MEASURING EVERYDAY RACIAL/ETHNIC DISCRIMINATION IN HEALTH SURVEYS

How Best to Ask the Questions, in One or Two Stages, Across Multiple Racial/Ethnic Groups?[1]

Salma Shariff-Marco

Division of Cancer Control and Population Sciences, National Cancer Institute

Nancy Breen[2]

Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute

Hope Landrine

Center for Health Disparities Research, East Carolina University

Bryce B. Reeve

Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute

Nancy Krieger

Harvard School of Public Health, Harvard University

Gilbert C. Gee

School of Public Health, University of California, Los Angeles

David R. Williams

Department of Society, Human Development, and Health, Harvard School of Public Health; Departments of African and African American Studies and *of Sociology, Harvard University*

Vickie M. Mays

School of Public Health, University of California, Los Angeles and *Center for Research, Education, Training, and Strategic Communication on Minority Health Disparities, University of California, Los Angeles*

Ninez A. Ponce

School of Public Health, University of California, Los Angeles and *UCLA Center for Health Policy and Research*

Margarita Alegría

Harvard Medical School, Harvard University

Benmei Liu

Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute

Gordon Willis

Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute

Timothy P. Johnson

Survey Research Laboratory, Department of Public Administration and *Institute for Health Research and Policy, University of Illinois at Chicago*

Abstract

While it is clear that self-reported racial/ethnic discrimination is related to illness, there are challenges in measuring self-reported discrimination or unfair treatment. In the present study, we evaluate the psychometric properties of a self-reported instrument across racial/ethnic groups in a population-based sample, and we test and interpret findings from applying two different widely-used approaches to asking about discrimination and unfair treatment. Even though we found that the subset of items we tested tap into a single underlying concept, we also found that different groups are more likely to report on different aspects of discrimination. Whether race is mentioned in the survey question affects both frequency and mean scores of reports of racial/ethnic discrimination. Our findings suggest caution to researchers when comparing studies that have used different approaches to measure racial/ethnic discrimination and allow us to suggest practical empirical guidelines for measuring and analyzing racial/ethnic discrimination. No less important, we have developed a self-reported measure of recent racial/ethnic discrimination that functions well in a range of different racial/ethnic groups and makes it possible to compare how racial/ethnic discrimination is associated with health disparities among multiple racial/ethnic groups.

Keywords: Discrimination, Racism, Prejudice, Race/Ethnicity, Differential Item Functioning, Surveys, Split-ballot, Psychometric

INTRODUCTION

Recent research shows that self-reported racial/ethnic discrimination is related to risk for health problems, including mortality (Barnes et al., 2008), psychiatric disorders (Kessler et al., 1999; Landrine et al., 2006), and cardiovascular disease (Lewis et al., 2006; Krieger and Sidney, 1996). Numerous review articles have found that discrimination is related to these and other health issues across many racial and ethnic populations (Mays et al., 2007; Paradies 2006; Pascoe and Smart Richman, 2009; Williams and Mohammed, 2009). Reviews also have suggested that many existing measures may not fully capture the experiences of diverse racial/ethnic groups (Bastos et al., 2010; Gee et al., 2009). A major concern is whether current measures comprehensively and reliably capture the forms of discrimination experienced by diverse groups, and whether these groups report discrimination in the same way. To examine and compare how self-reported discrimination is associated with health, it is important to have a measure that conceptually and empirically captures racial/ethnic discrimination for the range of experiences of diverse groups.

Racial discrimination has a longstanding history in the United States, affecting different racial/ethnic groups in both similar and different ways. For example, stud-

ies have found that Asians, Blacks,[3] and Latinos encounter similar levels of racial discrimination when purchasing a home (Turner and Skidmore, 1999). However, other studies show important group differences. For instance, Blacks are more likely to be perceived as dangerous than Asian Americans and other ethnic groups (Bobo 2001), and Asian Americans (including those born in the United States) are perceived to be "perpetual foreigners" (Tuan 1998), an issue not often faced by Blacks. Importantly, these differences are matters of degree.

To measure this heterogeneity of experiences of discrimination among different racial/ethnic groups in a single instrument and compare it among population groups requires balancing the tension between comprehensiveness and parsimony. Comprehensively assessing the forms of discrimination faced by specific groups would require a large set of items that are detailed for specific groups, and it is expected that groups would vary in their reporting of such items. At the same time, population-based studies have practical limits of respondent burden, cost, and time. Obviously, the survey items should perform reliably across groups and yield group differences that reflect real differences that are not due to survey design or other artifacts.

In this article, we address two challenging questions in measuring racial/ethnic discrimination on population health surveys. First, can a single parsimonious item set be used to measure racial/ethnic discrimination across multiple racial/ethnic groups, and second, what are the implications of directly asking about race when measuring racial/ethnic discrimination? To address these challenges we designed a multi-year research project in which we are using mixed methods to develop and test a concise, valid and reliable multidimensional instrument for measuring self-reported racial/ethnic discrimination in population-based health surveys for multicultural and multilingual populations (Shariff-Marco et al., 2009). Our Discrimination Module measures four dimensions of racial/ethnic discrimination—recent everyday discrimination; lifetime discrimination; appraisal of discrimination as stressful; and usual responses to discrimination. In this paper, we focus on a subset of items on recent discrimination that is a modified version of the Everyday Discrimination Scale (Williams et al., 1997) because this part of our empirical work is especially well suited to address these two challenges.

Building on previous research, our measure conceptualizes exposure to racial/ethnic discrimination as a psychosocial stressor, and we anticipate that it will have psychological, physiological, behavioral, and material consequences that can adversely affect health behaviors and health status (Clark et al., 1999; Epel et al., 2004; Geronimus 2001, 2006; Krieger 2000, 2010; Mays et al., 2007; Seeman et al., 2002; Wildsmith 2002; Williams et al., 2010). Everyday discrimination refers to chronic, day-to-day discrimination including recurrent character assaults—such as being treated as inferior or dishonest—that is part of the social fabric in the United States (Desmond and Emirbayer, 2009; Essed 1990). We use Differential Item Functioning (DIF) to learn whether our subset of items taps into a single underlying concept (Reeve and Fayers, 2005; Teresi and Fleishman, 2007; Thissen et al., 1993). Further, we use the DIF results to try to identify whether different racial/ethnic groups are more likely to report experiencing different aspects of discrimination.

Based on previous literature, we hypothesize that overall, Blacks would report the highest levels of discrimination, followed by Latinos, Asian Americans/Native Hawaiians/Pacific Islanders and American Indians/Alaska Natives (among whom we do not pose specific ordering), and the lowest reporting among Whites. We expect variation for specific items and that, after accounting for overall levels of discrimination by group, Blacks would be more likely to report being treated as threatening, while Asians and Latinos would be most likely to report discrimination based on language and accent.

TWO KEY CHALLENGES IN MEASURING EVERYDAY DISCRIMINATION

One major challenge lies in validating measures across diverse populations. Most discrimination measures have been developed and validated for Black populations (Kressin et al., 2008). Recent reviews have found that scales specifically designed for Black populations have not been rigorously tested when adapted for other racial/ethnic groups (Bastos et al., 2010; Kressin et al., 2008). For instance, these scales seldom ask about discrimination based on language and accent, yet this type of discrimination may be particularly salient for population groups with large proportions of immigrants like Asian Americans and Latinos (Gee et al., 2009).

Our second challenge is to assess how to ask about discrimination experiences related to race/ethnicity. Two approaches are common. One approach asks explicitly about experiences with racial/ethnic discrimination in one question (one-stage approach). A second approach divides this line of questioning into two parts and initially uses more neutral language by asking about "unfair treatment" (two-stage approach). Hence, these two approaches make fundamentally different assumptions about how best to query respondents. With the one-stage approach, the intent of the question is clearly focused on racial/ethnic discrimination. The two-stage approach allows participants to report on racial/ethnic as well as other types of discrimination. Both approaches have been shown to have good inter-item reliabilities, and both are correlated with health problems. However, these approaches tend to yield different levels of self-reported exposure (Brown 2001; Chae et al., 2008; Krieger et al., 2005), and it is unclear whether they are empirically equivalent.

Two competing hypotheses propose how the approaches may differ empirically. One hypothesis is that the two-stage approach would yield more accurate—and by implication higher—reporting of racial/ethnic discrimination because it does not require respondents to engage in the challenging cognitive task of attributing cause as they engage in the task of recalling, discerning, and reporting experiences of discrimination (Williams and Mohammed, 2009); the one-stage approach requires respondents to attribute discrimination to race/ethnicity at the same time as they report frequency, and this, the authors hypothesize, would result in underreporting experiences of discrimination. A separate and competing hypothesis suggests that including the attribution of race/ethnicity in the stem of the question will result in higher reports of racial/ethnic discrimination due to the highly sensitized terminology and because this approach avoids the challenge of attributing discrimination to other specific factors (e.g., race vs. gender) (Bastos et al., 2010).

To address these unanswered empirical questions, we conducted, through telephone surveys, a head-to-head comparison of the one-stage versus two-stage approaches used to measure self-reported recent everyday racial/ethnic discrimination. We compared these two approaches. For the two-stage approach, we further examined responses separately using only the first stage without specifying attribution to unfair treatment and using the second stage to specify attribution to racial/ethnic discrimination.

METHODS

Overview

A subsample of respondents to the 2007 California Health Interview Survey (CHIS) was selected to receive the Discrimination Module (DM). CHIS is a random digit

dial telephone survey. The total adult sample in 2007 was 51,048.[4] The overall response rate for adults (from the landline/list sample), 18.7%, is comparable to the California Behavioral Risk Factor Surveillance System telephone survey (California Health Interview Survey 2009). The overall response rate is the product of the screener and extended interview response rates, which were 35.5% and 52.8%, respectively in 2007.[5]

Based on prespecified sampling probabilities, 7505 adults were randomly assigned to the one- or two-stage versions of the DM questionnaire. Among these adults, 104 quit the CHIS study before being asked about discrimination, and another 435 did not provide codable information on race/ethnicity (e.g., reported "human race"), resulting in an analysis sample of 6966 respondents (2254 Whites; 1800 Latinos; 1006 Black or African Americans; 190 American Indian/Alaska Natives, 1231 Asian American Native Hawaiian or Pacific Islanders, and 485 Multiracial respondents).[6] Of these respondents, 3506 were assigned to the one-stage version, and 3460 were assigned to the two-stage version.

The CHIS 2007 Recent Everyday Racial/Ethnic Discrimination Module Field Test

We present findings on the eight items that assess recent everyday discrimination experiences in the past twelve months. Items were adapted from the Everyday Discrimination Scale (Williams et al., 1997), with changes in the wording of both specific items and response categories. For instance, the "courtesy" item from the original scale was found to be redundant with "respect" and dropped (Reeve et al., Forthcoming). We added a language item because previous research has found that language discrimination is an important type of racial discrimination faced by ethnic groups (Gee et al., 2009; Spencer and Chen, 2004). It is important to note that language discrimination can occur in the absence of factors related to immigration. For instance, Massey and Lundy (2001) found that individuals speaking "Black English Vernacular" or "Black Accented English" over the telephone faced more discrimination in renting an apartment than individuals speaking "White Middle-Class English." Further, our analyses from a separate behavior-coding study suggested that a simplified response category (0 = never to 3 = often) was appropriate for telephone-administered surveys (Reeve et al., forthcoming). Participants were asked the following questions: "In the past 12 months, how often have . . .

(1) you been treated with less respect than other people?
(2) you been treated unfairly at restaurants or stores?
(3) people criticized your accent or the way you speak?
(4) people acted as if they think you are not smart?
(5) people acted as if they are afraid of you?
(6) people acted as if they think you are dishonest?
(7) people acted as if they're better than you are?
(8) you been threatened or harassed?"

The two approaches are as follows:

One-stage: With this approach, each item above directly attributes discrimination to race/ethnicity in the stem. For example, "Have you been treated with less respect than other people *because [of your race/ethnicity]*?"

Two-stage: With this approach, each item above asks about unfair treatment in an initial question and then asks for attribution. For example, "Have you been treated

with less respect than other people?" If yes, the respondent is asked a series of questions to attribute the experiences of unfair treatment to ancestry or national origin, gender or sex, race or skin color, age, the way he or she speaks English, or some other reason (specify). For example, "Now, I'm going to ask you why you may have been treated unfairly. In the past 12 months, were you treated unfairly because of your ancestry or national origin?"

The two-stage approach yields two sets of results for everyday discrimination: One is unattributed (using information from the first stage only), and a second is attributed to race/ethnicity (using information from both the first and the second stage). To construct the second group, attributions of ancestry or national origin, race or skin color, and the way he or she speaks English were ascribed to *racial/ ethnic* discrimination. Responses attributing discrimination to something other than race/ethnicity or "never" experiencing everyday racial/ethnic discrimination were recoded to "no racial/ethnic discrimination" in order to maintain them in our analytic sample.

We categorized our data into three groups for analysis: (1) two-stage unattributed; (2) two-stage attributed to race/ethnicity; and (3) one-stage. We used the demographic items on CHIS to categorize the DM sample by race/ethnicity. Anyone who self-identified as Hispanic or Latino was categorized as "Latino." Others were identified as non-Hispanic White, Asian American Native Hawaiian or Pacific Islander (AANHPI), African American/Black, American Indian/Alaska Native (AI/AN) or Multiracial. AANHPIs were aggregated due to small samples. AI/AN were counted differently in order to maximize the likelihood of including them in this group. All respondents who made *any mention* of AI/AN were included in this category (Lee et al., 2009; Mays et al., 2003; Swan et al., 2006). Multiracial respondents formed their own category with the exception of AI/AN respondents.

Data Analysis

Psychometric Properties

We performed a confirmatory factor analysis (CFA) in our total sample and within each racial/ethnic group. Based on prior literature, we hypothesized a one-factor structure (Kessler et al., 1999; Krieger et al., 2005; Williams et al., 1997). Factor analyses were used to test this hypothesis on data without sampling weights using Mplus (version 4.21) software (Muthen and Muthen, 2007). Model fit was assessed by examining the comparative fit index (CFI > .95), Tucker-Lewis Index (TLI > .95), and the Root Mean Squared-Error of Approximation (RMSEA < .08). Inter-item reliability was assessed using Cronbach's alpha (Nunnally and Bernstein, 1994).

We then used Differential Item Functioning (DIF) analysis to determine if individual questions perform differently across racial/ethnic groups after controlling for other items measured in the construct. Scales containing items that exhibit DIF may have reduced validity for between-group comparisons because scores may be indicative of attributes other than those that the scale is designed to measure (Roussos and Stout, 1996; Teresi and Fleishman, 2007; Thissen et al., 1993). If an item displays DIF, the DIF item may be kept in the scale if it does not affect the overall score; or, if it does, the item may be removed (Reeve et al., 2007; Teresi and Fleishman, 2007).

DIF tests were performed comparing racial/ethnic groups in pairwise fashion using the Item Response Theory (IRT)-based likelihood-ratio (IRT-LR) method (Thissen

et al., 1993) with the IRTLRDIF (version 2.0b) software program (Thissen 2001). Each item in the scale was tested for DIF using the remaining seven items as anchors to adjust for the differences in experiences of discrimination. Sample sizes were too small to examine DIF for AI/ANs and Multiracial respondents even after recoding.

Prevalence Estimates

We computed weighted percentages and the weighted mean frequencies of recent everyday discrimination reported in the 2007 CHIS DM field test for each approach by race/ethnicity using SAS-callable SUDAAN (version 10, Research Triangle Institute 2008). Special weights were developed for the DM sample (details available from authors). Though missing data for the module items were small, ranging from 0.7 to 2.3 percent, to minimize loss of data for subsamples, missing data were imputed using the AutoImpute macro (Judkins et al., 2007, 2008; Piesse et al., 2005).

RESULTS

Table 1 presents sample characteristics, stratified by approach. Demographic characteristics (e.g., race/ethnicity or gender) were similar across the two approaches, indicating that the randomization yielded balanced samples.

Psychometric Findings

We obtained a good fit for a single factor solution (CFI = .99; TLI = .99; and RMSEA = .05) for both of the samples shown in Table 1. Factor analyses by racial/ethnic subgroups yielded similar results. These findings indicate that the subset of the eight DM questions tap into a single underlying dimension of recent discrimination.

Cronbach's alpha for the one-stage and the two-stage approach were 0.88 and 0.81, respectively. While both of these estimates are considered acceptable for *group* level measurement (> .70), the one-stage approach is closer to the acceptable standard for *individual* level assessment (>.90) (Nunnally and Bernstein, 1994).

Table 2 summarizes the pairwise DIF findings for the one-stage approach only for parsimony in presentation. DIF results were similar across all three groups (two-stage unattributed; two-stage attributed to race/ethnicity; and one-stage). Small DIF was identified for several items, but no single item had DIF that affected the overall group mean by more than .03 standardized units on a scale with a mean of zero and standard deviation of one. In short, across approaches and racial/ethnic groups, DIF was not large enough to consider deleting any item from the scale.

Even though no item has DIF large enough to delete it from the scale, some interesting pair-wise patterns in DIF were discerned:

African Americans and Latinos

DIF was found for questions (2) ". . . treated unfairly at restaurants/stores" and (5) ". . . people acted as if they are afraid of you." As the reported level of overall recently experienced racial discrimination increased, African Americans were more likely to endorse these two items than Latinos at any given level of discrimination. Small DIF was also detected pertaining to three other items where Latinos were more likely to

Table 1. Sample Characteristics for One- and Two-Stage Approaches,[a] 2007 CHIS Discrimination Module Field Test

Sample Characteristics	1-Stage[b]		2-Stage[c]	
	N	Weighted % (Mean)	N	Weighted % (Mean)
Total	3506	*100.0*	3460	*100.0*
Race/Ethnicity				
White	1154	*59.2*	1100	*58.8*
Latino	908	*18.6*	892	*18.3*
African American	516	*7.1*	490	*6.8*
American Indian/Alaska Native	115	*1.1*	75	*1.0*
Asian American/Native Hawaiian/ Pacific Islander	598	*11.3*	633	*11.4*
Multiracial	215	*2.7*	270	*3.8*
Gender				
Male	1378	50.7	1380	51.8
Female	2128	49.3	2080	48.2
Marital Status				
Married	1768	56.7	1741	56.5
Living with Partner	178	5.7	179	5.7
Wid/Div/Sep	878	13.5	847	13.5
Never Married	682	24.1	693	24.3
English Proficiency				
Very Well	2941	86.7	2908	88.8
Well	494	11.7	479	9.6
Not Very Well/Not at All	71	1.6	73	1.6
Born in USA				
Yes	2626	80.0	2557	80.3
No	880	20.1	903	19.7
Insurance Status				
Yes (Insured Past 12 Mos)	3046	85.2	3053	85.9
No	460	14.8	407	14.1
General Health				
Excellent	684	23.0	668	21.8
Very Good	1126	33.5	1144	35.9
Good	1085	30.3	1053	27.7
Fair	428	10.1	430	11.9
Poor	183	3.1	165	2.7
Age (mean, in years)	3506	45.1	3460	45.3
Education (mean, years completed)	3506	14.4	3460	14.3
Income (below the U.S. poverty line)	3506	5.6	3460	5.9
% Life in United States	3506	91.2	3460	91.2

[a]Chi-square tests and t-tests were performed to test the difference of the distributions of each sociodemographic variable in the two samples. No statistically significant difference was found at alpha=0.05 level.
[b]Explicit questions on racial discrimination.
[c]Questions about experiences conceptualized as discriminatory followed by attribution to racial/ethnic attributes.

endorse them: numbers (3), "criticized accent or the way you speak"; (4), ". . . think you are not smart"; and (7), ". . . acted as if they're better than you."

African Americans and AANHPIs

DIF was found for questions (1), ". . . treated with less respect"; (3), ". . . people criticized your accent or the way you speak . . ."; (7), ". . . acted as if they're better

Table 2. Results from Differential Item Functioning (DIF) for the One-stage Approach,[a,b] 2007 CHIS Discrimination Module Field Test

Recent Discrimination	African-American— Latino	African-American— AANHPI	AANHPI— Latino	White— African-American	White— AANHPI	White— Latino
1. Less Respect		B	B	W		W
2. Unfairly at Restaurants/Stores	A			A		
3. Criticized Accent or Speech	L	B	B		B	
4. People Think Not Smart	L		L			L
5. People Afraid of You	A	A		A	W	
6. People Think Dishonest		A			W	
7. People Think Better than You	L	B	L			L
8. Threatened/Harassed		B	B	W	W	W

[a]The letters represent the racial/ethnic groups that endorsed an item more frequently based on the DIF analyses: A-African American, B-Asian American/Native Hawaiian/Pacific Islander, L-Latino. W-White. Item Response Theory—Likelihood Ratio method used to test for DIF. All significant DIF findings are reported when p < .01.
[b]A small amount of DIF was identified for several items, but, between two groups, no single item had DIF that affected the overall group mean by .03 or more standardized units.

than you"; and (8), ". . . been threatened/harassed." The DIF findings suggest that for African Americans and AANHPIs with similar levels of discrimination, AANHPIs were more likely to endorse these items than African Americans. A small level of DIF was also found between African Americans and AANHPIs for questions (5), ". . . people acted as if they are afraid of you," and (6), ". . . think you are dishonest," with African Americans more likely to endorse them.

AANHPIs and Latinos

DIF was found for questions (1), ". . . you have been treated with less respect than other people . . ."; (3), ". . . people criticized your accent or the way you speak . . ."; and (8), ". . . been threatened or harassed . . .". AANHPIs were more likely to endorse these items than Latinos, despite having experienced the same level of discrimination. DIF was also found for two other items with Latinos more likely to endorse them: numbers (4), pertaining to ". . . think you are not smart," and (7), pertaining to ". . . acted as if they're better than you."

Whites and All Other Groups

DIF was found for question number (8), ". . . been threatened or harassed. . . ." Whites who reported recent everyday racial discrimination, after we controlled for differences in the frequency of racial discrimination they experienced, were more likely to endorse the threatened/harassed question than other groups.

Across the set of pairwise findings, several race/ethnicity specific patterns emerged. African Americans were more likely to endorse item number (5), "People acted as if they are afraid of you." Latinos were more likely to endorse item number (4), "People acted as if they think you are not smart"; or number (7), "People acted as if they're better than you." AANHPIs were more likely to endorse number (3), "People criticized your accent or the way you speak." Whites were more likely to endorse

number (8), ". . . been threatened or harassed." These patterns were also observed for the two-stage approach sample, except for Whites.

Results of Using the One-Stage and Two-Stage Approaches to Measure Everyday Discrimination Experiences

Tables 3 and 4 compare reports of recent discrimination for three groups: two-stage unattributed, two-stage attributed to race/ethnicity, and one-stage. Table 3 presents percentages of reports of any experience of recent discrimination, and Table 4 presents mean scores across the eight items. The first two columns of both tables show the estimates for respondents from the two-stage sample. The first column shows the percentage reporting any kind of discrimination using the first stage of the two-stage approach and the second column shows the percentage reporting discrimination attributed to their race or ethnicity as described in the methods section above. The third column shows the percentage of racial/ethnic discrimination reported by respondents from the one-stage sample who were directly asked whether they experienced recent racial/ethnic discrimination.

For the two-stage approach, 2638 respondents reported experiencing some form of recent everyday discrimination. Of them, 941 (or 27%) reported this was due to race/ethnicity. This varied by racial/ethnic group, with higher proportions attributing discrimination to race/ethnicity among the non-White groups: African Americans (62%); Latinos (43%); AANHPIs (45%); AI/ANs (37%); and Multiracial respondents (41%). Eleven percent of Whites attributed discrimination to racial/ethnic characteristics.

Table 3 shows the percentage of respondents who reported recent everyday discrimination by approach and by racial/ethnic group. As expected, the two-stage unattributed yielded the highest percentage of reports with at least 75% of respondents in each racial/ethnic group having endorsed at least one item. The highest percentage was among Multiracial respondents (87%) followed by African Americans (85%) and the lowest was among Whites (75%).

For discrimination due to race/ethnicity, the one-stage approach yielded a higher percentage of reports than the two-stage approach. Those who were asked using the one-stage approach reported more racial/ethnic discrimination—African Americans (83%), Whites (35%)—than those who were asked using the two-stage approach with attribution for racial/ethnic discrimination specified—African Americans (53%), Whites (9%).

Differences in self-reported exposures between racial/ethnic groups were smaller for two-stage unattributed than for either two-stage attributed to race/ethnicity or one-stage. Subgroups with small sample sizes (AI/ANs and Multiracial respondents) had wide, overlapping 95% CIs around their point estimates. Of the total fifteen pairwise t-tests we conducted, we found a higher number of statistically significant different pairs in discrimination estimates between racial/ethnic groups in the one-stage approach sample (ten pairs) than in the two-stage approach sample (five pairs). The two-stage attributed to race/ethnicity was more similar to the one-stage with nine pairs. This was especially true when comparing Whites to other populations of color.

Table 4 shows mean scores of recent everyday discrimination by approach and by racial/ethnic group. Mean scores may range from 0–3. For both the total and by racial/ethnic group, mean scores were highest for two-stage unattributed followed by one-stage and two-stage attributed to race/ethnicity. Across these groups, African Americans had the highest mean scores (0.76, 0.75, 0.53) and Whites had the lowest (0.42, 0.15, 0.08). The spread of scores between groups was larger in one-stage than

Table 3. Percentage of People Who Reported Any Recent Discrimination by Approach and Race/Ethnicity,[a] 2007 CHIS Discrimination Module Field Test

Racial/Ethnic Group	Two-Stage Unattributed Weighted % (95% CI)	Two-Stage Attributed to Race/Ethnicity Weighted % (95% CI)	One-Stage Weighted % (95% CI)
Total	77.4 (74.7,79.8)	20.6 (18.7,22.5)	48.8 (45.9,51.6)
White	74.5 (70.3,78.3)	8.5 (6.4,11.2)	35.2 (31.0,39.7)
Latino	79.2 (75.1,82.7)	33.9 (29.0,39.2)	66.6 (61.8,71.1)
African American	85.2 (81.1,88.6)	52.9 (46.9,58.8)	82.7 (77.8,86.8)
American Indian/ Alaska Native	81.8 (56.0,94.1)	30.5 (16.1,50.1)	47.1 (31.4,63.3)
Asian American/ Native Hawaiian/ Pacific Islander	81.0 (77.0,84.5)	36.7 (31.9,41.7)	67.8 (62.2,72.9)
Multiracial	87.2 (81.5,91.4)	35.8 (27.4,45.2)	54.6 (42.4,66.3)

[a]Pairwise t-test were performed to test the differences of the estimates reported above between different racial/ethnic groups within each approach. The pairs that have statistically significant difference in recent discrimination include:
For the one-stage: White-Latino, White-African American, White-AANHPI, White-Multiracial, Latino-African American, Latino-AI/AN, African American-AI/AN, African American/AANHPI, African American-Multiracial, AI/AN-AANHPI.
For the two-stage: White-African American, White-AANHPI, White-Multiracial, Latino-African American, Latino-Multiracial;
For the two-stage specified sample, White-Latino, White-African American, White-AI/AN, White-AANHPI, White-Multiracial, Latino-African American, African American-AI/AN, African American-AANHPI, African American-Multiracial.

in either two-stage analytic samples (attributed and unattributed). Each approach had a similar number of pairs of statistically significant differences.

DISCUSSION

Our study provides new evidence for understanding the two common approaches for measuring racial/ethnic discrimination among multiple racial/ethnic groups. Our results include three central findings. First, our measure of recent everyday discrimination showed small DIF across groups; the DIF that we found for specific racial/ethnic groups was in accordance with what we would expect based on the literature. This suggests that our measure is meaningful for use across different groups. Second, responses to the two-stage unattributed approach yield the highest levels of exposure to everyday discrimination (with no attribution to race/ethnicity) and smallest variation in reported exposure by racial/ethnic group. Third, estimates of self-reported racial discrimination are higher when using the one-stage approach than when using the two-stage approach with attribution to race/ethnicity specified.

Implications

Measurement Challenge 1: Is a Single Measure Valid for Measuring Everyday Discrimination Across Diverse Racial/Ethnic Groups?

Our DIF analysis suggested that our measure of recent everyday discrimination performed similarly across diverse racial/ethnic populations. All DIF was small, and

Table 4. Mean Scores for Recent Experiences of Discrimination by Approach and Race/Ethnicity,[a] 2007 CHIS Discrimination Module Field Test

Racial/Ethnic Group	Two-Stage Unattributed Weighted Mean Score (95% CI)	Two-Stage Attributed to Race/Ethnicity Weighted Mean Score (95% CI)	One-Stage Weighted Mean Score (95% CI)
Total	0.50 (0.47,0.52)	0.19 (0.17,0.21)	0.29 (0.27,0.31)
White	0.42 (0.38,0.46)	0.08 (0.05,0.10)	0.15 (0.12,0.17)
Latino	0.56 (0.51,0.62)	0.30 (0.25,0.35)	0.47 (0.42,0.53)
African American	0.76 (0.67,0.85)	0.53 (0.44,0.62)	0.75 (0.66,0.84)
American Indian/Alaska Native	0.64 (0.41,0.87)	0.28 (0.15,0.40)	0.40 (0.22,0.57)
Asian American/Native Hawaiian/Pacific Islander	0.55 (0.49,0.60)	0.33 (0.28,0.39)	0.45 (0.39,0.50)
Multiracial	0.72 (0.62,0.82)	0.37 (0.26,0.48)	0.39 (0.28,0.51)

[a]Pairwise t-test were performed to test the differences of the estimates reported above between different racial/ethnic groups within each approach—The pairs that have statistically significant difference in mean scores of recent discrimination include:
For the one-stage: White-Latino, White-African American, White-AI/AN, White-AANHPI, White-Multiracial, Latino-African American, African American-AI/AN, African American-AANHPI, African American-Multiracial;
For the two-stage: White-Latino, White-African American, White-AANHPI, White-Multiracial, Latino-African American, Latino-Multiracial, African American-AANHPI, AANHPI-Multiracial;
For the two-stage specified sample:, White-Latino, White-African American, White-AI/AN, White-AANHPI, White-Multiracial, Latino-African American, African American-AI/AN, African American-AANHPI, African American-Multiracial.

any DIF was cancelled out in the aggregate. The most useful DIF was found in items where there was precedent to anticipate a priori group differences. For example, Blacks were more likely to endorse the item ". . . they are afraid of you." Discrimination based on people's fear of one particular person or group, more salient for Blacks in our study, can be expected when groups are singled out and stereotyped as dangerous and criminal, such as African American males are in the United States (Feagin 2006; Krieger 2000; Pincus 2003; Williams and Neighbors, 2001). A national survey, for example, reported that, in 1990, more than 50% of White Americans stereotyped Black Americans as being prone to violence; White Americans' beliefs that other groups were prone to violence was under 20%, except for Hispanics at 38% (Williams and Neighbors, 2001).

Asians and Latinos were more likely to endorse "criticized accent or the way you speak." Our study confirmed more language and accent discrimination reported by AANHPIs and Latinos, and this is not surprising since these groups have a large proportion of immigrants from non-English-speaking countries (Cameron 1997; Spencer and Chen, 2004; Yoo et al., 2009, 2010). This form of discrimination may include both overt discrimination based on English-only policies, and patronizing assumptions, such as those revealed in statements like, "you speak English so well" to native speakers who are of Asian or Latino ancestry (Liang et al., 2004). This is consistent with similar points raised in the literature that Latinos and Asians are seen as perpetual foreigners (Gee et al., 2009; Tuan 1998).

The one unexpected finding was that Whites were more likely to endorse being threatened/harassed because of their race/ethnicity. It is unclear whether this pattern

revealed in our DIF analysis is valid versus an artifact due to measurement error, item construction, or other problems with the instrument. Given the extensive research documenting that racially motivated hate crimes are overwhelmingly directed against persons of color (Feagin 2006; National Research Council 2004), this item will require further study, building on the social science literature on reverse discrimination and the social construction of Whiteness (Feagin and O'Brien, 2003; Kluegel and Bobo, 2001; Pincus 2003).

Using the same measure to examine recent everyday discrimination across diverse racial/ethnic groups appears to be valid as long as the measure taps into the range of discrimination experienced by the populations studied. Thus, we recommend including an item pertaining to language discrimination because this item resonates with groups that include large proportions of immigrants from other countries where American English is not the primary spoken language. An area for further research is whether a language item would also resonate with Black Americans who speak Black Accented English or who are from countries where American English is not the primary spoken language. While the effect of DIF on the overall score was minimal on our eight-item scale, if the scale had fewer items, then the DIF for any of these items could have a larger impact. This is another area for further research.

Measurement Challenge 2: Do the One-Stage and Two-Stage Approaches for Measuring Everyday Discrimination Produce Similar Results?

Our study is also the first that we are aware of to directly compare the two approaches for measuring recent everyday racial/ethnic discrimination, using identically worded questions, in a single study. We compared results across three groups: two-stage unattributed; two-stage attributed to race/ethnicity; and one-stage. The data show that approach matters. As hypothesized by Williams and Mohammed (2009), the first stage of the two-stage approach yields a higher percent of respondents reporting any recent unfair treatment (overall, not attributed to race/ethnicity) than the one-stage approach that asks directly about racial/discrimination. This pattern holds when we look at the average level of reporting of discrimination. However, when we restrict the two-stage to focus on discrimination attributed to race/ethnicity, the one-stage approach yields a higher percentage of people reporting this discrimination, and a higher average. The latter finding supports the hypothesis presented by Bastos et al. (2010).

Together, these results underscore the difficulties that may arise when comparing results from studies that analyze data on recent "unfair treatment" without attribution with results from studies that analyze data from questions that explicitly ask about racial discrimination. The greatest danger is when studies using the two measures attempt to compare prevalence estimates of recent self-reported discrimination. These estimates will differ, although no way has yet been found to ascertain which estimate is more accurate. Accordingly, researchers are encouraged to state clearly which approach is used when describing and comparing estimates of self-reported discrimination.

Study Limitations

Proper interpretation of our study findings requires the consideration of several limitations. First, our sample size was not adequate to stratify our DIF analysis by multiple characteristics. For example, studying DIF for racial discrimination questions when controlling for other types of social characteristics (e.g., gender, educa-

tion, and acculturation) would be useful, since these characteristics may also affect self-reports of experiences of racial discrimination (Krieger 2000; Landrine et al., 2006; Paradies 2006; Teresi and Fleishman, 2007; Williams and Mohammed, 2009).

Second, we do not know whether respondents from different racial/ethnic groups bring the same meaning to the response categories. For example, when White, Black, and Latino respondents endorse "sometimes," do they have the same frequency in mind? Likewise, when these different groups report having been "threatened or harassed" due to their race/ethnicity, are they describing similar experiences of the same intensity or chronicity? Additional research is required to address these questions, both cognitively and in relation to their impact on measuring associations between racial/ethnic discrimination and health status.

Third, our comparisons of the two different approaches are based on group averages rather than direct comparisons of individual respondents. To directly compare how individuals responded to the two approaches, we could have designed our study to ask every respondent about discrimination using both approaches in a randomly assigned order. No studies have done this, although a few studies have administered different discrimination instruments (e.g., both one-stage and two-stage) to the same respondents (Brown 2001; Chae et al., 2008; Krieger et al., 2005). Doing so would have been impractical in the context of the CHIS, where the repetition of questions could frustrate respondents (the overall discrimination module has twenty-five to thirty-five items). However, such a within-respondent comparison would be an important complement to our current work.

A fourth limitation is that our study could not address the extent to which the perceived race/ethnicity of the interviewer affected levels of reported racial discrimination, and whether such effects varied across racial/ethnic group. Some research suggests that there are large race-of-interviewer effects for self-reports of discrimination by Blacks (Krysan and Couper, 2003). Other methodologies exist that might improve measures of people's exposure to racial discrimination and its health impact, such as use of the functional MRI (Mays et al., 2007) or the implicit association test (Carney et al., 2010; Krieger et al., 2010). These limitations point to areas for further research.

Lastly, the findings presented in this paper focus only on questions pertaining to everyday discrimination. We are mindful that everyday discrimination does not capture all aspects of racial/ethnic discrimination even at the individual level, much less at other levels (e.g., occupational segregation, residential segregation, etc.). However, our finding that explicitly attributing the discrimination to race/ethnicity within a single question does matter when asking about discrimination experiences is widely applicable to self-report measures that assess other dimensions of discrimination, including everyday discrimination.

CONCLUSION

Our study, based on a population-based sample including a wide range of racial/ethnic groups, provides important findings relevant to research on racial/ethnic discrimination and health.

First, we demonstrate that questions that explicitly ask about recent everyday racial/ethnic discrimination are different from questions that ask about recent everyday unfair treatment. While this latter finding is not surprising, it is important because results of studies using these two different approaches are often treated as equivalent (Albert et al., 2010; Lewis et al., 2010; Taylor et al., 2007; Tomfohr

et al., 2010). Our study empirically proves that this practice is not scientifically sound and confirms findings from researchers who have inferred this point from both reviewing the literature and also using different approaches to compare results based on one-stage versus two-stage questions (Bastos et al., 2010; Brown 2001; Chae et al., 2008; Krieger et al., 2005; Williams and Mohammed, 2009). We also found important differences in results that are computed from the two different approaches we studied. We found that frequencies are lower when racial/ethnic attribution is culled—in a second stage—from questions that initially ask about any type of unfair treatment than they are when racial/ethnic discrimination is detected by questions that ask directly about experiences of racial/ethnic discrimination in a single stage. These results call for greater rigor in comparing existing studies because the two approaches are not conceptually comparable or empirically equivalent. Underscoring this point, we documented that the two approaches yield different results even after adjusting the estimates for the two-stage approach by specifying attribution to race/ethnicity to ensure that we are measuring only racial/ethnic discrimination. Our results also suggest that more studies are needed to ascertain what constitutes accurate measurement of racial/ethnic discrimination. Second, and this is specific to the measure we tested, we found that both approaches have acceptable psychometric properties for use in diverse racial/ethnic groups, and also that including an item pertaining to discrimination based on language is appropriate.

Finally, our study reinforces that DIF is a useful tool for evaluating instruments across racial/ethnic groups. Our analyses, however, suggest that researchers should hypothesize a priori where they anticipate DIF and remind researchers that the detection of DIF does not necessarily indicate measurement bias. In some circumstances, small DIF is warranted. Hence, the presence of small DIF may be a way to validate assumptions about items, but this is only possible when assumptions are made explicit and a priori.

Our findings suggest caution to researchers when comparing studies that have used different approaches to measure racial/ethnic discrimination, and provide useful empirical guidelines for measuring and analyzing racial/ethnic discrimination. Of equal importance, we have developed a self-reported measure of racial/ethnic discrimination that functions well in a range of different racial/ethnic groups and makes possible comparing how racial/ethnic discrimination is associated with health disparities among multiple racial/ethnic groups.

Corresponding author: Salma Shariff-Marco, Cancer Prevention Institute of California, 2201 Walnut Avenue, Suite 300, Fremont, CA 94538. E-mail: salma.shariffmarco@cpic.org.

NOTES

2. The first two authors shared in the leadership of the project and bore equal responsibility in drafting this manuscript.
3. We will use the terms Black and African American interchangeably in this paper because, following the 1997 federal directive on reporting race and ethnicity (Office of Management and Budget 1997), the option provided for self-identification in the survey was Black or African American, and also because the literature we cite uses both terms.
4. The sample consisted of 33,410 Whites, 9077 Latinos, 4441 Asian Americans/Native Hawaiians/Pacific Islanders (AANHPIs), 2410 African Americans, 423 American Indians/Alaska Natives (AI/ANs) and 1287 respondents identifying as Multiracial.
5. The study was approved by the Institutional Review Board at UCLA and California State and, thus, exempted from IRB review at the National Institutes of Health (NIH).
6. The 435 respondents excluded from our study were similar to the final study sample for all the sociodemographic characteristics reported in Table 1, except for gender. There was a higher proportion of males among the excluded respondents.

REFERENCES

Albert, Michelle A., Yvette Cozier, Paul M. Ridker, Julie R. Palmer, Robert J. Glynn, Lynda Rose, Nitsan Halevy, and Lynn Rosenberg (2010). Perceptions of Race/Ethnic Discrimination in Relation to Mortality Among Black Women: Results From the Black Women's Health Study. *Archives of Internal Medicine*, 170(10): 896–904.

Barnes, Lisa L., Carlos F. Mendes de Leon, Tene T. Lewis, Julia L. Bienias, Robert S. Wilson, and Denis A. Evans (2008). Perceived Discrimination and Mortality in a Population-Based Study of Older Adults. *American Journal of Public Health*, 98(7): 1241–1247.

Bastos, Joao Luiz, Roger Keller Celeste, Eduardo Faerstein, and Aluisio J. Barros (2010). Racial Discrimination and Health: a Systematic Review of Scales With a Focus on Their Psychometric Properties. *Social Science & Medicine*, 70(7): 1091–1099.

Bobo, Lawrence (2001). Racial Attitudes and Relations at the Close of the 20th Century. In Neil J. Smelser, William J. Wilson, and Faith Mitchell (Eds.), *America Becoming: Racial Trends and Their Consequences, Volume 1*, pp. 264–301. Washington, DC: National Academy Press.

Brown, Tony N. (2001). Measuring Self-Perceived Racial and Ethnic Discrimination in Social Surveys. *Sociological Spectrum*, 21(3): 377–392.

California Health Interview Survey (2009). *CHIS 2007 Methodology Series Report 4: Response Rates*. Los Angeles, CA: UCLA Center for Health Policy Research.

Cameron, Christopher David Ruiz (1997). How the Garcia Cousins Lost Their Accents: Understanding the Language of Title VII Decisions Approving English-Only Rules As the Product of Racial Dualism, Latino Invisibility, and Legal Indeterminacy. *California Law Review*, 85: 1347–1393.

Carney, Dana R., Mahzarin R. Banaji, and Nancy Krieger (2010). Implicit Measures Reveal Evidence of Personal Discrimination. *Self and Identity*, 9(2): 162–176.

Chae, David H., David T. Takeuchi, Elizabeth M. Barbeau, Gary G. Bennett, Jane Lindsey, and Nancy Krieger (2008). Unfair Treatment, Racial/Ethnic Discrimination, Ethnic Identification, and Smoking Among Asian Americans in the National Latino and Asian American Study. *American Journal of Public Health*, 98(3): 485–492.

Clark, Rodney, Norman B. Anderson, Vernessa R. Clark, and David R. Williams (1999). Racism As a Stressor for African Americans. A Biopsychosocial Model. *American Psychologist*, 54(10): 805–816.

Desmond, Matthew and Mustafa Emirbayer (2009). What Is Racial Domination? *Du Bois Review: Social Science Research on Race*, 6(2): 335–355.

Epel, Elissa S., Elizabeth H. Blackburn, Jue Lin, Firdaus S. Dhabhar, Nancy E. Adler, Jason D. Morrow, and Richard M. Cawthon (2004). Accelerated Telomere Shortening in Response to Life Stress. *Proceedings of the National Academy of Sciences of the United States of America*, 101(49): 17312–17315.

Essed, Philomena (1990). *Understanding Everyday Racism: an Interdisciplinary Theory*. Newbury Park, CA: Sage Publications, Inc.

Feagin, Joe R. (2006). *Systematic Racism*. New York: CRC Press.

Feagin, Joe R. and Eileen O'Brien (2003). *White Men on Race: Power, Privilege, and the Shaping of Cultural Consciousness*. Boston, MA: Beacon Press.

Gee, Gilbert C., Annie Ro, Salma Shariff-Marco, and David Chae (2009). Racial Discrimination and Health Among Asian Americans: Evidence, Assessment, and Directions for Future Research. *Epidemiologic Reviews*, 31: 130–151.

Geronimus, Arline T. (2001). Understanding and Eliminating Racial Inequalities in Women's Health in the United States: the Role of the Weathering Conceptual Framework. *Journal of the American Medical Women's Association* 56(4): 133–136, 149–150.

Geronimus, Arline T., Margaret Hicken, Danya Keene, and John Bound (2006). "Weathering" and Age Patterns of Allostatic Load Scores Among Blacks and Whites in the United States. *American Journal of Public Health*, 96(5): 826–833.

Judkins, David, Tom Krenzke, Andrea Piesse, Zizhong Fan, and Wen-Chau Haung (2007). Preservation of Skip Patterns and Covariate Structure Through Semi-Parametric Whole Questionnaire Imputation. *Proceedings of the Joint Statistical Meetings of the American Statistical Association*, pp. 3211–3288.

Judkins, David, Andrea Piesse, and Tom Krenzke (2008). Multiple Semi-Parametric Imputation. *Proceedings of the Joint Statistical Meetings of the American Statistical Association*, pp. 48–58.

Kessler, Ronald C., Kristin D. Mickelson, and David R. Williams (1999). The Prevalence, Distribution, and Mental Health Correlates of Perceived Discrimination in the United States. *Journal of Health and Social Behavior*, 40(3): 208–230.

Kluegel, James R. and Lawrence D. Bobo (2001). Perceived Group Discrimination and Policy Attitudes: the Sources and Consequences of the Race and Gender Gaps. In Alice O'Connor, Chris Tilly, and Lawrence D. Bobo (Eds.), *Urban Inequality: Evidence From Four Cities*, pp. 163–214. New York: Russell Sage Foundation.

Kressin, Nancy R., Kristal L. Raymond, and Meredith Manze (2008). Perceptions of Race/Ethnicity-Based Discrimination: a Review of Measures and Evaluation of Their Usefulness for the Health Care Setting. *Journal of Health Care for the Poor and Underserved*, 19(3): 697–730.

Krieger, Nancy (2000). Discrimination and Health. In Lisa F. Berkman and Ichiro Kawachi (Eds.), *Social Epidemiology*, pp. 36–75. New York: Oxford University Press.

Krieger, Nancy (2010). The Science and Epidemiology of Racism and Health: Racial/Ethnic Categories, Biological Expressions of Racism, and the Embodiment of Inequality—an Ecosocial Perspective. In Ian Whitmarsh and David S. Jones (Eds.), *What's the Use of Race? Modern Governance and the Biology of Difference*, pp. 225–255. Cambridge, MA: The MIT Press.

Krieger, Nancy, Dana Carney, Katie Lancaster, Pamela D. Waterman, Anna Kosheleva, and Mahzarin Banaji (2010). Combining Explicit and Implicit Measures of Racial Discrimination in Health Research. *American Journal of Public Health*, 100(8): 1485–1492.

Krieger, Nancy, and Stephen Sidney (1996). Racial Discrimination and Blood Pressure: the CARDIA Study of Young Black and White Adults. *American Journal of Public Health*, 86(10): 1370–1378.

Krieger, Nancy, Kevin Smith, Deepa Naishadham, Cathy Hartman, and Elizabeth M. Barbeau (2005). Experiences of Discrimination: Validity and Reliability of a Self-Report Measure for Population Health Research on Racism and Health. *Social Science & Medicine*, 61(7): 1576–1596.

Krysan, Maria and Mick P. Couper (2003). Race in the Live and the Virtual Interview: Racial Deference, Social Desirability, and Activation Effects in Attitude Surveys. *Social Psychology Quarterly*, 66(4): 364–383.

Landrine, Hope, Elizabeth A. Klonoff, Irma Corral, Senaida Fernandez, and Scott Roesch (2006). Conceptualizing and Measuring Ethnic Discrimination in Health Research. *Journal of Behavioral Medicine*, 29(1): 79–94.

Lee, Sunghee, Delight E. Satter, and Ninez A. Ponce (2009). Effect of Race and Ethnicity Classification on Survey Estimates: Anomaly of the Weighted Totals of American Indians and Alaska Natives. *American Indian and Alaska Native Mental Health Research*, 16(3): 1–15.

Lewis, Tene T., Allison E. Aiello, Sue Leurgans, Jeremiah Kelly, and Lisa L. Barnes (2010). Self-Reported Experiences of Everyday Discrimination Are Associated With Elevated C-Reactive Protein Levels in Older African-American Adults. *Brain, Behavior, and Immunity*, 24(3): 438–443.

Lewis, Tene T., Susan A. Everson-Rose, Lynda H. Powell, Karen A. Matthews, Charlotte Brown, Kelly Karavolos, Kim Sutton-Tyrrell, Elizabeth Jacobs, and Deidre Wesley (2006). Chronic Exposure to Everyday Discrimination and Coronary Artery Calcification in African-American Women: the SWAN Heart Study. *Psychosomatic Medicine*, 68(3): 362–368.

Liang, Christopher T., Lisa C. Li, and Bryan S. Kim (2004). The Asian American Racism-Related Stress Inventory: Development, Factor Analysis, Reliability, and Validity. *Journal of Counseling Psychology*, 51(1): 103–114.

Massey, Douglas S. and Garvey Lundy (2001). Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings. *Urban Affairs Review*, 36(4): 452–469.

Mays, Vickie M., Susan D. Cochran, and Namdi W. Barnes (2007). Race, Race-Based Discrimination, and Health Outcomes Among African Americans. *Annual Review of Psychology*, 58: 201–225.

Mays, Vickie M., Ninez A. Ponce, Donna L. Washington, and Susan D. Cochran (2003). Classification of Race and Ethnicity: Implications for Public Health. *Annual Review of Public Health*, 24: 83–110.

Muthen, Linda K. and Bengt O. Muthen (2007). *Mplus (Version 4.21) Software*. Los Angeles, CA: Muthen & Muthen.

National Research Council (2004). *Measuring Racial Discrimination. Panel on Methods for Assessing Discrimination*. Washington, DC: The National Academies Press.

Nunnally, Jum C. and Ira H. Bernstein (1994). *Psychometric Theory*, 3ed. New York: McGraw Hill Co.

Office of Management and Budget (1997). Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, Office of Management and Budget Directive 15. *U.S. Census Bureau*. ⟨http://www.census.gov/population/www/socdemo/race/Ombdir15.html⟩ (accessed October 3, 2010).

Paradies, Yin (2006). A Systematic Review of Empirical Research on Self-Reported Racism and Health. *International Journal of Epidemiology*, 35(4): 888–901.

Pascoe, Elizabeth A. and Laura Smart Richman (2009). Perceived Discrimination and Health: a Meta-Analytic Review. *Psychological Bulletin*, 135(4): 531–554.

Piesse, Andrea, David R. Judkins, and Zizhong Fan (2005). Item Imputation Made Easy. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 2540–2546.

Pincus, Fred L. (2003). *Reverse Discrimination: Dismantling the Myth*. Boulder, CO: Lynne Rienner Publishers.

Reeve, Bryce B. and Peter Fayers (2005). Applying Item Response Theory Modelling for Evaluating Questionnaire Item and Scale Properties. In Peter Fayers and Ron Hays (Eds.), *Assessing Quality of Life in Clinical Trials: Methods and Practice*, 2ed., pp. 55–73. New York: Oxford University Press.

Reeve, Bryce B., Ron D. Hays, Jakob B. Bjorner, Karon F. Cook, Paul K. Crane, Jeanne A. Teresi, David Thissen, Dennis A. Revicki, David J. Weiss, Ronald K. Hambleton, Honghu Liu, Richard Gershon, Steven P. Reise, Jin-shei Lai, and David Cella (2007). Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*, 45(5 Suppl 1): S22–S31.

Reeve, Bryce B., Gordon Willis, Salma N. Shariff-Marco, Nancy Breen, David R. Williams, Gilbert C. Gee, Margarita Alegria, David T. Takeuchi, Martha S. Kudela, and Kerry Y. Levin (Forthcoming). Comparing Cognitive Interviewing and Psychometric Methods to Evaluate a Racial/Ethnic Discrimination Scale. *Field Methods*.

Research Triangle Institute (2008). *SUDAAN (Release 10.0)*. Research Triangle Park, NC: Research Triangle Institute.

Roussos, Louis and William Stout (1996). A Multidimensionality-Based DIF Analysis Paradigm. *Applied Psychological Measurement*, 20(4): 355–371.

Seeman, Teresa E., Burton H. Singer, Carol D. Ryff, Gayle Dienberg Love, and Lene Levy-Storms (2002). Social Relationships, Gender, and Allostatic Load Across Two Age Cohorts. *Psychosomatic Medicine*, 64(3): 395–406.

Shariff-Marco, Salma, Gilbert C. Gee, Nancy Breen, Gordon Willis, Bryce B. Reeve, David Grant, Ninez A. Ponce, Nancy Krieger, Hope Landrine, David R. Williams, Margarita Alegria, Vickie M. Mays, Timothy P. Johnson, and E. Richard Brown (2009). A Mixed-Methods Approach to Developing a Self-Reported Racial/Ethnic Discrimination Measure for Use in Multiethnic Health Surveys. *Ethnicity & Disease*, 19(4): 447–453.

Spencer, Michael S. and Juan Chen (2004). Effect of Discrimination on Mental Health Service Utilization Among Chinese Americans. *American Journal of Public Health*, 94(5): 809–814.

Swan, Judith, Nancy Breen, Linda Burhansstipanov, Delight E. Satter, William W. Davis, Timothy McNeel, and C. Matthew Snipp (2006). Cancer Screening and Risk Factor Rates Among American Indians. *American Journal of Public Health*, 96(2): 340–350.

Taylor, Teletia R., Carla D. Williams, Kepher H. Makambi, Charles Mouton, Jules P. Harrell, Yvette Cozier, Julie R. Palmer, Lynn Rosenberg, and Lucile L. Adams-Campbell (2007). Racial Discrimination and Breast Cancer Incidence in US Black Women: the Black Women's Health Study. *American Journal of Epidemiology*, 166(1): 46–54.

Teresi, Jeanne A. and John A. Fleishman (2007). Differential Item Functioning and Health Assessment. *Quality of Life Research*, 16(Suppl 1): 33–42.

Thissen, David L. (2001). *IRTLRDIF (v. 2.0b): Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio Tests for Differential Item Functioning*. Chapel Hill, NC: Thurstone Psychometric Laboratory.

Thissen, David, Lynne Steinberg, and Howard Wainer (1993). Detection of Differential Item Functioning Using the Parameters of Item Response Models. In Paul W. Holland and Howard Wainer (Eds.), *Differential Item Functioning*, pp. 67–113. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tomfohr, Lianne, Denise C. Cooper, Paul J. Mills, Richard A. Nelesen, and Joel E. Dimsdale (2010). Everyday Discrimination and Nocturnal Blood Pressure Dipping in Black and White Americans. *Psychosomatic Medicine*, 72(3): 266–272.

Tuan, Mia (1998). *Forever Foreigners or Honorary Whites? The Asian Ethnic Experience Today*. New Brunswick, NJ: Rutgers University Press.

Turner, Margery A. and Felicity Skidmore (Eds.) (1999). *Mortgage Lending Discrimination: A Review of Existing Evidence*. Washington, DC: The Urban Institute.

Wildsmith, Elizabeth M. (2002). Testing the Weathering Hypothesis Among Mexican-Origin Women. *Ethnicity & Disease*, 12(4): 470–479.

Williams, David R. and Selina A. Mohammed (2009). Discrimination and Racial Disparities in Health: Evidence and Needed Research. *Journal of Behavioral Medicine*, 32(1): 20–47.

Williams, David R., Selina A. Mohammed, Jacinta Leavell, and Chiquita Collins (2010). Race, Socioeconomic Status, and Health: Complexities, Ongoing Challenges, and Research Opportunities. *Annals of the New York Academy of Sciences*, 1186: 69–101.

Williams, David R. and Harold Neighbors (2001). Racism, Discrimination and Hypertension: Evidence and Needed Research. *Ethnicity & Disease*, 11(4): 800–816.

Williams, David R., Yan Yu, James S. Jackson, and Norman B. Anderson (1997). Racial Differences in Physical and Mental Health: Socio-Economic Status, Stress and Discrimination. *Journal of Health Psychology*, 2(3): 335–351.

Yoo, Hyung Chol, Gilbert C. Gee, Craig K. Lowthrop, and Joanne Robertson (2010). Self-Reported Racial Discrimination and Substance Use Among Asian Americans in Arizona. *Journal of Immigrant and Minority Health*, 12(5): 683–690.

Yoo, Hyung Chol, Gilbert C. Gee, and David Takeuchi (2009). Discrimination and Health Among Asian American Immigrants: Disentangling Racial From Language Discrimination. *Social Science & Medicine*, 68(4): 726–732.