data validity is also of the utmost importance, as evidence to date is lacking about whether respondents provide more valid data to sociodemographically similar or dissimilar interviewers. Research is likewise needed on which respondents feel more comfortable with, trust, and prefer interacting with homophilous or heterophilous interviewers, as these factors may influence the data that respondents provide. As per the recommendations of Groves [3], future research will ideally include large sample sizes, large corps of interviewer, interpenetrated survey designs and statistical techniques capable of accounting for the clustering of respondents by interviewers. These ideal study qualities may be difficult to achieve in many public health survey settings. However, even if lacking perfect design, the repeated investigation and reporting of interviewer effects, whether significant or null, will contribute to a significantly enhanced understanding of the magnitude and frequency of interviewer effects in public health research and practice over time.

Health practitioners should be cognizant that they may find themselves working in very specific cultural contexts, whether in the United States or abroad, that may bear upon the presence of interviewer effects. For example, Becker *et al.* [22] found that interviewer gender effects for a family planning survey in Nigeria were strongest in Islamic, culturally conservative areas and weakest in more socially liberal areas. In some cultures, certain types of interviewers or survey topics may

necessitate the presence of a male chaperone to gain access to interviewing women, and the presence of an observer may moderate the survey responses obtained. In addition, what is considered a sensitive topic in one culture may not be considered sensitive, or sensitive to the same degree, in another culture. Male interviewers administering a family planning survey in rural Nepal reported that they would be perceived as making sexual advances if they asked female respondents if they were pregnant and voiced fears of being attacked by respondents' husbands [73]. As a result, many male interviewers in the study skipped asking this question. Interviewer age effects may vary across cultures where elders are more or less esteemed, and social status effects may be more pronounced in cultures with more observable class hierarchies. Race and ethnicity are influential social characteristics in most cultures, but their exact effects are likely to vary by the race or ethnicity in question from culture to culture. Cultures also vary in their mistrust of and attitudes toward strangers [74]. Research on cross-cultural interviewer effects is still developing, but much progress is anticipated in the next decade.

Even with the advent of new technologies, interviewer-administered surveys are likely to persist. Further research on interviewer effects is therefore warranted to inform survey design and minimize undetected measurement error. If future research could identify what types of interviewer–respondent pairings result in the most accurate data for different survey designs, these dyads could be pursued through survey design by, for example, hiring only interviewers with a desired race, ethnicity or gender. However, the current state of knowledge is inadequate to guide the selection of interviewers based on sociodemographic characteristics. Public health professionals should instead aim to hire a large and diverse interviewing staff, to provide thorough and consistent interviewer training and monitoring and to measure and, if necessary, control for interviewer effects through multi-level statistical modeling. In so doing, they will rise to Hyman's challenge and contribute to the advancement of public health science.

Funding

National Cancer Institute (P50 CA101451).

Conflict of interest statement

None declared.

References

- Hyman HH, Cobb WJ, Feldman JJ *et al.* *Interviewing in Social Research*. Chicago, IL: University of Chicago Press, 1954.
- Brenner M. Response-effects of 'role restricted' characteristics of the interviewer. In: Dijkstra W, van der Zouwen J (eds). *Response Behaviour in the Survey-Interview*. New York: Academic Press, 1982, 91–130.
- Groves RM. *Survey Errors and Survey Costs*. Hoboken, NJ: John Wiley & Sons, Inc, 2004.
- Hagenaars JA, Heinen TG. Effects of role-independent interviewer characteristics on responses. In: Dijkstra W, van der Zouwen J (eds). *Response Behaviour in the Survey-Interview*. New York: Academic Press, 1982, 91–130.
- Dijkstra W. How interviewer variance can bias the results of research on interviewer effects. *Qual Quant* 1983; **17**: 179–87.
- Davis DW, Silver BD. Stereotype threat and race of interviewer effects in a survey on political knowledge. *Am J Pol Sci* 2003; **47**: 33–45.
- Krysan M, Couper MP. Race in the live and the virtual interview: racial deference, social desirability, and activation effects in attitude surveys. *Soc Psychol Q* 2003; **66**: 364–83.
- Schuman H, Converse JM. The effects of Black and White interviewers on Black responses in 1968. *Pub Opin Q* 1971; **35**: 44–68.
- Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. New York: Cambridge University Press, 2000.
- Williams JA, Jr. Interviewer role performance: a further note on bias in the information interview. *Pub Opin Q* 1968; **32**: 287–94.
- Malat JR, van Ryn M, Purcell D. Race, socioeconomic status, and the perceived importance of positive self-presentation in health care. *Soc Sci Med* 2006; **62**: 2479–88.
- Davis DW. Nonrandom measurement error and race of interviewer effects among African Americans. *Pub Opin Q* 1997; **61**: 183–207.
- Groves RM, Magilavy LJ. Measuring and explaining interviewer effects in centralized telephone surveys. *Pub Opin Q* 1986; **50**: 251–66.
- Freeman J, Butler EW. Some sources of interviewer variance in surveys. *Pub Opin Q* 1976; **40**: 79–91.
- Dailey RM, Claus RE. The relationship between interviewer characteristics and physical and sexual abuse disclosures among substance users: a multilevel analysis. *J Drug Issues* 2001; **31**: 867–88.
- Fendrich M, Johnson T, Shaligram C *et al.* The impact of interviewer characteristics on drug use reporting by male juvenile arrestees. *J Drug Issues* 1999; **29**: 37–58.
- Heeb JL, Gmel G. Interviewers' and respondents' effects on self-reported alcohol consumption in a Swiss health survey. *J Stud Alcohol* 2001; **62**: 434–42.
- Kane EW, Macaulay LJ. Interviewer gender and gender attitudes. *Pub Opin Q* 1993; **57**: 1–28.
- Wilson SR, Brown NL, Mejia C *et al.* Effects of interviewer characteristics on reported sexual behavior of California Latino couples. *Hisp J Behav Sci* 2002; **24**: 38–62.
- Groves RM, Fultz NH. Gender effects among telephone interviewers in a survey of economic attitudes. *Sociol Methods Res* 1985; **14**: 31–52.
- Stock JS, Hochstim JR. A method of measuring interviewer variability. *Pub Opin Q* 1951; **15**: 322–34.
- Becker S, Feyisetan K, Makinwa-Adebusoye P. The effect of the sex of interviewers on the quality of data in a Nigerian family planning questionnaire. *Stud Fam Plann* 1995; **26**: 233–40.
- Carpenter JS, Andrykowski MA, Cordova MJ *et al.* Do participants' reports of symptom prevalence or severity vary by interviewer gender? *Nurs Res* 1999; **48**: 276–9.
- Fry RPW, Rozewicz LM, Crisp AH. Interviewing for sexual abuse: reliability and effect of interviewer gender. *Child Abuse Negl* 1996; **20**: 725–9.
- Landis JR, Sullivan D, Sheley J. Feminist attitudes as related to sex of the interviewer. *Pac Sociol Rev* 1973; **16**: 305–14.
- Kish L. Studies of interviewer variance for attitudinal variables. *J Am Stat Assoc* 1962; **57**: 92–115.
- Hanson RH, Marks ES. Influence of the interviewer on the accuracy of survey results. *J Am Stat Assoc* 1958; **53**: 635–55.
- Hox JJ, de Leeuw ED, Kreft IGG. The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model. In: Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S (eds). *Measurement Errors in Surveys*. New York: John Wiley & Sons, Inc, 1991, 439–61.
- Summers GF, Hammonds AD. Effect of racial characteristics of investigator on self-enumerated responses to a Negro prejudice scale. *Soc Forces* 1966; **44**: 515–8.
- Lenski GE, Leggett JC. Caste, class, and deference in the research interview. *Am J Sociol* 1960; **65**: 463–7.
- Williams JA, Jr. Interviewer-respondent interaction: a study of bias in the information interview. *Sociometry* 1964; **27**: 338–52.
- Anderson BA, Silver BD, Abramson PR. The effects of the race of the interviewer on race-related attitudes of Black respondents in SRC/CPS National Election Studies. *Pub Opin Q* 1988; **52**: 289–324.
- Campbell BA. Race-of-interviewer effects among Southern adolescents. *Pub Opin Q* 1981; **45**: 231–44.
- Schaeffer NC. Evaluating race-of-interviewer effects in a national survey. *Sociol Methods Res* 1980; **8**: 400–19.
- Freitag CB, Barry JR. Interaction and interviewer bias in a survey of the aged. *Psychol Rep* 1974; **34**: 771–4.
- Hatchett S, Schuman H. White respondents and race-of-interviewer effects. *Pub Opin Q* 1975–1976; **39**: 523–8.

37. Welch S, Comer J, Steinman M. Interviewing in a Mexican-American community: an investigation of some potential sources of response bias. *Pub Opin Q* 1973; **37**: 115–26.
38. Carr LG. The Srole items and acquiescence. *Am J Sociol* 1971; **36**: 287–93.
39. Johnson TP, Parsons JA. Interviewer effects on self-reported substance use among homeless persons. *Addict Behav* 1994; **19**: 83–93.
40. Webster C. Hispanic and Anglo interviewer and respondent ethnicity and gender: the impact on survey response quality. *J Mark Res* 1996; **33**: 62–72.
41. Zimmerman MA, Caldwell CH, Bemat DH. Discrepancy between self-report and school-record grade point average: correlates with psychosocial outcomes among African American adolescents. *J Appl Soc Psychol* 2002; **32**: 86–109.
42. Cotter PR, Cohen J, Coulter PB. Race-of-interviewer effects in telephone interviews. *Pub Opin Q* 1982; **46**: 278–84.
43. Finkel SE, Guterbock TM, Borg MJ. Race-of-interviewer effects in a preelection poll: Virginia 1989. *Pub Opin Q* 1991; **55**: 313–30.
44. Reese SD, Danielson WA, Shoemaker PJ *et al.* Ethnicity-of-interviewer effects among Mexican-Americans and Anglos. *Pub Opin Q* 1986; **50**: 563–72.
45. Davis DW. The direction of race of interviewer effects among African-Americans: Donning the black mask. *Am J Pol Sci* 1997; **41**: 309–22.
46. Singer E, Frankel MR, Glassman MB. The effect of interviewer characteristics and expectations on response. *Pub Opin Q* 1983; **47**: 68–83.
47. Wolford ML, Brown RE, Marsden A *et al.* Bias in telephone surveys of African Americans: the impact of perceived race of interviewer on responses. 1995. Paper presented at the Annual Meeting of the American Statistical Association, Orlando, FL.
48. Livert D, Kadushin C, Schulman M *et al.* Do interviewer-responder race effects impact the measurement of illicit substance use and related attitudes? 1998. Paper presented at the 53rd Annual Conference of the American Association for Public Opinion Research, St Louis, MO.
49. Benney M, Riesman D, Star SA. Age and sex in the interview. *Am J Sociol* 1956; **62**: 143–52.
50. Colombotos J, Elinson J, Loewenstein R. Effect of interviewers' sex on interview responses. *Milbank Mem Fund Q* 1969; **47**: 227–31.
51. DeLamater JD. Methodological issues in the study of premarital sexuality. *Sociol Methods Res* 1974; **3**: 30–61.
52. Johnson WT, Delamater JD. Response effects in sex surveys. *Pub Opin Q* 1976; **40**: 165–81.
53. Hutchinson KL, Wegge DG. The effects of interviewer gender upon response in telephone survey-research. *J Soc Behav Pers* 1991; **6**: 573–84.
54. Johnson TP, Moore RW. Gender interactions between interviewer and survey respondents: issues of pornography and community standards. *Sex Roles* 1993; **28**: 243–61.
55. Grimes MD, Hansen GL. Response bias in sex-role attitude measurement. *Sex Roles* 1984; **10**: 67–72.
56. Lueptow LB, Moser SL, Pendleton BF. Gender and response effects in telephone interviews about gender characteristics. *Sex Roles* 1990; **22**: 29–42.
57. Pollner M. The effects of interviewer gender in mental health interviews. *J Nerv Ment Dis* 1998; **186**: 369–73.
58. Catania JA, Binson D, Canchola J *et al.* Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Pub Opin Q* 1996; **60**: 345–75.
59. Huddy L, Billig J, Bracciodieta J *et al.* The effect of interviewer gender on the survey response. *Polit Behav* 1997; **19**: 197–220.
60. Tourangeau R, Couper MP, Steiger DM. Humanizing self-administered surveys: experiments on social presence in Web and IVR surveys. *Comput Hum Behav* 2003; **19**: 1–24.
61. Ehrlich JS, Riesman D. Age and authority in the interview. *Pub Opin Q* 1961; **25**: 39–56.
62. Ford K, Norris AE. Effects of interviewer age on reporting of sexual and reproductive behavior of Hispanic and African American youth. *Hisp J Behav Sci* 1997; **19**: 369–76.
63. Stember H, Hyman H. Interviewer effects in the classification of responses. *Pub Opin Q* 1949–1950; **13**: 669–82.
64. Feldman JJ, Hyman H, Hart CW. A field study of interviewer effects on the quality of survey data. *Pub Opin Q* 1951–1952; **15**: 734–61.
65. Cleary PD, Mechanic D, Weiss N. The effect of interviewer characteristics on responses to a mental health interview. *J Health Soc Behav* 1981; **22**: 183–93.
66. McGlone MS, Aronson J, Kobrynowicz D. Stereotype threat and the gender gap in political knowledge. *Psychol Women Q* 2006; **30**: 392–8.
67. Olson K, Peytchev A. Effect of interviewer experience on interview pace and interviewer attitudes. *Pub Opin Q* 2007; **71**: 273–86.
68. Benney M, Geiss P. Social class and politics in Greenwich. *Br J Sociol* 1950; **1**: 310–27.
69. Dohrenwend BS, Colombotos J, Dohrenwend BP. Social distance and interviewer effects. *Pub Opin Q* 1968; **32**: 410–22.
70. Katz D. Do interviewers bias poll results? *Pub Opin Q* 1942; **6**: 248–68.
71. Clarke P, Sproston K, Thomas R. An investigation into expectation-led interviewer effects in health surveys. *Soc Sci Med* 2003; **56**: 2221–8.
72. Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. *Pub Opin Q* 1987; **51**: 201–19.
73. Axinn WG. The influence of interviewer sex on responses to sensitive questions in Nepal. *Soc Sci Res* 1991; **20**: 303–18.
74. Weinreb AA. The limitations of stranger-interviewers in rural Kenya. *Am Soc Rev* 2006; **71**: 1014–39.

Received on February 18, 2009; accepted on July 27, 2009