Interviewer effects in public health surveys


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

© The Author 2009. Published by Oxford University Press. All rights reserved.
For permissions, please email: journals.permissions@oxfordjournals.org
doi:10.1093/her/cyp046
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, \( \rho_{int}^{*} \) (rho), is often used to measure interviewer error associated with a survey statistic. Values for \( \rho_{int}^{*} \) range from 0 to 1. Calculations of \( \rho_{int}^{*} \) across studies suggest average \( \rho_{int}^{*} \) 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 \( \rho_{int}^{*} \) [13]. For example, a \( \rho_{int}^{*} \) 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 \( \rho_{int}^{*} \) 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 multilevel 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>
<thead>
<tr>
<th>Authors</th>
<th>Mode</th>
<th>Number of interviewers</th>
<th>Racial/ethnic composition of interviewers</th>
<th>Number of respondents</th>
<th>Race or ethnicity of respondents</th>
<th>Random(^b) assignment of interviewers</th>
<th>Effort(^b) to account for interviewer clustering</th>
<th>Queried racial or ethnic attitudes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenski and Leggett [30]</td>
<td>FTF</td>
<td>NR</td>
<td>White</td>
<td>624</td>
<td>Black and White</td>
<td>NR</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Summers and Hammonds [29]</td>
<td>FTF</td>
<td>NR</td>
<td>Black and White</td>
<td>NR</td>
<td>White</td>
<td>NR</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Williams [31]</td>
<td>FTF</td>
<td>22</td>
<td>Black and White</td>
<td>840</td>
<td>Black</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Williams [10]</td>
<td>FTF</td>
<td>21</td>
<td>Black and White</td>
<td>452</td>
<td>Black</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Carr [38]</td>
<td>FTF</td>
<td>6</td>
<td>Black and White</td>
<td>151</td>
<td>Black</td>
<td>NR</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Schuman and Converse [8]</td>
<td>FTF</td>
<td>42</td>
<td>Black and White</td>
<td>619</td>
<td>Black</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Hatchett and Schuman [36]</td>
<td>FTF</td>
<td>16</td>
<td>Black and White</td>
<td>106</td>
<td>White</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Welch et al. [37]</td>
<td>FTF</td>
<td>6</td>
<td>Mexican American and Anglo</td>
<td>178</td>
<td>Mexican American</td>
<td>NR</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Freitag and Barry [35]</td>
<td>FTF</td>
<td>54</td>
<td>Black and White</td>
<td>724</td>
<td>White</td>
<td>NR</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Campbell [33]</td>
<td>FTF</td>
<td>12</td>
<td>Black and White</td>
<td>944</td>
<td>Black and White</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Schaeffer [34]</td>
<td>FTF</td>
<td>NR</td>
<td>Black and White</td>
<td>Varies</td>
<td>Black and White</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cotter et al. [42]</td>
<td>Phone</td>
<td>12</td>
<td>Black and White</td>
<td>542</td>
<td>Black and White</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Singer et al. [46]</td>
<td>Phone</td>
<td>35</td>
<td>Black and White</td>
<td>1014</td>
<td>NR</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Anderson et al. [32]</td>
<td>FTF</td>
<td>NR</td>
<td>Black and White</td>
<td>1389</td>
<td>Black</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Reese et al. [44]</td>
<td>Phone</td>
<td>15</td>
<td>Hispanic and Anglo</td>
<td>1004</td>
<td>Hispanic and Anglo (White and Black)</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Davis [12]</td>
<td>Phone</td>
<td>76</td>
<td>Black and White</td>
<td>1150</td>
<td>Black</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Davis [45]</td>
<td>Phone</td>
<td>76</td>
<td>Black and White</td>
<td>1150</td>
<td>Black</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Finkel et al. [43]</td>
<td>Phone</td>
<td>NR</td>
<td>Black and White</td>
<td>252</td>
<td>White</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Johnson and Parsons [39]</td>
<td>FTF</td>
<td>14</td>
<td>Black and White</td>
<td>481</td>
<td>Black and White</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Fendrich et al. [16]</td>
<td>FTF</td>
<td>127</td>
<td>Black, White, Hispanic and Other</td>
<td>3978</td>
<td>NR</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Wolford et al. [47]</td>
<td>Phone</td>
<td>NR</td>
<td>NR</td>
<td>1206</td>
<td>Black</td>
<td>NR</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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, 30]. Black and White respondents may also be more likely to report physical abuse to White interviewers than Black interviewers [15]. One study notes that interviewers race and ethnicity interact with characteristics such as gender among Anglo and Hispanic respondents [40]. Most non-racial items in these studies were not influenced by interviewer race or ethnicity [39–41]. However, only one of these surveys [40] contained explicitly reported interviewers race and ethnicity—telephone surveys.

The first set of telephone studies indicates that interviewer race and ethnicity can affect data from surveys. Table I. Continued.

Table I. Continued

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

NR = not reported; FTF = face-to-face.

The space limitations of publications often prohibit sufficient description of study methodologies. Any misrepresentation of these studies is regretted.
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 \( \rho_{int}^* \) values with male interviewers, but the number of male interviewers in their study was too small to explore

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

\(^a\)NR = not reported; FTF = face-to-face; IVR = interactive voice recognition.

\(^b\)The 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

5. Dijkstra W. How interviewer variance can bias the results of research on interviewer effects. Qual Quant 1983; 17: 179–87.

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