# Chapter 12
# Correlates of Differences in Interactional Patterns among Black and White Respondents


**Jennifer Dykema, Dana Garbarski, Nora Cate Schaeffer, Isabel Anadon, and Dorothy Farrar Edwards**

## Introduction

While survey research remains one of the most important methodologies through which researchers collect data about key characteristics of populations, fundamental features of the survey interview may increase variable error, decreasing the precision of estimates. These features can also lead to systematic differences in respondents' answers across the spectrum of racial, ethnic, or other socially defined cultural groups, compromising researchers' ability to make group comparisons. In this chapter, we describe patterns of how answers to standardized survey questions about participating in medical research, some of which focus on race-related topics, occur during interviews with members from different racial groups, whose distinctive experiences with the topics of the questions may differentially affect how they cognitively process the items. We use interviewer-respondent interaction as a vehicle to examine and understand this processing.

J. Dykema (✉) · N. C. Schaeffer · I. Anadon · D. F. Edwards
University of Wisconsin-Madison, Madison, WI, USA
e-mail: dykema@ssc.wisc.edu

D. Garbarski
Loyola University Chicago, Chicago, IL, USA



P. S. Brenner (ed.), *Understanding Survey Methodology*, Frontiers in Sociology and Social Research 4, https://doi.org/10.1007/978-3-030-47256-6_12

## Participation in Medical Research: A Legacy of Mistrust among African Americans

Despite a mandate from the National Institutes of Health (NIH) to improve the inclusion of women and racial/ethnic minorities in research (NIH 2008), African Americans and other underrepresented groups continue to have very low rates of participation in medical research studies (Brown and Topcu 2003; George et al. 2014; Luebbert and Perez 2016). A growing body of literature specifically addresses the problems of recruitment and retention of minority participants in health-related research (Branson et al. 2007). The results of both qualitative and quantitative studies have identified many common barriers to participation, as well as issues unique to specific communities. The majority of these studies have focused on Black and African Americans, with fewer publications describing the attitudes and beliefs of Latinos and American Indians; however, the same themes consistently emerge across the groups. One of the most commonly cited factors includes a fear and mistrust of medical researchers based on episodes of unethical treatment by such investigators or discrimination associated with government sponsored programs.

The impact of unethical research practices has left a lingering sense of mistrust of biomedical research in the African American community in particular (Corbie-Smith 1999, Corbie-Smith et al. 1999, Corbie-Smith 2004). Mistrust of academic and research institutions are the most significant attitudinal barriers to research participation reported by African Americans and other groups underrepresented in biomedical research (George et al. 2014; Hoyo et al. 2003; Luebbert and Perez 2016). The etiology of mistrust is complex and multifaceted. One of the most frequently cited reasons for negative attitudes towards research are the historic violations of research ethics best exemplified by the Tuskegee Syphilis Trials (Bates et al. 2005; Gamble 1993; Shavers-Hornaday et al. 1997; Thomas and Quinn 1991). The negative consequences and resulting perceptions that followed Tuskegee and other well-known studies continues to influence research participation today; however, some researchers argue that awareness of Tuskegee alone does not predict mistrust of the medical care system (Branson et al. 2007; Scharff et al. 2010; Shavers et al. 2001).

One common consequence of the Tuskegee study is concern by subjects that they will be denied treatment for the health conditions under investigation. The belief expressed by some African Americans that HIV/AIDS was created in a laboratory and deliberately released in the Black community is plausibly a long-term consequence of that community's knowledge about the Tuskegee study (Washington 2006). Attitudinal studies suggest that mistrust of clinical investigators is highly influenced by perceived racial disparities in health, limited access to health care, and negative encounters with health care providers (Boulware et al. 2003; Halbert et al. 2006). Several investigators have found that Blacks are more likely than age-, education-, and gender-matched Whites to believe that research findings will be used to reinforce negative stereotypes about their racial/ethnic group (Goldman et al. 2008; Schulz et al. 2003), or will expose them to unnecessary risks (Branson et al.

2007; Corbie-Smith et al. 1999). For example, Corbie-Smith and colleagues worked with the Roper polling organization to collect survey data from a nationally representative sample of African Americans and White Americans. They reported that 79% of the African American respondents believed that they (or people like them) might be used as guinea pigs without consent and 63% of African Americans believe that they actually have been used in medical studies without consent.

It is within this social and historical context that we designed and administered the Voices Heard Survey, a standardized survey interview developed to identify barriers and facilitators to participating in medical studies designed to identify genetic, physiological, behavioral, environmental, or lifestyle markers of disease or disease risk.

## Standardized Interviewing and Interviewer-Respondent Interaction

Survey data are overwhelmingly gathered using standardized interviewing, which aims to control interviewer variability (Hyman 1975 [1954]; O'Muircheartaigh and Campanelli 1998; Schaeffer 1991; Schaeffer et al. 2010; West and Blom 2017). The rules of standardization most commonly referred to are those offered by Fowler and Mangione (1990): read questions as written; probe inadequate answers non-directively; record answers without discretion; and be interpersonally nonjudgmental regarding the substance of answers. If survey questions are clearly written and fit the target population, standardized interviews should consist of a series of "paradigmatic" question-answer sequences (Schaeffer and Maynard 1996, 2008), in which the interviewer reads the question as scripted and the respondent provides an answer to the question that is codable (e.g., "yes" or "no" for a yes/no question); optionally, the interviewer may acknowledge the respondent's answer (e.g., "thank you") before moving on to the next question. However, answers to survey questions are interactional accomplishments, and nonparadigmatic question-answer sequences arise for many reasons. These include respondents' displays of problems comprehending the meanings of questions and the terms they contain, difficulties respondents encounter mapping responses that summarize their attitudes and experiences onto the response categories provided, and a poor fit between the content of questions and respondents' knowledge or past experiences (Dykema et al. 1997; Holbrook et al. 2006).

Motivation for examining interviewer-respondent interaction is provided by the interactional model of the question-answer sequence that we developed in prior work (see Fig. 12.1; Dykema et al. 2020). The model summarizes paths that link the practices of standardization and conversation, characteristics of questions, respondents, and interviewers, cognitive processing of the survey participants, and the production of survey answers. The model is informed by a variety of sources: evidence of interviewer variance, which motivates the practices of standardization;

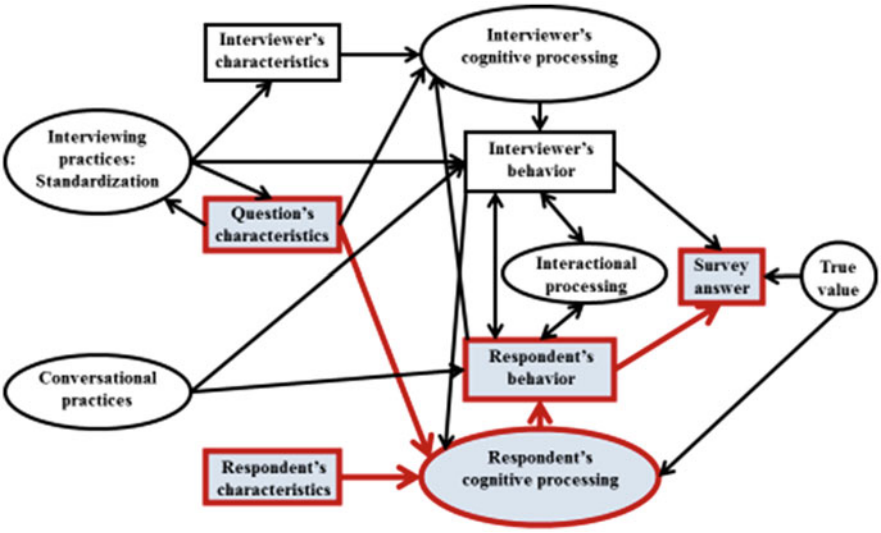


**Fig. 12.1** International model of the question-answer sequence adpted from Dykema et al. 2020

evidence that features of interaction such as response latency are associated with cognitive processing (Schaeffer and Dykema 2011a); and ways in which conversational practices affect standardized measurement (Garbarski et al. 2011, 2016; Schaeffer et al. 2010; Schaeffer and Maynard 2002; Schwarz 1996).

In Fig. 12.1, we highlight features of the model that are particularly relevant for this chapter. The model posits that for a given survey question, a respondent's cognitive processing is directly influenced by the question's characteristics, such as its topic, sensitivity, format, and complexity, and the respondent's characteristics, such as their past experiences and socio-demographic attributes including race and ethnicity. A respondent's cognitive processing, in turn, affects their behavior including how they respond, what they say, and what they offer as an answer. Data obtained in the survey interview are thus accomplished through the interplay of the instrument, respondent, and interviewer (Krosnick 2011; Schaeffer and Dykema 2011b; van der Zouwen and Smit 2004). Interviewer-respondent interaction is the vehicle through which the various characteristics of questions, respondents, and interviewers affect cognitive processing and data quality, with the verbal and nonverbal behavioral displays produced during the interaction providing inferences about the quality of the data generated (Dykema et al. 1997; Fowler 2011; Fowler and Cannell 1996; Ongena and Dijkstra 2007; Schaeffer and Dykema 2011a, b).

During the course of answering survey questions, respondents produce several distinct categories of talk including codable answers, uncodable answers, requests for repetition or clarification, and conversational elements (see Table 12.1). As noted previously, the survey interview is designed to obtain a codable answer. To be codable, an answer must occur after the respondent has heard the entire question,

**Table 12.1** Examples of units-of-talk provided by respondents during the course of answering the selection questions included in the current study

| Unit-of-talk | Definition | Examples |
|---|---|---|
| Codable answers | | |
| Exact repetition | Answer that is an exact repetition of one of the response categories | "extremely hard" for question 34 |
| Kernel | Answer that is a "kernel" of a category (i.e., a word or phrase that uniquely and unambiguously identifies a single category) | "extremely" for "extremely hard" for question 34 |
| Category reference | Answer that references a specific category based on the location of the category relative to the other categories | "the last one" for "extremely" for question 34 |
| Uncodable answers | | |
| Hypothetical response category | Answer is a response option that could hypothetically be included along the response dimension but is not included as one of the response categories offered | "little likely" for question 1, "pretty often" for question 40 |
| Repeat or paraphrase part of the question | Answer repeats part of the question verbatim or paraphrases part of question in a way that does not provide new information | "I would answer the questions" for question 1 |
| Repeat response dimension | Answer repeats some part or all of the response dimension but without the intensifier and so does not uniquely identify a single response category | "likely" for question 1, "often" for question 39 |
| Report | Answer does not only repeat or paraphrase part of the question but provides relevant information that is stated as an answer but is not codable | "they seem like they care when I've done them" for question 35, "everybody's just a number when they are doing it they aren't thinking about them as people" for question 41 |
| Requests | | |
| Request repetition or clarification | Comment or question requesting repetition or clarification of a term, phrase, or some part of the question or response categories | "can you read/repeat that," "what are the five options," "are we talking about researchers in the United States" for question 32 |
| Conversational elements | | |
| Apology | Word or phrase that conveys the act of being sorry | "I apologize," "excuse me," "I'm sorry" |
| Comment | Unscripted talk, often evaluative in nature, about the question, respondent, interviewer, or interviewing situation | "it's hard to answer that," "that's a good question," "I'm losing my focus" |
| Elaboration | Additional information offered along with a codable answer to explain the answer provided and | "because I can't stand blood" for question 3, "depends on the |

(continued)

**Table 12.1** (continued)

| Unit-of-talk | Definition | Examples |
|---|---|---|
| | sometimes preceded by "because," "depends" or "if" | medication they're testing for" for question 6 |
| Exclamation | Word or phrase that expresses sudden surprise, anger, excitement, happiness, or other emotion | "boy," "dear," "geez," "gosh," "shoot," "wow" |
| Laughter | Freestanding laughter (laughter that occurs between words) or laugh tokens (particles of laughter that occur within words or phrases) | |
| Mitigator | Word or phrase that reduces the exactness, precision, or certainty of another utterance or that itself expresses uncertainty | "about," "just," "kind of," "I would say," "I don't know," "maybe" |
| Token | Particles of speech that indicate a delay or disruption in the actor's cognitive processing | "ah," "aw," "eh," "er," "hm," "huh," "mm," "uh," "um" |

adequately answer the question, and match the response format of the question (e.g., one of the response categories or the format on the screen or paper) (Schaeffer et al. 2020).

All of the questions in the current study have a similar response format consisting of a set of ordered categories, commonly referred to as a rating scale (see Appendix A). For this type of response format, an answer is codable if it is an exact repetition of one of the response categories (e.g., "extremely hard" when the response categories are "not at all hard, a little hard, somewhat hard, very hard, extremely hard"), a "kernel" of a response category, a word that uniquely and unambiguously identifies a single category (e.g., "extremely" uniquely and unambiguously identifies the category "extremely hard"), or a category reference, a reference to a specific category based on the positioning of the category relative to the other categories (e.g., "the last one" as a reference to "extremely hard").

In lieu of providing a codable answer, respondents may offer an uncodable answer, an answer that is provided in an attempt to respond to the survey question, but that cannot be coded using the response format or response categories offered. As described in more depth in Table 12.1, uncodable answers in this study take many forms including hypothetical response categories, repetitions or paraphrases of the question, repeating the response dimension without indicating a unique response category, and reports. Alternatively, because they did not hear or did not understand some part or all of a question or the response categories, respondents may refrain from answering a question and instead request repetition or clarification. Finally, respondents may provide different kinds of conversational elements along with codable answers, uncodable answers, and requests (Garbarski et al. 2016). These conversational elements include such varied behaviors as apologies, comments, elaborations, exclamations, laughter, mitigators, and tokens.

The presence (or absence) of any of these categories of talk likely varies based primarily on characteristics of questions but also on characteristics of respondents and to a lesser extent, because they are trained to be standardized, characteristics of interviewers (e.g., Olson and Smyth 2015). While uncodable answers, requests, and conversational elements arise for many reasons, including everyday conversational practices, they frequently occur when respondents encounter difficulty answering questions and may signal a problem with cognitive processing. For example, interactional behaviors that evince uncertainty (e.g., mitigators) or problems with response processing (e.g., reports) appear to increase when respondents must integrate conflicting information about their health, such as the presence of disease but high physical functioning, when answering a question on self-rated health (Garbarski et al. 2011). Because many of these behaviors are often associated with survey data that are of lower quality as indicated by being less reliable or valid (Dykema et al. 1997; Schaeffer and Dykema 2011b) or with response patterns that are undesirable, studying these behaviors may tell us something about the quality of the data we are collecting.

## Racial/Ethnic Variation in Survey Response Processing

Coding interaction between survey participants to study cultural variation in how respondents behave is a "relatively new innovation" (Johnson et al. 2019, p. 272). Although research is limited, some evidence suggests intensive study of interviewer-respondent interaction may index differences across racial/ethnic groups in how respondents process survey concepts or how they exhibit comprehension or mapping difficulties. For example, Holbrook et al. (2006) evaluated questions about health events and behaviors used in federal population surveys and found certain racial and ethnic minority groups showed more interactional behaviors associated with comprehension problems—such as requests for clarification—than did non-Hispanic Whites. These differences suggest the meaning of concepts may vary across groups in such a way that respondents from minority groups have difficulty comprehending questions when the language or concepts are fitted to the dominant group; however, this study found no differences across groups in behaviors that indicate mapping difficulties such as providing inadequate or imprecise answers.

Johnson et al. (2015) examined levels of interactional indicators of possible measurement problems, such as interruptions, requests for clarification, and problems answering, displayed by respondents during the administration of questions about self-reported racial and ethnic discrimination. The questions studied used two different approaches to measuring discrimination: a one-stage approach in which questions directly focused on discrimination based on race/ethnicity (e.g., "... how often have you been treated with less respect than other people because you are RACE/ETHNICITY") versus a two-stage approach in which questions asked about non-race-related treatment first (e.g., "how often have you been treated with less respect than other people") followed by a list of reasons including race/ethnicity

(e.g., "because of your race or skin color"). Overall, results indicated that while the two-stage approach was associated with lower odds of respondent problems than the one-stage form, the interactional indicators did not vary by race/ethnicity with the exception that the odds of exhibiting problems answering were lower among Latino than White respondents.

More recently, Johnson et al. (2019) sought to determine whether respondents from diverse racial and ethnic backgrounds and interviewed in multiple languages would display similar levels of comprehension and mapping difficulties when responding to questions deliberately designed to evoke such difficulties. For example, questions posed comprehension problems by asking about nonexistent objects (e.g., "how frequently have you visited a serrerium") and presented mapping challenges by mismatching the response format projected by the question (e.g., "Does it ever snow at the equator?" which projects a yes/no response) and response options (e.g., "never, occasionally, sometimes, or frequently?"). Overall, findings indicated that the levels of behaviors indicating difficulties demonstrated by the groups were remarkably consistent. An exception was that in contrast to non-Hispanic Whites, Korean-Americans interviewed in English produced lower levels of mapping problems for questions written to elicit such difficulties.

## Current Study

We use interaction coding to examine differences between Black and White respondents answering sets of questions on varied topics including the likelihood of participating in medical research studies that collect different kinds of measures (e.g., blood, saliva) and questions about trust in medical researchers that are or are not focused on race. Trust is a central concept in many disciplines including sociology, survey methodology, and medicine (e.g., Dillman et al. 2014). Past research demonstrates that trust in medical researchers varies across racial and ethnic groups. For African Americans this distrust is rooted in the legacy of historical atrocities that have been perpetrated against them and that make up the collective memory of many African Americans. Distrust also stems from knowledge of and experience with a system of health research and health care that produces and reproduces unequal access, experiences, and treatment of individuals in that group (Corbie-Smith et al. 2002; Feagin and Bennefield 2014; Scharff et al. 2010). Consequently, we posit the questions on trust in medical research that focus on race are a better cultural fit for the Black respondents, in that they will be more likely to have had experiences and knowledge that align with what the questions are asking. We predict White respondents will be more likely to display behavioral indicators of problems for the race-focused questions about trust because they ask about concepts and use language that is less familiar for this group. We predict the other question sets—used here as controls for comparison to some extent—will be associated with similar levels of indicators of problems with cognitive processing for both racial groups.

## Methods

### *Sample*

The Voices Heard computer-assisted telephone survey sought to interview a total of 400 individuals from Wisconsin, equally distributed among the following racial and ethnic groups: White, Black, Latino, and American Indian. We employed a quota sampling strategy for the study because the costs of screening to identify members in the non-White groups would have been prohibitively expensive. The quota sample consisted primarily of volunteers; however, to supplement the volunteer sample, a targeted list containing names was provided by a vendor of consumer data (see Appendix B for more detail on the sample). Interviewers conducted 410 usable interviews (in English only) between October 2013 and March 2014 with respondents in the four subgroups defined by their race and ethnicity.

Respondents were categorized into the four racial/ethnic groups based on self-reports to a series of questions about their perceived racial and ethnic identities. The series began with a yes/no question asking respondents if they are "Hispanic or Latino." Respondents answering "yes" to this question were classified as "Latino" regardless of how they answered a follow-up question about their race. To assess race, respondents were asked, "Which one or more of the following would you say is your race: White, Black or African American, American Indian, Alaska Native, Asian, or Native Hawaiian or Other Pacific Islander?" Interviewers were instructed to record all of the categories offered by the respondent; respondents provided up to three categories. Respondents were classified as: "White," if they answered "no" to the question on Hispanic origin and reported no other racial categories; "Black," if they answered "no" to the question on Hispanic origin and reported "Black" only or "Black" and "White" as their race; or "American Indian" if they answered "no" to the question on Hispanic origin and reported "American Indian" alone or in combination with one or more other racial categories. In addition, one respondent who failed to answer the question on Hispanic origin, but reported "American Indian" as their race, was classified as "American Indian."

The average time to complete the interview was 25.21 minutes. We produced digital recordings for 371 interviews; 24 interviews were not recorded because the respondent refused and 15 were lost due to poor quality or recording errors. We limit our analysis to a comparison between the Black ($n = 90$) and White ($n = 94$) respondents because of the well-documented differences between these two groups in their experiences with, attitudes toward, and knowledge about the health care system.

## Questionnaire and Items

The primary objective of the survey was to measure respondents' perceptions of the barriers and facilitators to participating in medical research studies that collect biomarkers, such as saliva, blood and tissue, and to document whether there were important differences among groups defined by their race and ethnicity. The telephone interview was part of a larger research effort that involved key informant and cognitive interviews with members of populations underrepresented in biomedical research. Questions covered the following topics: ratings of the likelihood to participate in specific types of medical research studies such as those that collect biomarkers; ratings of the likelihood to participate in medical research studies depending on the characteristics of the person making the request (e.g., "a member of your community"); evaluation of things medical researchers do to encourage participation (e.g., provide results or incentives); evaluation of things that sometimes concern people about participating in medical research; views toward medical researchers (e.g. how much trust or mistrust respondents have); measures of general health status, health-related quality of life, health behaviors, chronic conditions, and health care utilization; general knowledge of research procedures; and sociodemographic characteristics.

The current analysis focuses on three sets of questions (see Appendix A). The first set of questions is from a battery of items that uses the same response categories for each question and asks respondents to rate their self-assessed likelihood of participating in medical research studies that involve answering questions, giving samples of saliva, blood, tissue, or cerebrospinal fluid, or participating in a clinical trial. The second and third set of questions are from a twelve-item scale about trust in medical researchers adapted from previously administered instruments (Dykema et al. 2019). The response categories for these questions vary depending on the underlying dimension in the question (e.g., "never" to "extremely often" for frequency-based questions versus "not at all" to "extremely" for intensity-based questions). Within this twelve-item scale about trust in medical researchers, we make a further distinction about whether the questions focus on race or not. Thus, the second set of questions are race-focused trust in medical research questions, and the third set of questions are non-race-focused trust in medical research questions.

## Systematic Transcriptions and Interaction Coding

Three transcribers listened to the audio recordings and created systematic transcriptions based on procedures we developed in previous work and which we describe in some detail here. Transcribers recorded all of the interaction that occurred between the interviewer and respondent for a given question. Within a question, interaction was segmented into turns, a unit-of-talk from one actor—the interviewer or respondent—that was not broken up by talk from the other actor. A turn-of-talk reached

**Table 12.2** Example of systematic transcription for Questions 36 and 37, Case 10032, White male

| ID | Question | Actor | Turn | Interruption | Overlap | Laugh Token |
|---|---|---|---|---|---|---|
| 10032 | 36 | i | when selecting participants for their most risky studies how likely are medical researchers to select minorities not at all likely a little likely somewhat likely very likely or extremely likely? | | | |
| | | r | I I I I think that's a really weird question I'm just going to say not likely | | 1 | |
| | | i | {L} ok um not at all likely or a little likely I guess I have to ask you | | 1 | |
| | | r | not at all likely | | 1 | |
| | | i | ok | | | |
| | 37 | i | how often do medical researchers hide information about the possible risks of participating in medical research studies never rarely sometimes very often or extremely often? | | | |
| | | r | well I mean a as as an express eh obviously I'd had no answer I mean I would hope it's never but I don't know | | | |
| | | i | ok don't know | | | |

Notes: "i" = interviewer, "r" = respondent, {L} = freestanding laughter

completion when the other actor began talking either because the original actor's talk concluded or the current actor interrupted the original actor.

For each interview, transcribers began with a template formatted in Excel containing a row that displayed the exact wording of each question (see Table 12.2). Transcribers listened to the audio and recorded any departures interviewers made in administering the question exactly as worded. In subsequent rows of the Excel sheet, transcribers recorded talk produced by the interviewer or respondent before the interviewer moved on to the next question. In addition to recording talk verbatim, transcribers also wrote out tokens (e.g., "ah"), coded whether the respondent interrupted the interviewer's initial reading of the question, and recorded whether the turn contained overlapping talk, freestanding laughter (laughter that occurs between words), or laugh tokens (particles of laughter that occur within words or phrases).

Coding the turns-of-talk was done in Stata using the electronic transcripts by a member of the project team in consultation with other members of the team. Using commands to read string variables, the primary coder identified strings of text capturing different units of talk, such as those described in Table 12.1. These string

functions allowed us to parse talk into discrete coding units. As an example, in response to the race-focused Question 36, "When selecting participants for their most risky studies, how likely are medical researchers to select minorities: not at all likely, a little likely, somewhat likely, very likely, or extremely likely?", a respondent answered "that's a tricky one cause it depends what they're studying ah I would say a little likely." We coded this respondent's turn into categories representing a comment on the question ("that's a tricky one"), an elaboration ("cause it depends what they're studying"), a token ("ah"), a mitigator ("I would say"), and a codable answer ("a little likely").

## Measures

We examine five outcomes previous research has found to be associated with lower data quality and that are hypothesized to be indicators of potential cognitive problems respondents have when processing a survey question. Note that these are not mutually exclusive indicators of processing difficulties but different ways to conceptualize separate but related features of the response process and potential breakdowns in cognitive processing (Garbarski et al. 2011). Our first outcome indicated whether the question-answer sequence contained more than three turns, a sign the respondent may have had difficulty answering the question and the interviewer intervened by following up in order to obtain a codable answer. Second, we examine whether the respondent failed to provide a codable answer during their first turn of talk. As noted, respondents routinely include other non-standardized talk with a codable answer, particularly in the course of thinking out loud and formulating a response. As long as this talk did not contradict the respondent's final answer, it was included as part of a codable answer. Third, we code whether the respondent requested to have all or part of the question or response categories repeated or requested clarification of a term or phrase in their first turn. These requests happened in sequences with a codable answer, but were more likely to occur in lieu of providing a codable answer. Our fourth outcome marked whether the respondent's talk included an affective element such as laughter, a laugh token, or an exclamation (e.g., "gosh," "oh boy," or "wow") during their first turn of talk. Finally, we examine whether the respondent's initial turn of talk included a token, such as "ah," "er," "uh," or "um." These particles likely indicate a delay or disruption in the actor's cognitive processing (e.g., Bortfeld et al. 2016).

Although the primary respondent characteristic of interest is the respondent's race, we also include gender, age, and education (high school education or less, some college, college or more) as control variables in the multivariate models.

## Analytic Strategy

Our unit of analysis is the question-answer sequence. A question-answer sequence began with the interviewer's administration of the question and ended with the last utterance spoken before the next question was read, typically the respondent's final answer or a statement by the interviewer acknowledging the respondent's answer (e.g., "ok"). The analysis examines 3301 question-answer sequences produced by the 184 respondents answering the 18 questions; ten question-answer sequences are omitted because the recording for the question was not audible.

To account for the complicated crossed and nested structure of the data, we implement a mixed-effects model with a variance structure that uses crossed random effects. Initial models included random effects for interviewers, questions, and respondents (nested within interviewers and crossed with questions). However, results indicated that including all three random effects resulted in the models being overfitted and the estimate of the interviewer effect being close to zero, and so we removed the random intercept for the interviewer. Respondent characteristics and question set (i.e., trust in medical research questions that are race-focused, those that are not race-focused, and likelihood to participate in medical research questions) are modeled as fixed effects which are nested within and crossed with the random effects. Each of the dependent variables are binary; logit models were computed in Stata using the meqrlogit function. Because of our relatively small sample of respondents, we describe results with a p-value of less than .10 as marginally significant and .05 or less as statistically significant.

## Results

Descriptive statistics are shown in Table 12.3. Approximately 23% of the question-answer sequences contained more than three turns. Respondents did not provide a codable answer in their first turn in 19% of the sequences. They requested a repetition or clarification in nearly 8% of all of their first turns, and their response included an affective element in the first turn in 5% of the question administrations. Tokens were fairly common, occurring in the first turn in 25% of the sequences. The quota sample yielded approximately equal numbers of Black and White respondents, slightly more women than men, and a roughly equal distribution of respondents in the three educational categories.

In a series of bivariate analyses, we examine whether there was a difference between Black and White respondents in the likelihood of producing the outcomes of interest for each of the question sets. Table 12.4 presents results from separate multilevel logistic regression models in which the interactional outcome is regressed on the respondents' race, separately for each question set.

For the models predicting question-answer sequences with more than three turns, no codable answer, and requests for repetition or clarification, results indicate that

**Table 12.3** Descriptive statistics for interactional outcomes, respondent characteristics, and question sets

| | Mean or Percent | Standard Deviation | Minimum | Maximum | n |
|---|---|---|---|---|---|
| Interactional outcomes | | | | | |
| More than 3 turns (vs. less) | 23.21 | | 0 | 1 | 3301 |
| No codable answer (vs. codable answer) | 19.21 | | 0 | 1 | 3301 |
| Any request for repetition or clarification (vs. none) | 7.66 | | 0 | 1 | 3301 |
| Affective response (vs. not) | 4.85 | | 0 | 1 | 3301 |
| Token (vs. none) | 25.05 | | 0 | 1 | 3301 |
| Respondent characteristics | | | | | |
| Race | | | | | |
| Black | 48.91 | | | | 90 |
| White | 51.09 | | | | 94 |
| Gender | | | | | |
| Male | 44.02 | | | | 81 |
| Female | 55.98 | | | | 103 |
| Education | | | | | |
| High school or less | 35.87 | | | | 66 |
| Some college | 28.26 | | | | 52 |
| College or more | 35.87 | | | | 66 |
| Age (in years) | 44.70 | 16.74 | 18.00 | 90.00 | 184 |
| Question sets | | | | | |
| Trust questions: Race-focused | 27.78 | | | | 5 |
| Trust questions: Non-race-focused | 38.89 | | | | 7 |
| Likelihood to participate | 33.33 | | | | 6 |

when answering the race-focused trust questions, White respondents were significantly (or marginally so for requests) more likely to produce longer sequences, uncodable answers, and marginally more likely to produce requests than Black respondents. In contrast, the levels of these outcomes did not differ between White and Black respondents for the non-race-focused trust questions or the likelihood-to-participate questions.

The pattern of results for affective elements and tokens was slightly different (Table 12.4). White respondents were (marginally) more likely to display an affective element than Black respondents for both the race- and non-race-focused trust questions, while levels were the same for the likelihood-to-participate questions. The only question set to show a difference between White and Black respondents for tokens was for the non-race-focused trust questions, for which White respondents were more likely to produce one or more tokens while answering.

Next, we examine whether the levels of differences for the Black and White respondents in the bivariate models are statistically significant across the question

**Table 12.4** Bivariate, multilevel logistic regression analyses of interactional outcomes on respondents' race within question set

| | More than 3 turns | | | | | | No codable answer | | | | | | Request repetition or clarification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trust race-focused | | Trust non-race-focused | | Likelihood to participate | | Trust race-focused | | Trust non-race-focused | | Likelihood to participate | | Trust race-focused | | Trust non-race-focused | | Likelihood to participate | |
| Variables | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE |
| Respondent characteristics | | | | | | | | | | | | | | | | | | |
| White (vs. Black) | 0.462* | 0.228 | −0.073 | 0.218 | 0.071 | 0.144 | 0.904*** | 0.282 | 0.090 | 0.243 | −0.025 | 0.874 | 0.665+ | 0.345 | 0.203 | 0.243 | −0.062 | 0.234 |
| Intercept | −1.956*** | 0.258 | −2.015*** | 0.230 | −0.631*** | 0.107 | −2.415*** | 0.294 | −2.341*** | 0.232 | −1.067*** | 0.149 | −3.430*** | 0.388 | −3.045*** | 0.232 | −2.571*** | 0.268 |
| Random-effects parameters | | | | | | | | | | | | | | | | | | |
| Question-level variance | 0.157 | 0.124 | 0.166 | 0.110 | 0.003 | 0.016 | 0.141 | 0.118 | 0.112 | 0.081 | 0.053 | 0.048 | 0.133 | 0.130 | 0.263 | 0.180 | 0.162 | 0.133 |
| Respondent-level variance | 0.768 | 0.280 | 0.814 | 0.258 | 0.186 | 0.106 | 1.540 | 0.435 | 1.089 | 0.321 | 0.205 | 0.123 | 1.644 | 0.664 | 0.601 | 0.333 | 0.045 | 0.297 |
| Model fit statistics | | | | | | | | | | | | | | | | | | |
| N | 918 | | 1288 | | 1095 | | 918 | | 1288 | | 1095 | | 918 | | 1288 | | 1095 | |
| Wald chi-square | 4.11* | | 0.11 | | 0.24 | | 10.25** | | 0.14 | | 0.03 | | 3.72+ | | 0.62 | | 0.07 | |
| Log likelihood | −432.18 | | −525.81 | | −714.17 | | −412.52 | | −481.10 | | −628.82 | | −254.65 | | −320.94 | | −287.86 | |

(continued)

**Table 12.4** (continued)

| | Affective element | | | | | | Token | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trust race-focused | | Trust non-race-focused | | Likelihood to participate | | Trust race-focused | | Trust non-race-focused | | Likelihood to participate | |
| Variables | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE |
| Respondent characteristics | | | | | | | | | | | | |
| White (vs. Black) | 1.039+ | 0.563 | 0.830+ | 0.428 | 0.416 | 0.329 | 0.394 | 0.242 | 0.533* | 0.214 | 0.347 | 0.221 |
| Intercept | −5.342*** | 0.838 | −4.680*** | 0.519 | −3.426*** | 0.357 | −1.738*** | 0.191 | −1.778*** | 0.196 | −1.244*** | 0.179 |
| Random-effects parameters | | | | | | | | | | | | |
| Question-level variance | 0.020 | 0.080 | 0.023 | 0.082 | 0.125 | 0.113 | 0.000 | 0.000 | 0.075 | 0.059 | 0.029 | 0.035 |
| Respondent-level variance | 4.437 | 2.693 | 1.908 | 1.068 | 1.526 | 0.593 | 1.150 | 0.324 | 1.020 | 0.251 | 1.176 | 0.284 |
| Model fit statistics | | | | | | | | | | | | |
| N | 918 | | 1288 | | 1095 | | 918 | | 1288 | | 1095 | |
| Wald chi-square | 3.40+ | | 3.75+ | | 3.29 | | 2.65 | | 6.19* | | 2.46 | |
| Log likelihood | −141.58 | | −178.40 | | −307.69 | | −468.85 | | −655.76 | | −635.64 | |

sets. Table 12.5 presents results from multivariate models that include interaction terms for race by question set; the models also control for respondents' sociodemographic characteristics. To facilitate interpretation of the results, Fig. 12.2 provides the estimated marginal predicted probability of each outcome by race and question set. For the model predicting question-answer sequences with more than three turns, we find that White respondents answering the race-focused trust questions are more likely to require more than three turns to answer the question compared to Black respondents ($b = 0.355$, $p < .10$; coefficient for "White" because the reference group is Black respondents answering the race-focused trust questions), although the effect is attenuated compared to the results in Table 12.4 by controlling for respondents' sociodemographic characteristics. Further, the interaction between race and question set is significant for non-race-focused trust questions ($b = -0.492$, $p < .05$), indicating that the effect of race in the race-focused trust questions is significantly different from the effect of race in the non-race-focused trust questions. The interaction of race with the likelihood-to-participate questions is not significant, indicating that the effect of race in the race-focused trust questions is not different from the effect of race in the likelihood-to-participate questions.

A similar but stronger pattern of results is shown for question-answer sequences that fail to result in a codable answer in the first turn. White respondents are significantly more likely not to provide a codable answer than Black respondents to race-focused trust questions ($b = 0.643$, $p < .01$). Both of the interaction terms are significant, indicating that the effect of race in the race-focused trust questions is significantly different from the effect of race in the non-race-focused trust questions and the likelihood-to-participate questions.

Requests for repetition or clarification mirror the previous results but are not as strong, likely due to the fact that such requests rarely occur. The effect of race for the race-focused trust questions is marginally significant ($b = 0.537$, $p < .10$) as is the interaction term for race by the non-race-focused trust questions ($b = -0.651$, $p < .10$).

Turning to the results predicting the presence of an affective element in the first term, while the effect of race for the race-focused trust questions is marginally significant ($b = 0.537$, $p < .10$), the effect of race in the race-focused trust questions is not different from the effect of race in the non-race-focused trust questions or the likelihood-to-participate questions. Finally, net of other respondent characteristics, race is no longer a significant predictor of tokens.

## Discussion

Overall, we find that non-Hispanic Whites display more interactional behaviors and patterns that may indicate comprehension and mapping difficulties than do Black respondents for questions about trust in medical researchers that invoke race. In bivariate analyses, this pattern holds for the indicators of multiple turns of talk, uncodable answers, requests for repetition or clarification, and the presence of

**Table 12.5** Multivariate, multilevel logistic regression analyses of interactional outcomes on respondent characteristics, question sets, and interaction terms

| | More than 3 turns | | No codable answer | | Request repetition or clarification | | Affective element | | Token | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE |
| Respondent characteristics | | | | | | | | | | |
| White (vs. Black) | 0.355+ | 0.214 | 0.643** | 0.229 | 0.537+ | 0.306 | 0.785+ | 0.430 | 0.329 | 0.248 |
| Female (vs. male) | 0.093 | 0.137 | 0.067 | 0.154 | 0.260 | 0.208 | 0.402 | 0.251 | −0.139 | 0.183 |
| Education | | | | | | | | | | |
| High school or less | – | – | – | – | – | – | – | – | – | – |
| Some college | −0.105 | 0.167 | 0.234 | 0.186 | 0.476+ | 0.254 | 0.182 | 0.317 | 0.463* | 0.226 |
| College or more | −0.119 | 0.160 | −0.111 | 0.181 | 0.344 | 0.242 | 0.320 | 0.290 | 0.413 | 0.217 |
| Age | 0.009* | 0.004 | 0.011* | 0.005 | 0.004 | 0.006 | 0.015+ | 0.008 | −0.007 | 0.006 |
| Question sets | | | | | | | | | | |
| Trust questions: race-focused | – | – | – | – | – | – | – | – | – | – |
| Trust questions: non-race-focused | −0.054 | 0.275 | −0.099 | 0.268 | −0.017 | 0.383 | −0.180 | 0.472 | −0.072 | 0.204 |
| Likelihood to participate | 1.155*** | 0.271 | 0.921*** | 0.263 | 0.203 | 0.390 | 1.011* | 0.424 | 0.487* | 0.205 |
| Interaction terms | | | | | | | | | | |
| White × non-race-focused trust questions | −0.492* | 0.240 | −0.666** | 0.252 | −0.388 | 0.339 | −0.077 | 0.510 | 0.157 | 0.227 |
| White × likelihood-to-participate questions | −0.336 | 0.222 | −0.778*** | 0.235 | −0.651+ | 0.348 | −0.520 | 0.449 | −0.027 | 0.226 |
| Intercept | −2.242*** | 0.322 | −2.656*** | 0.353 | −3.705*** | 0.497 | −5.347*** | 0.612 | −1.526*** | 0.376 |
| Random-effects parameters | | | | | | | | | | |
| Question-level variance | 0.102 | 0.047 | 0.103 | 0.408 | 0.223 | 0.099 | 0.125 | 0.082 | 0.038 | 0.025 |
| Respondent-level variance | 0.415 | 0.088 | 0.544 | 0.108 | 0.757 | 0.203 | 0.791 | 0.275 | 1.010 | 0.167 |
| Model fit statistics | | | | | | | | | | |
| N | 3301 | | 3301 | | 3301 | | 3301 | | 3301 | |
| Wald chi-square | 51.63*** | | 43.24*** | | 10.72 | | 25.77** | | 26.11** | |
| Log likelihood | −1657.98 | | −1506.86 | | −851.90 | | −590.69 | | −1705.65 | |

+p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001

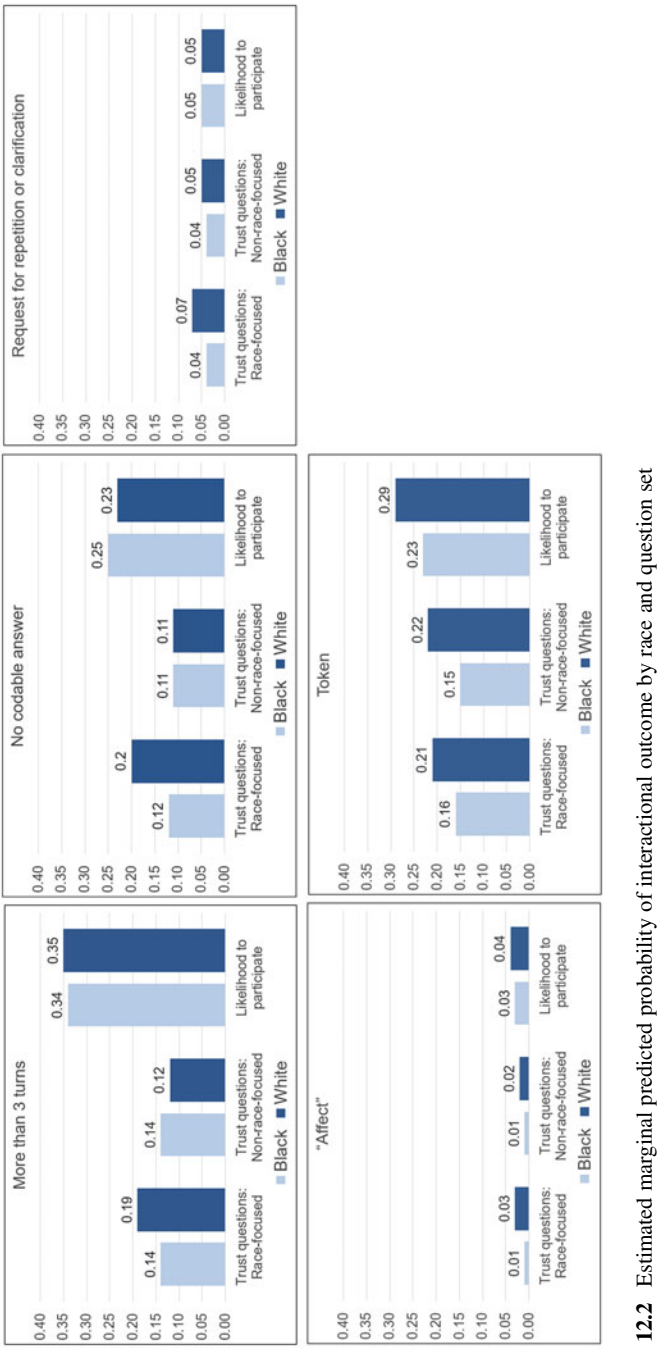


**Fig. 12.2** Estimated marginal predicted probability of interactional outcome by race and question set

affective elements. In multivariate analyses that include interaction terms for race by question set and control for other socio-demographic characteristics, we find that the difference between Blacks and Whites answering the race-focused trust questions is significantly different from the race effect (1) for the non-race-focused questions for the model predicting more than three turns; (2) for both the non-race-focused questions and likelihood-to-participate questions for the model predicting no codable answer in the first turn; and (3) for the likelihood-to-participate questions for the model predicting requests in the first turn. For affective elements, neither of the interaction terms are significant, possibly due to the fact that these are very infrequent behaviors.

Tokens occur relatively frequently—respondents utter them in their first turn in a quarter of the question-answer sequences examined—and they are associated with a different interactional trajectory. In bivariate analyses, tokens are more commonly produced by White respondents for the non-race-focused trust questions (Table 12.5), but this effect is no longer significant when other characteristics of respondents are controlled and neither of the interaction terms is significant in the multivariate models. Interestingly, there is a significant association between education and tokens. It may be that the delay and disfluency tokens analyzed here are not associated with cognitive processing difficulties; rather, they reflect ways in which speakers "hold the floor" while they are thinking in ways that vary across sociodemographic groups. The implications of this finding require future theorizing and research.

We speculate that the differences between Black and White respondents in interactional patterns in this study arise because awareness of the concept of trust in medical researchers varies across these groups. Non-Hispanic Whites are less familiar with the concept of (dis)trust in medical researchers compared to racial/ethnic minority groups who are more likely to have been exposed to such considerations both personally and during interactions with members of their communities (Corbie-Smith et al. 2002; Feagin and Bennefield 2014; Scharff et al. 2010). Lack of general familiarity and personal experiences may have exacerbated cognitive processing difficulties when the questions focus on race for the White respondents. Although the evidence is limited and the findings somewhat mixed, these results and some previous research suggest potential differences across groups in the response processing of survey questions (Schoua-Glusberg 2011; Warnecke et al. 1997). These findings add to a small body of work that explores whether the interaction that unfolds between interviewers and respondents in standardized interviews varies across respondents from different racial or ethnic groups (Holbrook et al. 2006; Johnson et al. 2015; Johnson et al. 2019).

Our study was limited in several regards. First, the sample sizes for each of the groups under study was relatively small and may have decreased our ability to detect statistically significant differences between the groups, especially for the rarer behaviors of requesting repetition or clarification and providing affective elements as part of one's response. Second, due to cost constraints, respondents were not recruited randomly, which limits the generalizability of our sample to a larger population. Third, the questions we examine were not randomly sampled from a

population of questions that focus on race versus not, and so the conclusions we draw may be limited to the comparisons tested in this study. Fourth, the five outcomes we examine were selected because they have been shown to be associated with lower data quality in previous research and they feature characteristic types of talk that respondents display when answering survey questions. However, we may have overlooked behaviors characteristics of Black respondents that have not been described in the literature. Furthermore, our indicators of problems in cognitive processing are not independent of each other. For example, when respondents fail to produce a codable answer or request clarification, interviewers are trained to follow up, with the result that the question-answer sequence will by necessity contain more than three turns.

Fifth, we use interaction coding as a vehicle to study differences in how respondents from different racial groups process surveys items. As a criterion that indexes data quality, behavior from interviewer-respondent interaction has the advantages that it can be observed and coded for all questions from any interviewer-administered instrument and provides information about the performance of individual items in actual operational setting. Further, the rich quantitative and qualitative data produced during the question-answer sequences can be coded reliably from transcriptions, particularly using the methodology advanced in the current study of systematically coding strings of text. In addition, although recording, transcribing, and coding interactions is not inexpensive, it may be less expensive than other designs for assessing data quality, although we are not aware of studies that compare question-testing methodologies in terms of costs.

There are, of course, disadvantages to the methodology. Some response processing problems may be internal to the participants and leave no trace in the interaction. As Johnson et al. (2019, p. 274) document "respondents may in some instances elect to answer difficult or unclear questions without revealing any misunderstandings or other confusion about them." While research is limited, it does appear, however, that respondents from different racial and ethnic backgrounds demonstrate similar levels of problem indicators when responding to questions intentionally written to evoke comprehension and mapping difficulties (Johnson et al. 2019). Although past research demonstrates they are often associated with measures of validity and reliability (Schaeffer and Dykema 2011b), the behavioral outcomes we examine are only proxy measures of response error, and we lack external criteria to determine whether the behaviors we examine predict measurement error. It is possible that the behaviors we examine may be influenced by factors that are not direct influences on data quality (see Palmieri 2016).

The requirements of standardized measurement may pose challenges for respondents from different social and cultural backgrounds—and these challenges may affect measurement differently for these groups, compromising the use of surveys as a tool for comparative research. In so far as the differential rate of problematic behaviors is associated with measurement error, this error complicates estimates of differences across groups and may lead to incorrect conclusions about the overall levels of and differences in an outcome of interest among groups, such as apparent differences in health across groups even though true differences do not exist or

apparent similarities when true differences exist, using health as an example. Previous research on interviewer-respondent interaction during the standardized survey interview has focused on documenting what features of this interaction reveal about response processing and data quality. A new generation of research has begun exploring what a detailed analysis of interaction may tell us about how respondents from various cultural groups, including those that differ based on their race and ethnicity, respond to questions that vary in their topics and characteristics. Products of this research have important implications for the practice of survey research. A detailed analysis of interviewer-respondent interaction can: (1) inform best practices for writing survey questions and designing survey forms for respondents with varying backgrounds and characteristics, (2) identify features of interaction, such as those that indicate uncertainty, to use as control variables to augment analysis, and (3) highlight techniques to improve interviewer training, for example, by using evidence of what participants actually do to inform decisions about when and how interviewers should intervene in the question-answer process to obtain codable answers (Garbarski et al. 2016; Schaeffer et al. 2016).

**Acknowledgements** The collection of survey data for the Voices Heard Survey was funded by NIMHD grant P60MD003428 (PD: A. Adams). Project: Increasing Participation of Underrepresented Minorities in Biomarker Research (PI: D. Farrar Edwards). This study is based upon work supported by the National Science Foundation (grant number SES-1853094 to J. Dykema and D. Garbarski]. Project: Effects of Interviewers, Respondents, and Questions on Survey Measurement. Additional support was provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, the Charles Cannell Fund in Survey Methodology at the University of Michigan, the University of Wisconsin Survey Center (UWSC), which receives support from the College of Letters and Science at the University of Wisconsin-Madison, and the facilities of the Social Science Computing Cooperative and the Center for Demography and Ecology (NICHD core grant P2C HD047873). The authors thank Steven Blixt for comments on earlier drafts and Russell Diamond for advice on the analysis. Opinions expressed here are those of the authors and do not necessarily reflect those of the sponsors or related organizations.

## Appendix A: Exact Question Wordings by Topic

| Question Number | Question Stem | Response Categories |
|---|---|---|
| **Likelihood to participate in medical research** | | |
| 1 | If a medical researcher asked you to participate in a medical research study by **answering questions about yourself**, how likely would you be to participate | very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely |
| 2 | If a medical researcher asked you to participate in a medical research study by **giving a sample of your saliva**, how likely would you be to participate | very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely |

(continued)

| Question Number | Question Stem | Response Categories |
|---|---|---|
| 3 | If a medical researcher asked you to participate in a medical research study by **giving a sample of your blood**, how likely would you be to participate | (very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely) |
| 4 | Tissue is located in the human body and is made up of cells. Small pieces of tissue can be taken from the body by a health care professional. If a medical researcher asked you to participate in a medical research study by **giving a sample of your tissue**, how likely would you be to participate | very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely |
| 5 | Cerebrospinal fluid is a fluid that surrounds your brain. It can be collected by inserting a small needle into your lower back, a procedure called a lumbar puncture or spinal tap. If a medical researcher asked you to participate in a medical research study by **giving a sample of your cerebrospinal fluid**, how likely would you be to participate | (very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely) |
| 6 | A clinical trial is a study that tests new drugs or treatments. If a medical researcher asked you to participate in a **clinical trial**, how likely would you be to participate | (very likely, somewhat likely, neither likely nor unlikely, somewhat unlikely, or very unlikely) |
| **Trust in medical research: Non-race-focused** | | |
| 32 | All things considered, how much do you trust medical researchers | none, a little, some, quite a bit, or a great deal |
| 34 | How hard do medical researchers work to make sure that the participants in their studies are safe | not at all hard, a little hard, somewhat hard, very hard, or extremely hard |
| 35 | To what extent do medical researchers care more about the findings of their research than they do about their participants | not at all, a little, somewhat, quite a bit, or a great deal |
| 37 | How often do medical researchers hide information about the possible risks of participating in medical research studies | never, rarely, sometimes, very often, or extremely often |
| 39 | How often do medical researchers tell participants everything they need to know about the risks of participating in their studies | never, rarely, sometimes, very often, or extremely often |
| 41 | How hard do medical researchers work to make sure they keep information from participants private and secure | not at all hard, a little hard, somewhat hard, very hard, or extremely hard |
| 43 | How often do medical researchers want to know more than they need to know | never, rarely, sometimes, very often, or extremely often |

(continued)

| Question Number | Question Stem | Response Categories |
|---|---|---|
| **Trust in medical research: Race-focused** | | |
| 33 | When they are conducting research, how often do medical researchers have the best interests of participants from your racial or ethnic group in mind | never, rarely, sometimes, very often, or extremely often |
| 36 | When selecting participants for their most risky studies, how likely are medical researchers to select minorities | not at all likely, a little likely, somewhat likely, very likely, or extremely likely |
| 38 | How often do medical researchers treat participants from your racial or ethnic group like guinea pigs in their studies | never, rarely, sometimes, very often, or extremely often |
| 40 | How often do medical researchers treat participants from your racial or ethnic group the same as participants from other racial or ethnic groups | never, rarely, sometimes, very often, or always |
| 42 | How concerned are you that the information collected in medical research studies could be used to confirm or promote stereotypes | not at all concerned, a little concerned, somewhat concerned, very concerned, or extremely concerned |

## Appendix B: Sample Description

The following table shows the distribution of completed interviews by race/ethnicity for the volunteer and vendor lists.

*Volunteer list*. For the volunteer sample, members of the project team recruited 471 (n = 46 White, n = 137 Black, n = 144 Latino, and n = 144 American Indian) individuals through connections they built with leaders in specific racial and ethnic communities, by visiting churches and community centers, by attending events sponsored by specific racial or ethnic groups (e.g., pow-wows), and by posting flyers at targeted locations in communities. Project staff collected names, demographic data (e.g., race and ethnicity), and contact information (e.g., phone numbers) for these potential respondents, and all individuals identified through these channels were contacted and asked to participate in the study.

*Vendor list*. A total of 8075 records were purchased from Infogroup, a business and consumer data provider. Infogroup filtered data from their databases based on a surname algorithm and geo-coding that would supposedly help target individuals living in diverse communities in Wisconsin. In addition, Infogroup filtered records to accrue only those with high-deliverability for direct mail and those with active telephone numbers. From the list of records, a total of 700 cases (7 replicates of 100 cases each, consisting overall of 100 White, 200 Black, 200 Latino, and 200 American Indian targeted individuals) were fielded for calling.

| Number of Completed Cases by Race/Ethnicity of List Source | | |
|---|---|---|
| **Race/Ethnicity** | **Volunteer List** | **Vendor List** |
| White | 29 | 73 |
| Black | 103 | 3 |
| Latino | 93 | 7 |
| American Indian | 101 | 1 |
| Total | 326 | 84 |

## References


Bates, B. R., Lynch, J. A., Bevan, J. L., & Condit, C. M. (2005). Warranted concerns, warranted outlooks: A focus group study of public understandings of genetic research. *Social Science & Medicine, 60*(2), 331–344. https://doi.org/10.1016/j.socscimed.2004.05.012

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2016). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123–147. https://doi.org/10.1177/00238309010440020101.

Boulware, L. E., Cooper, L. A., Ratner, L. E., LaVeist, T. A., & Powe, N. R. (2003). Race and trust in the health care system. *Public Health Reports, 118*(4), 358–365. https://doi.org/10.1093/phr/118.4.358.

Branson, R. D., Davis, K., & Butler, K. L. (2007). African Americans' participation in clinical research: Importance, barriers, and solutions. *The American Journal of Surgery, 193*(1), 32–39. https://doi.org/10.1016/j.amjsurg.2005.11.007.

Brown, D. R., & Topcu, M. (2003). Willingness to participate in clinical treatment research among older African Americans and Whites. *The Gerontologist, 43*(1), 62–72. https://doi.org/10.1093/geront/43.1.62.

Corbie-Smith, G. (1999). The continuing legacy of the Tuskegee Syphilis Study: Considerations for clinical investigation. *The American Journal of the Medical Sciences, 317*(1), 5–8. https://doi.org/10.1016/S0002-9629(15)40464-1.

Corbie-Smith, G. (2004). Minority recruitment and participation in health research. *North Carolina Medical Journal, 165*(6), 385–387.

Corbie-Smith, G., Thomas, S. B., & St. George, D. M. M. (2002). Distrust, race, and research. *Archives of Internal Medicine, 162*(21), 2458–2463. https://doi.org/10.1001/archinte.162.21.2458.

Corbie-Smith, G., Thomas, S. B., Williams, M. V., & Moody-Ayers, S. (1999). Attitudes and beliefs of African Americans toward participation in medical research. *Journal of General Internal Medicine, 14*(9), 537–546. https://doi.org/10.1046/j.1525-1497.1999.07048.x.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: Wiley.

Dykema, J., Garbarski, D., Wall, I. F., & Edwards, D. F. (2019). Measuring trust in medical researchers: Adding insights from cognitive interviews to examine agree-disagree and construct-specific survey questions. *Journal of Official Statistics, 35*(2), 353-386. https://doi.org/10.2478/jos-2019-0017.

Dykema, J., Lepkowski, J. M., & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 287–310). New York, NY: Wiley. https://doi.org/10.1002/9781118490013.ch12.

Dykema, J., Schaeffer, N. C., Garbarski, D., & Hout, M. (2020). The role of question characteristics in designing and evaluating survey questions. In P. Beatty, D. Collins, L. Kaye, J. Padilla, G. Willis, & A. Wilmot (Eds.), *Advances in questionnaire design, development, evaluation, and testing* (pp. 449–470). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781119263685.ch6.

Feagin, J., & Bennefield, Z. (2014). Systemic racism and U.S. health care. *Social Science & Medicine, 103*, 7–14. https://doi.org/10.1016/j.socscimed.2013.09.006.

Fowler, F. J., Jr. & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 15–36). San Francisco, CA: Jossey-Bass.

Fowler, F. J., Jr. & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park: Sage.

Fowler, F. J., Jr. (2011). Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Miller, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality* (pp. 5–21). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118037003.ch2.

Gamble, V. N. (1993). A legacy of distrust: African Americans and medical research. *American Journal of Preventive Medicine, 9*(6), 35–38.

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2011). Are interactional behaviors exhibited when the self-reported health question is asked associated with health status? *Social Science Research, 40*(4), 1025–1036. https://doi.org/10.1016/j.ssresearch.2011.04.002.

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2016). Interviewing practices, conversational practices, and rapport: Responsiveness and engagement in the standardized survey interview. *Sociological Methodology, 46*(1), 1–38. https://doi.org/10.1177/0081175016637890.

George, S., Duran, N., & Norris, K. (2014). A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *American Journal of Public Health, 104*(2), e16–e31. https://doi.org/10.2105/ajph.2013.301706.

Goldman, R. E., Kingdon, C., Wasser, J., Clark, M. A., Goldberg, R., Papandonatos, G. D., et al. (2008). Rhode Islanders' attitudes towards the development of a statewide genetic biobank. *Personalized Medicine, 5*(4), 339–359. https://doi.org/10.2217/17410541.5.4.339.

Halbert, C. H., Armstrong, K., Gandy, O. H., Jr., & Shaker, L. (2006). Racial differences in trust in health care providers. *Archives of Internal Medicine, 166*(8), 896–901. https://doi.org/10.1001/archinte.166.8.896.

Holbrook, A., Cho, Y. I., & Johnson, T. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly, 70*(4), 565–595. https://doi.org/10.1093/poq/nfl027.

Hoyo, C., Reid, M. L., Godley, P. A., Parrish, T., Smith, L., & Gammon, M. (2003). Barriers and strategies for sustained participation of African-American men in cohort studies. *Ethnicity & Disease, 13*(4), 470–476.

Hyman, H. H. (1975 [1954]). *Interviewing in social research*. Chicago, IL: The University of Chicago.

Johnson, T. P., Holbrook, A., Cho, Y. I., Shavitt, S., Chavez, N., & Weiner, S. (2019). Examining the comparability of behavior coding across cultures. In T. P. Johnson, B. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)* (pp. 271–291). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118884997.ch13.

Johnson, T. P., Shariff-Marco, S., Willis, G. B., Cho, Y. I., Breen, N., Gee, G. C., et al. (2015). Sources of interactional problems in a survey of racial/ethnic discrimination. *International Journal of Public Opinion Research, 27*(2), 244–263. https://doi.org/10.1093/ijpor/edu024.

Krosnick, J. A. (2011). Experiments for evaluating survey questions. In J. Madans, K. Miller, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of*

*data quality* (pp. 215–238). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118037003.ch14.

Luebbert, R., & Perez, A. (2016). Barriers to clinical research participation among African Americans. *Journal of Transcultural Nursing, 27*(5), 456–463. https://doi.org/10.1177/1043659615575578.

NIH. (2008). Guidelines on the inclusion of women and minorities as subjects in clinical research–Amended, October, 2001.

Olson, K., & Smyth, J. D. (2015). The effect of CATI questions, respondents, and interviewers on response time. *Journal of Survey Statistics and Methodology, 3*(3), 361–396. https://doi.org/10.1093/jssam/smv021.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society, Series A, 161*, 63–77. https://doi.org/10.1111/1467-985X.00090.

Ongena, Y. P., & Dijkstra, W. (2007). A model of cognitive processes and conversational principles in survey interview interaction. *Applied Cognitive Psychology, 21*(2), 145–163. https://doi.org/10.1002/acp.1334.

Palmieri, M. (2016). Is verbal interaction coding a reliable pretesting technique? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 129*(1), 64–77. https://doi.org/10.1177/0759106315615512.

Schaeffer, N. C. (1991). Conversation with a purpose–or conversation? Interaction in the standardized interview. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 367–392). New York: Wiley. https://doi.org/10.1002/9781118150382.ch19.

Schaeffer, N. C., & Dykema, J. (2011a). Questions for surveys: Current trends and future directions. *Public Opinion Quarterly, 75*(5), 909–961. https://doi.org/10.1093/poq/nfr048.

Schaeffer, N. C., & Dykema, J. (2011b). Response 1 to Fowler's chapter: Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Miller, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality* (1st ed., pp. 23–39). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118037003.ch3.

Schaeffer, N. C., Dykema, J., Coombs, S. M., Schultz, R. K., Holland, L., & Hudson, M. L. (2020). General interviewing techniques: Developing evidence-based practices for standardized interviewing. In K. Olson, J. D. Smyth, J. Dykema, A. Holbrook, F. Kreuter, & B. T. West (Eds.), *Interviewer effects from a total survey error perspective* (pp. 33-46). Boca Raton, FL: CRC Press Taylor & Francis Group.

Schaeffer, N. C., Dykema, J., & Garbarski, D. (2016). *Answers to standardized questions: Recognizing "codable" answers and maintaining standardization and rapport*. Paper presented at the International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC), July, Chicago, IL.

Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and interviewing. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 437–470). Bingley, UK: Emerald Group Publishing Limited.

Schaeffer, N. C., & Maynard, D. W. (1996). From paradigm to prototype and back again: Interactive aspects of cognitive processing in survey interviews. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 65–88). San Francisco, CA: Jossey-Bass.

Schaeffer, N. C., & Maynard, D. W. (2002). Standardization and interaction in the survey interview. In J. Holstein & J. Gubrium (Eds.), *Handbook of Interviewing* (pp. 577–601). Thousand Oaks, CA: Sage.

Schaeffer, N. C., & Maynard, D. W. (2008). The contemporary standardized survey interview for social research. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 31–57). Hoboken, NJ: Wiley. https://doi.org/10.1002/9780470183373.ch2.

Scharff, D. P., Mathews, K. J., Jackson, P., Hoffsuemmer, J., Martin, E., & Edwards, D. (2010). More than Tuskegee: Understanding mistrust about research participation. *Journal of Health Care for the Poor and Underserved, 21*(3), 879–897. https://doi.org/10.1353/hpu.0.0323.

Schoua-Glusberg, A. (2011). Response 2 to Fowler's chapter: Coding the behavior of interviewers and respondents to evaluate survey questions. In J. Madans, K. Miller, A. Maitland, & G. Willis (Eds.), *Question evaluation methods: Contributing to the science of data quality* (pp. 41–48). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118037003.ch4.

Schulz, A., Caldwell, C., & Foster, S. (2003). "What are they going to do with the information?" Latino/Latina and African American perspectives on the human genome project. *Health Education & Behavior, 30*(2), 151–169. https://doi.org/10.1177/1090198102251026.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. New York, NY: Psychology Press.

Shavers, V. L., Lynch, C. F., & Burmeister, L. F. (2001). Factors that influence African-Americans' willingness to participate in medical research studies. *Cancer, 91*(S1), 233–236. https://doi.org/10.1002/1097-0142(20010101)91:1+<233::Aid-cncr10>3.0.Co;2-8.

Shavers-Hornaday, V. L., Lynch, C. F., Burmeister, L. F., & Torner, J. C. (1997). Why are African Americans under-represented in medical research studies? Impediments to participation. *Ethnicity & Health, 2*(1-2), 31–45. https://doi.org/10.1080/13557858.1997.9961813.

Thomas, S. B., & Quinn, S. C. (1991). The Tuskegee Syphilis Study, 1932 to 1972: Implications for HIV education and AIDS risk education programs in the Black community. *American Journal of Public Health, 81*(11), 1498–1505. https://doi.org/10.2105/ajph.81.11.1498.

van der Zouwen, J., & Smit, J. H. (2004). Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences: A diagnostic approach. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 109–130). New York: Wiley. https://doi.org/10.1002/0471654728.ch6.

Warnecke, R. B., Johnson, T. P., Chávez, N., Sudman, S., O'Rourke, D. P., Lacey, L., & Horm, J. (1997). Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology, 7*(5), 334–342. https://doi.org/10.1016/S1047-2797(97)00030-6.

Washington, H. A. (2006). *Medical apartheid: The dark history of medical experimentation on black Americans from colonial times to the present*. New York: Random House.

West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175–211. https://doi.org/10.1093/jssam/smw024.