Assessing Racial Health Inequality in Older Adulthood: Comparisons From Mixed-Mode Panel Interviews

Jessica A. Kelley-Moore
University of Maryland, Baltimore County.

**Objectives.** Estimates of the extent of health disparities among Black and White older adults are not consistent across studies. The purpose of this study was to systematically compare responses from Black and White older adults in telephone and face-to-face interviews in order to determine whether estimates of racial health inequality vary by survey interview mode.

**Methods.** By using data from a mixed-mode panel study, I compared estimates of changing health inequality for Black and White older adults collected from face-to-face and telephone interviews. I calculated trajectories of physical disability by using latent growth models across seven waves of data.

**Results.** Face-to-face interviews yielded consistently higher reports of disability relative to telephone interviews of the same persons. Black adults had significantly greater and increasing disability than did White adults for both interview modes. After adjusting for covariates, I found that Black and White older adults had parallel disability trajectories in face-to-face interviews but a widening gap in disability over time in telephone interviews.

**Discussion.** Researchers should judiciously consider whether estimates of racial health inequality—and change in disability more broadly—may be misleading because of interview-mode effects.

SOCIAL surveys have long been used in order to assess the existence and extent of inequality in society across a number of domains, including income, status attainment, and health. Researchers typically make use of respondents’ answers to survey questions in order to document inequality by gender, race or ethnicity, or other ascriptive social categories. The assessment of racial health inequality on a wide range of indicators often relies on respondent reports in surveys. There is inconsistency, however, across studies regarding the extent and nature of this inequality, particularly over time where it is unclear whether the racial health gap is shrinking or remaining constant. One potential source of variation that may be overlooked is group differences in response behavior across interview modes. Given the growing use of telephone surveys, could the extent of the racial gap in various quality-of-life outcomes have been underestimated in recent years? If yes, then the underestimation would appear as social change—falsely suggesting the amelioration of racial inequality.

Some research has shown that a discrepancy exists in responses between face-to-face and telephone interviews, the difference is greater in potentially vulnerable populations, such as older adults and racial minorities (Aquilino, 1994; Pruchno & Hayden, 2000). If this is the case, then what researchers know about the extent of racial health inequality, particularly in older adulthood, may be biased. Despite the importance of the issue, relatively few studies have addressed it.

The present research systematically compares responses of Black and White older adults who were interviewed in telephone and face-to-face interviews as part of a seven-wave panel study in order to determine if the estimates of racial inequality vary by interview mode. The analysis permits an examination of both initial level of disability and change in disability over time. In this article, I first consider recent estimates of racial health inequality among older adults and then turn to the potential consequences of interview mode for the accurate assessment of inequality.

**Measuring Inequality Accurately: The Case of Racial Health Disparity**

Research on health disparities is a crucial step in guiding efforts that will reduce and ultimately eliminate racial health inequality. By most measures, older Black Americans have poorer health (including higher mortality risk), more chronic and acute health conditions, and greater physical disability than do Whites (Martin & Soldo, 1997; National Center for Health Statistics, 2003). Studies have also shown that the Black/White discrepancy in life expectancy at age 65 has widened in the past half century (e.g., National Center for Health Statistics). Estimates of this disparity have been relatively stable across studies because information is often based on death records and census counts.

The measurement of the racial gap for other indicators of health status is more difficult, though, owing to differences in measurement, sampling, and analytic techniques. Widely varying estimates of the same health domain across studies make it difficult to test competing conceptual or theoretical explanations of health inequality. For instance, some longitudinal studies on physical disability have shown that the racial disability gap in later life continues to increase over time (Clark, 1997; Liao, McGee, Cao, & Cooper, 1999), consistent with the cumulative disadvantage perspective (Dannefer, 1987). Other studies have argued that this inequality is holding steady.
or even diminishing (Kelley-Moore & Ferraro, 2004; Mendes de Leon et al., 1997), supporting persistent inequality or selective survival explanations of health disparities. These inconsistent results make it impossible to determine definitively the extent of racial disparities in older adulthood.

Careful inspection of many of the studies of the racial gap in disability reveals that some were conducted with a mixed interview-mode design (Kelley-Moore & Ferraro, 2004; Mendes de Leon et al., 1997), whereas others used data that were collected in face-to-face interviews only (Ferraro, Farmer, & Wybraniec, 1997) or in telephone interviews only (Clark, 1997). Is it possible that the interview mode itself might shape estimates of the extent of racial health inequality? This question is important because claims of growing convergence in Black/White health inequality could actually be an artifact of interview mode. I will draw from the literature on surveying vulnerable populations in order to address this question and gauge its import on other groups of respondents.

Response Behavior of Vulnerable Populations

Assessing racial health inequality requires having adequate numbers of disadvantaged persons in the sample in order to ensure statistical power in comparisons (e.g., Black–White comparisons). Oversampling and procedures that heighten response rates among disadvantaged groups (e.g., use of bilingual interviewers) have greatly helped social science research on inequality. Once sampling is complete, however, there remain considerations regarding the validity of the responses across interview modes.

Previous research has revealed that discrepancies in reports from survey research may be greatest across interview modes for vulnerable populations such as older adults (Pruchno & Hayden, 2000; Weinberger et al., 1994), poor persons (Sullivan et al., 1995), and racial or ethnic minorities (Aquilino, 1994). Most research has shown that face-to-face interviews yield greater estimates of negative health behaviors, functional limitations, and depression relative to telephone interviews of these same populations. Additionally, belonging to more than one socially disadvantaged group simultaneously may increase the discrepancy in responses (Nebot et al., 1994).

When examining differences in responses across interview modes, particularly for vulnerable populations, one finds two distinct sources of variation. First, persons who are interviewed face-to-face may be different from persons who are interviewed over the telephone. A number of the studies that have compared responses across interview modes used samples that were not designed to investigate interview-mode effects. Although some surveys had planned mixed-mode designs (Fowler, Roman, & Di, 1998; Taylor, Wilson, & Wakefield, 1998), others resulted from samples with nonrandom interview-mode assignment. Such modifications were typically enacted to keep reluctant or hard-to-reach participants in the sample by engaging in face-to-face interviews in lieu of the assigned telephone mode (Nebot et al., 1994; Pruchno & Hayden, 2000; Sullivan et al., 1995). The endogeneity of interview mode, however, may lead to comparisons that confound data collection mode with vulnerability—persons with lower incomes, less education, and poorer health might be more likely to be retained in the sample with an alternative mode of interview. Whatever the case, nonrandom selection into an interview mode may obscure differences across modes.

The second source of variation is that some respondents may provide incongruent self-reports of health across interview modes. In telephone interviews, older adults are more likely to report “don’t know” or to require additional interviewer assistance, perhaps reflecting difficulty hearing or other vulnerabilities (Herzog & Rodgers, 1988). When the data are collected in a face-to-face encounter, the interviewer is able to employ visual aids and respond to non-verbal cues. Additionally, in the age of telemarketing and scams, telephone surveys—even from legitimate sources—may not engender trust among respondents. Indeed, minority group members and older adults tend to be suspicious of telephone callers who ask sensitive or personal questions and thus may be more likely to censor or temper their responses over the telephone (Bell, 1984). The physical presence of the interviewer may lend legitimacy to the study and create an atmosphere of trust that is not as easily achieved over the telephone.

There is also increasing concern that many measures that were once considered universal or global may not actually elicit comparable responses from different racial or ethnic groups (Ford, Havstad, & Hill, 2000). Respondent answers to questions about health status may reflect cultural norms about health disclosures, historic mistrust of the medical system, or cohort differences in sharing private information (Gamble, 1997). Knowing that there are Black/White differences in rates of participation and information disclosure, it is plausible that there may also be racial differences in response behavior across interview modes, potentially yielding inaccurate estimates of the amount of inequality in a population.

In light of these differences in response behavior, the aim of this study was to systematically examine whether interview mode influences estimates of physical disability for Black and White older adults. With data from a seven-wave panel study, the present research estimates the initial level and trajectory of disability over time. I have organized the analysis around three major questions. First, are there differences between Black and White older adults in the disability trajectory by interview mode? Second, if there are differences by interview mode, which mode reflects greater disability? Third, do these differences hold after adjusting the disability trajectories for key covariates? These three questions are important steps in answering the broader question of whether current estimates of racial health inequality may be misleading because of interview-mode effects.

METHODS

Sample

The present study used data from the Established Populations for Epidemiologic Studies of the Elderly, a multistage random sample of older adults in five counties of North Carolina (Cornoni-Huntley et al., 1993). The baseline sample, composed of 4,162 persons aged 65 or older, had a substantial subsample of Black older adults (54%). Respondents participated in seven annual interviews between 1986 and 1992. Interviews for Waves 1, 4, and 7 were face-to-face and took place in the home. These face-to-face interviews were longer in length than
subsequent telephone interviews and included more questions on more topics such as depression and clinical measurements of blood pressure. Researchers conducted the interviews for the remaining waves (Waves 2, 3, 5, and 6) over the telephone. All interviews at baseline took place in the home. At each of the following waves, however, some of the respondents could not be interviewed by the selected method and agreed to participate with the alternative data collection procedure. I eliminated these respondents from the present analysis because nonrandom selection of interview mode may obscure differences between modes. In the Wave 2 interview, 336 respondents agreed to be interviewed by the selected method and agreed to participate but who did not answer the ADL questions and interviewed but who did not answer the ADL questions from the present analyses 86 respondents who were alive being unable to perform all seven tasks (14). I eliminated all seven ADLs to create a continuous measure of physical disability ranging from needing no help with any task (0) to being unable to perform all seven tasks (14). I eliminated from the present analyses 86 respondents who were alive and interviewed but who did not answer the ADL questions at one or more waves. Reliability of the summary score is high (α = .89).

I used a number of covariates in these models in order to adjust the estimates of disability over time. There were four indicators of health status. Morbidity was the sum of seven prevalent health conditions: heart failure or heart attack, stroke, hypertension, broken bones, hip fracture, diabetes, and cancer. The summed morbidity scores ranged from 0 conditions to 7 conditions. I measured incontinence with an ordinal variable ranging from never having difficulty holding urine (1) to always having difficulty (5). I identified current smokers and past smokers in respective binary variables where 1 = the name of the variable and 0 = all others.

I measured socioeconomic status with education and income. Education ranged from completing fewer than eight years (1) to post college (6). Current total household income had 11 categories, from less than $1,000 per year (1) to greater than $40,000 per year (11). Of participants who were eligible for the final sample, 406 did not report their income at baseline. I estimated an algorithm that predicted values for these cases and imputed whole-number values. Demographic variables included female gender, Black, rural, currently married, currently

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coding</th>
<th>Total Sample (N = 2,387)</th>
<th>Black Adults (n = 1,310)</th>
<th>White Adults (n = 1,077)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability, W1</td>
<td>0 = none; 14 = high</td>
<td>.56 (1.57)</td>
<td>.61 (1.62)</td>
<td>.51 (1.51)</td>
</tr>
<tr>
<td>Disability, W2</td>
<td>0 = none; 14 = high</td>
<td>.50 (1.61)</td>
<td>.52 (1.58)</td>
<td>.47 (1.64)</td>
</tr>
<tr>
<td>Disability, W3</td>
<td>0 = none; 14 = high</td>
<td>.53 (1.58)</td>
<td>.56 (1.61)</td>
<td>.50 (1.55)</td>
</tr>
<tr>
<td>Disability, W4</td>
<td>0 = none; 14 = high</td>
<td>.83 (2.05)</td>
<td>.86 (2.02)</td>
<td>.79 (2.09)</td>
</tr>
<tr>
<td>Disability, W5</td>
<td>0 = none; 14 = high</td>
<td>.82 (2.00)</td>
<td>.92 (2.08)**</td>
<td>.70 (1.90)**</td>
</tr>
<tr>
<td>Disability, W6</td>
<td>0 = none; 14 = high</td>
<td>.88 (2.12)</td>
<td>.99 (2.22)**</td>
<td>.76 (2.00)**</td>
</tr>
<tr>
<td>Disability, W7</td>
<td>0 = none; 14 = high</td>
<td>1.11 (2.50)</td>
<td>1.23 (2.60)**</td>
<td>.95 (2.34)**</td>
</tr>
<tr>
<td>Female</td>
<td>1 = female; 0 = male</td>
<td>.65</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Black</td>
<td>1 = Black; 0 = White</td>
<td>.54</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Age</td>
<td>64–105*</td>
<td>73.55 (6.72)</td>
<td>73.62 (6.85)</td>
<td>73.48 (6.57)</td>
</tr>
<tr>
<td>Lives in rural area</td>
<td>1 = yes; 0 = otherwise</td>
<td>.44</td>
<td>.44</td>
<td>.45</td>
</tr>
<tr>
<td>Education</td>
<td>1 = ≤ 8 years; 6 = postcollege</td>
<td>2.03 (1.45)</td>
<td>1.74 (1.37)**</td>
<td>2.37 (1.47)**</td>
</tr>
<tr>
<td>Income</td>
<td>1 = &lt; $1,000; 11 = &gt; $40,000</td>
<td>5.36 (2.32)</td>
<td>4.53 (1.88)**</td>
<td>6.36 (2.42)**</td>
</tr>
<tr>
<td>Married</td>
<td>1 = yes; 0 = no</td>
<td>.38</td>
<td>.34***</td>
<td>.43***</td>
</tr>
<tr>
<td>Widowed</td>
<td>1 = yes; 0 = no</td>
<td>.49</td>
<td>.51**</td>
<td>.47**</td>
</tr>
<tr>
<td>Religious service attendance</td>
<td>1 = never; 6 = &gt; once per week</td>
<td>4.00 (1.72)</td>
<td>4.15 (1.57)**</td>
<td>3.82 (1.86)**</td>
</tr>
<tr>
<td>Lives alone</td>
<td>1 = yes; 0 = otherwise</td>
<td>.38</td>
<td>.35***</td>
<td>.41***</td>
</tr>
<tr>
<td>Incontinence</td>
<td>1 = never; 5 = all the time</td>
<td>2.02 (1.12)</td>
<td>2.01 (1.12)</td>
<td>2.03 (1.13)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>1 = yes; 0 = otherwise</td>
<td>.17</td>
<td>.16</td>
<td>.18</td>
</tr>
<tr>
<td>Past smoker</td>
<td>1 = yes; 0 = otherwise</td>
<td>.26</td>
<td>.23***</td>
<td>.30***</td>
</tr>
<tr>
<td>Morbidity, W1</td>
<td>Count of 6 conditions</td>
<td>1.28 (1.06)</td>
<td>1.27 (1.05)</td>
<td>1.29 (1.06)</td>
</tr>
</tbody>
</table>

Notes: Table data are means; standard deviations are presented in parentheses. Significant differences between Black and White adults were tested with t tests for continuous and ordinal variables and chi-square tests for binary variables.

*aOne subject was 64 at the baseline interview.
**Estimates are significantly different to p < .01.
***Estimates are significantly different to p < .001.
widowed, and lives alone. All of these were binary variables where $1$ = the name of the variable and $0$ = all others. Age was measured in years from 64 to 105. Religious service attendance was an ordinal variable ranging from never attend (1) to attend more than once per week (6).

### Analyses

Analyses for the present study proceeded in two main stages. First, I calculated unadjusted intercepts and slopes for disability by regressing disability on time. I estimated models for three groups: all seven interviews, face-to-face interviews only, and telephone interviews only. I plotted these slopes by survey wave for the total sample and then separately for Black and White older adults. In the second stage of analysis I adjusted the intercepts and slopes of disability for covariates in order to determine whether any racial disparities observed in the unadjusted disability trajectories persisted. By using latent growth models in structural equations, I calculated intercepts and slopes in three models: first for all seven waves together, and then separately for face-to-face interviews and telephone interviews (Mcaristle & Epstein, 1987). This stage of analysis was important because failing to adjust estimates for known covariates and predictors of disability may have led to overestimates of racial disparities in physical function over time. The models predicting the intercept and slope of disability each had one instrumental variable in order to avoid over-identification. Given the skewed distribution of the disability measures, I estimated all structural equation models on an asymptotic matrix of polyserial and polychoric correlations.

Attrition (primarily due to death) was significant in this sample of older adults. I adjusted the models with covariates for potential selection bias due to nonrandom attrition. This two-stage process uses a hazard rate instrument based on the inverse Mills ratio expressing the likelihood of not remaining in the study for all seven waves (Heckman, 1979). First, by using a probit equation, I estimated the likelihood of completing all seven waves of the study. Second, based on that likelihood, I calculated an inverse Mills ratio for each case so that high values indicated a strong likelihood of not completing the study. I entered this variable into the substantive model as a covariate (Berk, 1983).

### Results

The first research question asked if there were differences in the trajectory of disability by interview mode. Figure 1 shows the plot of the raw mean disability scores over time and three regression lines: one using all seven waves, one for face-to-face interviews only, and one for telephone interviews only. Participants reported lower disability scores in the telephone interviews immediately following a face-to-face interview. What appear to be the sharpest increases in disability occur between a telephone interview and a subsequent face-to-face interview. The plot of the actual disability scores (unadjusted) demonstrates downward curves between interviews for Waves 1 and 4 and again for interviews between Waves 4 and 7.

The regression line for all seven waves fell between the three face-to-face interviews and the four telephone interviews. All of the former were above the line, and the latter were below the line. For the purposes of illustration, I extended the disability trajectory calculated with telephone interviews to Waves 1 and 7 by estimating expected values on the regression line. As the results in Table 2 show, the $R^2$ value for the regression equation for all interviews was .85 and the intercept was .46. I calculated a second regression line using only the face-to-face interviews. The intercept was higher than for the first regression line, showing that face-to-face interviews consistently yielded higher disability estimates. The $R^2$ value of the second regression line was 1.00, which was a perfect linear association between disability and time, increasing an average of .09 disability units per year. The final regression line used only those interviews conducted over the telephone. Its intercept was .37, which was lower than the corresponding estimates of the intercepts for the lines using all seven waves and solely the face-to-face interviews. The $R^2$ value for the telephone-only equation was .96.

As both Table 2 and Figure 1 demonstrate, the trajectory of disability over time fit better for single-mode analyses than for all seven interviews together. Responses from face-to-face interviews had a consistently higher level of reported disability than responses from telephone interviews. An interesting finding is that there was a slight convergence of these trajectories over time. The slope was steepest for
indicated that Black older adults had significantly greater
increase in disability of .07 each year whereas Black older
adults were at .37 and White older adults were at .53. These were followed by both of the intercepts for the lines using all interviews. The bottom two intercepts were the telephone interview intercepts: Black older adults were at .59 and White older adults at .53. These were followed by both of the intercepts for the lines using all interviews. The bottom two intercepts were the telephone interview intercepts: Black older adults were at .59 and White older adults were at .53. There was little difference across racial groups in the level initially reported by interview mode. In fact, there was no significant difference between Black and White older adults on the intercepts for the face-to-face interviews. Likewise, there was no significant difference by race for the intercepts on telephone interviews.

The slopes of these lines were markedly different by race, however. When using all interviews, there was an increasing gap in physical disability over time between groups, with Black older adults experiencing apparent increases in disability at a faster rate than White older adults. Indeed the differences in the slopes of disability in Table 3 were larger for all three lines among Black older adults. For example, using just the face-to-face interviews, White older adults experienced an average increase in disability of .07 each year whereas Black older adults experienced an average increase of .10 in the same time period. Each of these slopes within interview mode was significantly different by race. Even though the intercepts were grouped by interview mode, the final disability trajectories indicated that Black older adults had significantly greater increases in disability over time—despite the mode of interview used.

The face-to-face interviews among White older adults (their highest line) were lower than the telephone interviews for Black older adults (their lowest line). The eventual convergence of telephone and face-to-face interviews occurred among Black older adults but not among White older adults.

The next step in the analysis was to determine whether the intra-individual differences in the disability trajectory by interview mode remained after controlling for covariates. Table 4 presents the model fit statistics and parameter estimates for three latent growth curve models of disability over time: one for all interviews, one for face-to-face interviews only, and one for telephone interviews only. Measures of model fit indicated that all three models fit the data well. The goodness-of-fit index and incremental fit index were both 1.00 for nearly all models. The root mean square error of approximation was well below the recommended threshold of .05 for all three models, with the lowest at .004 for the face-to-face interviews model (Kelloway, 1998).

The adjusted intercepts for disability presented in Table 4 were all slightly smaller than the unadjusted intercepts presented in Table 2 for the total sample. For example, after adjusting for covariates, the intercept in the model for all interviews was .46, compared with its unadjusted estimate of .51. In the model with all interviews, the adjusted slope (.08) was not significantly different from the unadjusted slope (.10). However, in the model with just the face-to-face interviews, the unadjusted slope (.09) was steeper than its adjusted slope (.06). This was a statistically significant difference ($p < .05$). Finally, the adjusted slope for the model with telephone interviews only (.09) was less steep than its unadjusted slope (.11), but it was not significantly different. The muted difference in slopes of disability across adjusted and unadjusted latent growth models is primarily due to the fact that the intercept of disability predicted the slope of disability in both models. This adjustment for baseline disability accounted for much of the variation in slopes over time.

Table 3. Model Statistics for the Trajectory of Disability Over Time
Separately for Black and White Older Adults, North Carolina Established Populations for the Epidemiologic Studies of the Elderly, 1986–1992

<table>
<thead>
<tr>
<th>Variable</th>
<th>Equation</th>
<th>$R^2$</th>
<th>$F$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black adults</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All interviews</td>
<td>$y = .47 + .11x^a$</td>
<td>.87</td>
<td>33.45 (1,5)</td>
</tr>
<tr>
<td>Face to face</td>
<td>$y = .59 + .10x^b$</td>
<td>.99</td>
<td>80.08 (1,1)</td>
</tr>
<tr>
<td>Telephone</td>
<td>$y = .36 + .13x^c$</td>
<td>.96</td>
<td>52.20 (1,2)</td>
</tr>
<tr>
<td>White adults</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All interviews</td>
<td>$y = .44 + .08x$</td>
<td>.80</td>
<td>20.30 (1,5)</td>
</tr>
<tr>
<td>Face to face</td>
<td>$y = .53 + .07x$</td>
<td>.98</td>
<td>40.33 (1,1)</td>
</tr>
<tr>
<td>Telephone</td>
<td>$y = .37 + .08x$</td>
<td>.98</td>
<td>84.79 (1,2)</td>
</tr>
</tbody>
</table>

*Significantly different slope compared to White adults for all interviews ($p < .01$).
*Significantly different slope compared to White adults for face-to-face interviews only ($p < .01$).
*Significantly different slope compared to White adults for telephone interviews only ($p < .01$).
Possible reasons for these differences between the unadjusted and adjusted slopes, as well as the consistency of the intercepts across models, can be seen in the estimates of the covariates displayed in Table 4. In the model predicting disability using all seven interviews, living alone and attending religious services more frequently predicted lower intercepts of disability. Persons with greater incontinence and those with more total prevalent morbidity had higher intercepts of disability. Individuals who were at greatest risk of attrition also had significantly greater disability intercepts. Older persons and those at highest risk of attrition had greater increases in disability over time.

In the model using only the face-to-face interviews, persons who attended religious services more frequently had lower overall intercepts of disability. Consistent with the model using all seven interviews, persons who had greater problems with continence and persons with greater morbidity had higher overall disability intercepts. Slopes of disability over time were steeper for older adults, individuals with more incontinence, and persons at highest risk of attrition. More frequent attendance at religious services predicted less steep disability trajectories. In the face-to-face interviews, two new covariates emerged as strong predictors of slower trajectories of disability over time: having higher income and being married. Neither of these was significant in the first model, but, after eliminating the telephone interviews, both covariates became significant. Finally, persons at highest risk of attrition had steeper disability slopes.

The third model estimated the intercept and slope of disability using only the telephone interviews. Consistent with the model using only face-to-face interviews, persons who attend religious services less often, individuals with greater incontinence, and persons with greater morbidity had higher disability intercepts. As for the slope, Black respondents and individuals who were older experienced the greatest increases in disability over time. Some of the same significant predictors appeared in this model as in the earlier models: religious service attendance, incontinence, and morbidity. For the model that estimates only telephone interviews, persons at highest risk of attrition had significantly higher disability intercepts and higher slopes over time.

In order to test for significantly different predictors across interview modes, I estimated a multi-group latent growth model by treating the face-to-face interviews as Group 1 and the telephone interviews as Group 2. I then reestimated the model with covariates. Each covariate that had been a significant predictor of the intercept or slope for at least one interview mode was individually constrained to be equal across models. Several of the constrained relationships produced a significantly higher chi-square value for one degree of freedom, indicating a worse fit and significantly different slope estimates across groups. Having lower incomes and not being married predicted significantly steeper disability trajectories over time in face-to-face interviews relative to telephone interviews. Religious service attendance was associated with lower disability over time in telephone interviews but not in face-to-face interviews. The influence of morbidity and incontinence on disability trajectories did not differ significantly by interview mode.

**DISCUSSION**

Accurate documentation of change in health domains is critical to assessing whether any progress has been made in
reducing racial health inequality. The present study used panel data from Black and White older adults with a mixed-mode design in order to examine the intra-individual differences in disability between face-to-face and telephone interviews. The results showed that face-to-face interviews yield consistently higher estimates of disability than those procured from telephone interviews. Regarding the unadjusted estimates, Black older adults had higher initial levels and steeper trajectories of disability than did White older adults in both telephone and face-to-face interview modes. Likewise, in both interview modes the racial disability gap widened, but the disparity was greater in telephone interviews over time. Adjusted for covariates, the disability trajectories of Black and White older adults were no longer significantly different in face-to-face interviews, but the racial disability gap continued to widen over time in telephone interviews. There are both theoretical and methodological implications for these disparate findings across interview modes.

One goal of research on racial health disparities in older adulthood is to accumulate evidence across health domains that will allow tests of competing conceptual frameworks of health inequality over the life course. There is disagreement across these frameworks regarding the degree of Black–White health disparities in older adulthood, with some researchers hypothesizing a growing health gap (e.g., Dannefer, 1987) and others arguing that the gap is remaining constant or even shrinking (Kelley-Moore & Ferraro, 2004). The results of this study indicate that interview mode presents an often unrecognized source of variability in response to health questions, leading to quite different conclusions concerning racial health disparities. The conclusion from just the telephone surveys is that health disparities continue to increase among Black and White older adults, consistent with the cumulative disadvantage perspective. However, there was no such widening among just the face-to-face interviews. Conclusions from the latter interviews would support a shrinking disability gap or, at the very least, persistent health inequality in older adulthood.

A mixed-mode panel design further complicates the study of racial health disparities. With a face-to-face baseline interview followed by a telephone interview, it would be difficult to show an increase in inequality because the follow-up estimates would likely be lower than those projected from the face-to-face interview(s). Researchers should cautiously consider claims of shrinking disparities from mixed-mode interview studies. The Established Populations for Epidemiologic Studies of the Elderly provides consistent evidence within mode that disability is increasing. The Black–White disability gap increases in these data, but this generalization would be hard to support without face-to-face follow-ups integrated into the study design.

Although in the present study the estimates of disability were lower by telephone, the difference between face-to-face and telephone disability estimates narrowed somewhat over time, especially among the Black older adults. For White respondents, the two telephone interviews following the face-to-face interviews always led to the appearance of less disability. Not so for Black respondents. For Waves 1–3, the pattern was similar to that for White participants, but by Waves 5 and 6, Black respondents’ disability reports were equal to or higher than the Wave 4 estimates gathered in the face-to-face interview. This intriguing finding shows that the responses to disability questions across interview modes became more concordant for Black older adults, perhaps due to the growing rapport that accompanied repeated interviewing. Generally, one of the keys to the inclusion and retention of minority participants in research, particularly Black older adults, is the development of trust. This, coupled with a familiarity with data collection procedures, may diminish the discrepancy in responses across interview modes (Nápolis-Springer et al., 2000). Black participants may begin a study with greater concern about the use of information obtained, but once rapport is established in a panel study, telephone interviews may work as well as face-to-face interviews for assessing health problems among Black older adults. For White persons, however, there appears to be no substantial gain in rapport by repeated interviewing.

Survey interview mode has further implications for the measurement of physical function among older adults, independent of race. Considering just the first three interviews would give the appearance of a decline in disability. Separating the data by interview mode, however, shows that this apparent decline is actually due to the use of telephone interviews following a face-to-face interview. Nevertheless, it is difficult to discern real change from interview-mode effects in the absence of repeated measures from each mode. Although disability tends to be progressive over time, recovery of ADL functioning is not uncommon among older adults as a result of the healing of broken hips, physical therapy after a stroke, or long-term adjustment to an amputation. Sharp spikes in disability can also occur with the onset of new health conditions, a fall, or terminal drop. The question is not whether change occurs, but how much variation in these scores can be attributed to actual change in function versus an artifact of a mixed-mode study design.

Indeed, several covariates differed significantly across modes, indicating that the predictors of disability were not identical for face-to-face and telephone interviews. Having lower incomes and not being married led to greater disability over time, but just for face-to-face interviews. In telephone interviews, religious service attendance predicted significantly lower disability levels and trajectories over time relative to the in-home interviews. Thus, the greatest discrepancies in reports of disability were among persons who were most socially vulnerable (low income, not married, infrequent religious service attendance), which is consistent with other studies (Pruchno & Hayden, 2000; Weinberger et al., 1994). Individuals with a greater number of health problems or with incontinence reported similar disability levels across interview modes. In fact, older age no longer predicted significantly higher initial levels of disability in either interview mode once morbidity was included in the model. This may indicate that declining health, including physical function, is a salient concern that outweighs other potential sources of variation.

The differences between face-to-face and telephone interviews observed in this study may be more far reaching than appears on the surface due to planned mixed-mode designs of some studies. Consider two national, federally funded panel surveys, the Longitudinal Study on Aging and the Longitudinal Study on Aging II. In both instances, the baseline survey was a face-to-face interview, conducted in tandem with the National Health Interview Survey, but researchers administered
all of the follow-ups by telephone. Based on the present findings, the only way one can support an assertion of disability decline is by mixing—then ignoring—modes of interviewing. To give another example, the panel study of Asset and Health Dynamics Among the Oldest Old switches from telephone to face-to-face interviews when the participant turns 80 years old. It may be judicious in that study to interview persons of advanced age in a face-to-face interview, but based on the findings of the present study, one should not be surprised to see an apparent increase in disability when switching from a telephone to a face-to-face interview.

Nonrandom missing data can affect the measurement of racial health disparities in any panel study, but it is particularly important for mixed-mode panel studies. Analyses that include only those who are healthy enough and willing to complete the panel study in the assigned interview mode may underestimate the extent of disparity over time. In order to counter this, I adjusted the estimates for two potential sources of selection bias: nonrandom attrition from the study and the requirement of an alternate interview mode (e.g., the need for a face-to-face interview instead of the assigned telephone interview). I found that respondents who participated in the assigned interview mode for all seven waves had lower initial levels and less steep trajectories of disability over time. Black respondents had a higher risk of both mortality and requiring an alternate interview mode, suggesting that the surviving and participating Black older adults were a more selectively healthy subsample. However, even after I had adjusted for selection bias, Black older adults continued to have significantly steeper trajectories of disability over time than did White older adults when examining just the telephone interviews. A more detailed examination of the impact of nonrandom attrition on disability appears in Kelley-Moore and Ferraro (2004).

Readers should consider two major limitations to this study when interpreting these results. First, the face-to-face and telephone interviews were not of equal length. Interviews for Waves 1, 4, and 7, which researchers conducted in the home, were longer and more in-depth than those for Waves 2, 3, 5, and 6, which investigators conducted by telephone. The differences in responses across waves could have been partially due to the respondents feeling more comfortable talking about physical function because there had been more questions leading up to that section of the interview. Future studies should confirm the present results using questionnaires that are of identical length and format across modes and waves. Second, the time lag between each wave was one year. Among older adults, especially the oldest participants, health can change drastically in one year. Two types of studies would be helpful in further examining the conclusions of this study derived from 1-year follow-ups. First, studies with shorter follow-up periods would be helpful because they would reduce the amount of change that would likely occur. Second, studies that ask about intervening events (e.g., illness or treatment episodes) could help identify changes that are plausibly related to the outcome rather than measurement procedures.

When measuring racial health inequality, researchers have often taken for granted the comparability of telephone and face-to-face interview modes. The present results show that mode of data collection can lead to variations not only in person-specific responses across modes of data collection but also in the calculation of entire trajectories over time, resulting in disparate conclusions about the persistence of inequality over time. Although racial health disparities grew over time within each interview mode for these data, analyses that ignore the effect of interview mode on subject responses would lead to an underestimate of racial inequality.

Acknowledgments

The data used in this article were made available by the Interuniversity Consortium for Political and Social Research. Neither the collector of the original data nor the Consortium bears any responsibility for the analyses or interpretations presented here. Special thanks to Ken Ferraro for his guidance in this project. I also appreciate the comments of Shalon Irving, Timothy Owens, Tetyana Pylypiv, John Stahura, Roland J. Thorpe, Jr., Janet Wilmuth, and Douglas Wolf on an earlier version of this article and the willingness of Dr. Linda K. George to share supplementary data.

Address correspondence to Jessica A. Kelley-Moore, Sociology and Anthropology, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250. E-Mail: jkm@umbc.edu

References


Received June 27, 2005
Accepted January 3, 2006
Decision Editor: Charles F. Longino, Jr., PhD