Practice of Epidemiology

Racial/Ethnic Differences in Responses to the Everyday Discrimination Scale: A Differential Item Functioning Analysis

Tené T. Lewis*, Frances M. Yang, Elizabeth A. Jacobs, and George Fitchett

* Correspondence to Dr. Tené T. Lewis, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, 4th Floor, New Haven, CT 06520 (e-mail: tene.lewis@yale.edu).

The authors examined the impact of race/ethnicity on responses to the Everyday Discrimination Scale, one of the most widely used discrimination scales in epidemiologic and public health research. Participants were 3,295 middle-aged US women (African-American, Caucasian, Chinese, Hispanic, and Japanese) from the Study of Women's Health Across the Nation (SWAN) baseline examination (1996–1997). Multiple-indicator, multiple-cause models were used to examine differential item functioning (DIF) on the Everyday Discrimination Scale by race/ethnicity. After adjustment for age, education, and language of interview, meaningful DIF was observed for 3 (out of 10) items: "receiving poorer service in restaurants or stores," "being treated as if you are dishonest," and "being treated with less courtesy than other people" (all P's < 0.001). Consequently, the "profile" of everyday discrimination differed slightly for women of different racial/ethnic groups, with certain "public" experiences appearing to have more salience for African-American and Chinese women and "dishonesty" having more salience for racial/ethnic minority women overall. "Courtesy" appeared to have more salience for Hispanic women only in comparison with African-American women. Findings suggest that the Everyday Discrimination Scale could potentially be used across racial/ethnic groups as originally intended. However, researchers should use caution with items that demonstrated DIF.

African Americans; Asian Americans; bias (epidemiology); European continental ancestry group; Hispanic Americans; prejudice; psychometrics; questionnaires

Abbreviations: CFI, comparative fit index; df, degrees of freedom; DIF, differential item functioning; EDS, Everyday Discrimination Scale; MIMIC, multiple-indicator, multiple-cause; RMSEA, root mean square error of approximation; SWAN, Study of Women's Health Across the Nation.

A growing body of research has linked self-reported experiences of discrimination to negative physical and mental health outcomes (1–12). Findings have been particularly pronounced among African-American populations (5, 13–15) but have been observed among Hispanics, Asian Americans, and Caucasians as well (9, 16–19). A major limitation of this research has been the overreliance on scales designed for use with African-American populations and a dearth of measurement instruments designed to assess discriminatory experiences across other populations (20). Because African Americans, Asian Americans, Hispanics, and Caucasians have different histories of discrimination and oppression in the United States, it is possible that the

experience of discrimination may differ for these groups as well.

One of the most widely used scales in epidemiologic and public health research is the Everyday Discrimination Scale (EDS) (6, 21). Based on the construct defined by Essed (22, 23), "everyday" experiences of discrimination are defined as "a range of events, many of which appear to be 'trivial' or even 'normal'. . . Certain rights, respect, and recognition, which whites take for granted in their own lives, are denied to people of color" (22, pp. 258–259). The EDS was designed to assess discriminatory treatment across a variety of domains; thus, items on the scale are framed in the context of general mistreatment, without reference

to race, ethnicity, gender, or other demographic/personal characteristics (21). Because the items themselves are framed rather generically, the scale has been used to measure discriminatory experiences for persons from a variety of racial/ethnic backgrounds (including Caucasians) (9, 16, 17, 24–28).

It is important to note that although the scale is widely used across racial/ethnic groups, items on the EDS were primarily based on qualitative data from interviews with African-American women in the United States and black women in the Netherlands (22, 23). Thus, the extent to which these experiences are equally relevant for persons of other racial/ethnic backgrounds is unclear. In some instances, investigators using this scale across racial/ethnic groups report one set of findings for one group (e.g., African Americans) and a different set of findings for another group (e.g., Caucasians) (5, 13–15, 17, 29). Hence, despite the potential universality of items on the EDS, it is possible that the items actually function differently for different racial/ethnic groups. To date, however, few studies have investigated this issue.

The current study was designed to examine differences in item functioning on the EDS by race/ethnicity in a sample of over 3,000 African-American, Hispanic, Chinese, Japanese, and Caucasian women from the Study of Women's Health Across the Nation (SWAN) cohort. In general, differential item functioning (DIF) can be conceptualized as a form of measurement bias, where individuals respond to items on a scale as a function of some attribute other than what the scale is designed to measure (30). In the current analysis, investigating differences in item functioning allows us to determine whether there is measurement bias in the EDS by race/ethnicity. Such bias could result in a form of exposure misclassification, where women of different racial/ethnic backgrounds are classified as higher (or lower) on everyday discrimination than they actually are. This misclassification could subsequently result in over- or underestimates of discrimination and health associations in epidemiologic research and ultimately impact the inferences that are drawn from studies of discrimination and health in a multiracial/ethnic context.

We used multiple-indicator, multiple-cause (MIMIC) models to examine DIF on the EDS by race/ethnicity in the SWAN cohort. Although there are other analytic approaches for examining DIF (30–34), these approaches do not allow for the statistical control of other variables. MIMIC models allow for the statistical control of potential demographic confounders such as age and education, which are known correlates of reports of discrimination (11, 35–38).

We did not have a priori hypotheses about specific racial/ethnic differences at the individual item level. However, given the origins of the scale, we sought to determine more broadly whether there were significant differences in item functioning for African-American women in particular compared with women of other racial/ethnic backgrounds or for ethnic minority women in general (African-American, Hispanic, Chinese, and Japanese) compared with Caucasian women. We were particularly interested in determining whether the "profile" of everyday discrimination (i.e., the set of experiences that comprise everyday discrimination) differed for women of different racial/ethnic groups.

MATERIALS AND METHODS

Participants

Participants were 3,302 women from SWAN, a multisite, multiethnic study of the natural history of the menopausal transition. Details of the SWAN study design have been previously published (39). Briefly, SWAN includes 7 community sites; at each site, investigators recruited Caucasian women and women from 1 other racial/ethnic minority group. African-American women were recruited in Boston, Massachusetts; Chicago, Illinois; Detroit, Michigan; and Pittsburgh, Pennsylvania. Hispanic women were recruited in Newark, New Jersey. Chinese women were recruited in Oakland, California, and Japanese women were recruited in Los Angeles, California.

The current analyses utilized data from the SWAN baseline examination (1996–1997). Women were eligible for SWAN if they were aged 42–52 years and self-identified as a member of one of the targeted racial/ethnic groups. Additional criteria included having an intact uterus and at least 1 ovary and reporting having had a menstrual period in the preceding 3 months. Women who were pregnant, were breastfeeding, or reported using exogenous hormones in the 3 months preceding the baseline examination were ineligible.

Study procedures were approved by the institutional review board at each site, and all women provided informed consent.

Measures

Race/ethnicity. Race/ethnicity was self-reported as African-American (referent), Caucasian, Hispanic, Chinese, or Japanese.

Discrimination. Discrimination was assessed with the Detroit Area Study EDS (21). This scale asked participants to indicate how often they had experienced various forms of day-to-day mistreatment over the previous 12 months. Examples include "You are treated with less respect than other people," "You receive poorer service than other people at restaurants or stores," and "People act as if they think you are not smart." Although the most commonly used version of the EDS includes only 9 items, the version used in SWAN has an additional, 10th item: "People ignore you or act as if you aren't there." This item was retained in SWAN because it is consistent with experiences included in the original work on everyday discrimination by Essed (22, 23) and is similar to items on other scales assessing day-to-day discriminatory treatment (40, 41). Each of the 10 types of discrimination was assessed with a 4-point scale (1 = never, 2 = rarely, 3 = sometimes, 4 = often). Scores were summed and averaged, resulting in a possible overall score of 1–4. The EDS has demonstrated good internal consistency (5, 37, 38, 42–44), stability over time (5), and convergent and divergent validity (44, 45) in prior studies.

Covariates. Covariates were age (years), interview language (non-English vs. English), and education (years). Education was used as a marker of socioeconomic status rather than income, because of potential regional differences in the "buying power" of a given income.

Analyses

Descriptive analyses. Descriptive statistics were used to characterize the sample in terms of race/ethnicity, age, education, language of interview, and level of everyday discrimination. Analysis of variance and chi-squared tests were conducted to test for racial/ethnic differences in sample characteristics.

We conducted a series of exploratory factor analyses to establish the dimensionality of the EDS in order to determine whether the items were represented by a single, dominant factor or several small factors. Unidimensionality was considered present if the eigenvalue of the first factor was more than 3.5 times the size of the eigenvalue of the second and subsequent factors (46). We also conducted confirmatory factor analyses to confirm the dimensionality of the models found in the exploratory factor analysis. Model modification indices (47) were used to improve the fit by modeling the largest correlations between the factor indicators each time, until no additional improvements to the fit of the model were identified.

DIF analyses. DIF analyses were used to test whether items on the EDS functioned differently for African-American women compared with women of other racial/ethnic groups. DIF has traditionally been used in educational research to identify poorly functioning test items (48, 49). It has also been used in psychological research to identify particular ''profiles'' on a construct of interest by sociodemographic characteristics (i.e., differences in depressive symptoms by gender). In keeping with the latter usage, DIF analyses were used in the current study to determine whether the profile of everyday discrimination differed by race/ethnicity.

MIMIC models were used to examine DIF on the EDS by race/ethnicity. In DIF analyses, MIMIC models function as a special case of structural equation modeling, in which a latent construct (i.e., the actual, unobserved amount of everyday discrimination a given participant experiences) intervenes between an observed background variable (race/ethnicity) and a set of observed response variables (responses to items on the EDS). When there is no DIF, any association between the observed background variable (race/ethnicity) and the observed response variables (responses to everyday discrimination items) is primarily attributable to the association between the background variable and the latent construct. In other words, if there is no DIF, race/ethnicity would be associated with exposure to everyday discrimination (the latent construct), and these actual experiences of everyday discrimination would, in turn, determine responses to items on the EDS. If race/ethnicity is in fact associated with responses to items on the EDS *over and above* the amount of actual exposure to everyday discrimination, DIF is present.

The basic MIMIC model constructed for this study is detailed in Figure 1. The right side of the model displays the relation between the latent construct (everyday discrimination) and the individual items on the EDS that are ''indicators'' of the latent construct (pathway ''a,'' as one example in Figure 1). This is the measurement model, equivalent to a factor analysis. The left side of the model (pathway ''b'') is the structural model, that is, regression estimates of the relation between race/ethnicity (the background
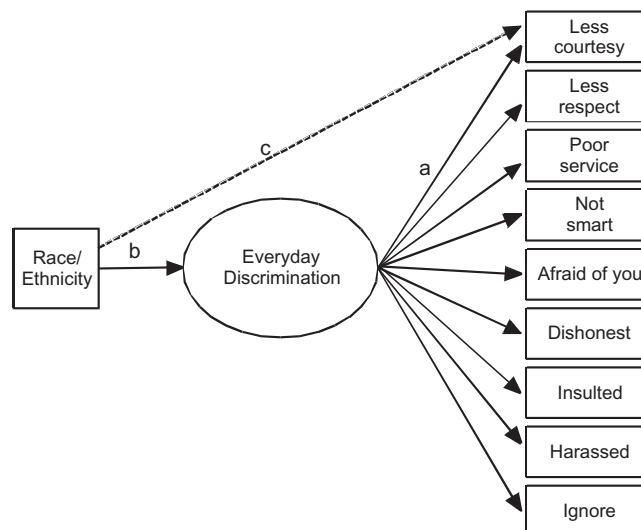
Figure 1. Multiple-indicator, multiple-cause (MIMIC) model for everyday discrimination, Study of Women's Health Across the Nation, 1996–1997. Pathway ''a'' is an individual item loading on the everyday discrimination latent construct (factor) from a factor analysis; pathway ''b'' is the regression estimate of the association between race/ethnicity and the latent construct of everyday discrimination; and pathway ''c'' is the regression estimate of a direct effect (differential item functioning) of race/ethnicity on an individual item on the Everyday Discrimination Scale.

variable) and everyday discrimination (the latent construct). The dashed line (pathway ''c'') indicates the possible presence of a direct effect, or DIF, which is conceptually equivalent to a regression estimate of the direct relation between race/ethnicity (background variable) and an individual item on the EDS that is not mediated through the everyday discrimination latent construct.

In building the MIMIC models, we began with a no-DIF base model (no direct effects of race/ethnicity or other covariates on individual items). Using modification indices, we iteratively identified DIF effects that would significantly improve model fit (50, 51). The initial models examined the effects of race/ethnicity only on the latent everyday discrimination construct and the individual items on the EDS. A second set of models added adjustments for age and education, and final models added an additional control for interview language. Sensitivity analyses were also conducted, limiting the sample to English-speaking women only. Results were comparable; thus, we retained the full sample for all analyses. All models utilized the weighted least-squares mean- and variance-adjusted estimator, which implements a multivariate probit model for the DIFFTEST procedure.

Final analyses were then rerun to express DIF effects as odds ratios (exponentiated regression coefficients from logistic regression models), which may have more utility in medical and public health research. Because DIF is sensitive to sample size, we used the Cole et al. (52) criteria for ''meaningful'' DIF: statistically significant odds ratios greater than 2.0 or less than 0.5. These analyses used a multivariate logit parameterization with robust maximum likelihood methods.

Table 1.   Characteristics of Participants (Mean or Percentage) by Race/Ethnicity, Study of Women's Health Across the Nation, 1996–1997

	Race/Ethnicity					P Value
	African-American (n = 931)	Caucasian (n = 1,547)	Chinese (n = 250)	Hispanic (n = 286)	Japanese (n = 281)	
Age, years	45.8 (2.7)[a]	45.8 (2.7)	46.0 (2.6)	45.9 (2.8)	46.2 (2.7)	0.23
Education, years	14.6 (2.3)	15.6 (2.1)	14.5 (3.2)	10.9 (3.7)	15.1 (1.8)	<0.0001
% interviewed in English	100	99.7	56.4	32.5	60	<0.0001
Everyday discrimination score[b]	1.92 (0.53)	1.69 (0.43)	1.87 (0.47)	1.24 (0.38)	1.59 (0.48)	<0.0001

[a] Numbers in parentheses, standard deviation.

[b] Possible range: 1 to 4; higher scores indicate higher reports of everyday discrimination.

All models were assessed with the root mean square error of approximation (RMSEA) (53, 54) and the comparative fit index (CFI) (55, 56). The RMSEA provides a measure of discrepancy per model degree of freedom and approaches zero as fit improves. Browne and Cudek (53) recommended rejecting models with RMSEA values greater than 0.1. The CFI ranges between 0 and 1; values greater than 0.95 generally indicate adequate fit (57, 58). All statistical analyses were conducted using SPSS, version 18 (SPSS Inc., Chicago, Illinois), and Mplus, version 6.11 (Los Angeles, California) (59).

RESULTS

Complete data were available for 3,295 women, of whom 931 (28.2%) were African-American, 1,547 (46.8%) were Caucasian, 250 (7.6%) were Chinese, 286 (8.5%) were Hispanic, and 281 (8.5%) were Japanese. Table 1 presents the distribution of study variables by race/ethnicity. On average, Caucasian women were the most educated (≥ 16 years), while Hispanic women were the least educated (11 years). All other groups reported, on average, 14.5–15.1 years of education. Almost all of the Caucasian and African-American women were interviewed in English, and more than half of the Chinese and Japanese women were interviewed in English, while only one-third of Hispanic women were interviewed in English. As previously reported in SWAN, there were significant racial/ethnic differences in reports of everyday discrimination (26), with African-American and Chinese women reporting the highest levels of everyday discrimination and Hispanic women reporting the lowest levels. Response frequencies, mean values, and standard deviations for each of the 10 items on the EDS for the full sample are presented in Appendix Table 1, and polychoric correlations for the full sample are presented in Appendix Table 2.

Factor analysis

Although a prior study of a small subset of SWAN women ($n = 363$) identified a 2-factor solution (29), exploratory factor analysis of the 10 items on the EDS in the full SWAN cohort revealed that a single-factor solution was a reasonable fit to the data, as all items had large and positive loadings on one factor, with the first eigenvalue being over 3.5 times the second eigenvalue (46). The first eigenvalue was 6.04, the second eigenvalue was 0.96, the next eigenvalue was 0.72,

and subsequent eigenvalues ranged from 0.14 (for the 10th eigenvalue) to 0.51 (for the fourth eigenvalue). The largest drop in eigenvalue was between the first and second values. The goodness-of-fit statistics were: CFI = 0.95, RMSEA = 0.14, $\chi^2 = 2,252$, and degrees of freedom (df) = 35 ($P < 0.001$).

The CFI supported a single-factor, unidimensional model; however, because our RMSEA of 0.14 was slightly higher than 0.10, we also fitted both 2- and 3-factor models using exploratory factor analysis. The 2-factor model had a slightly better fit (fit statistics: CFI = 0.98, RMSEA = 0.09, $\chi^2 = 745$, and df = 26 ($P < 0.001$)), but the fit statistics for the 3-factor model were not considerably better than those for the 2-factor model (CFI = 1.00, RMSEA = 0.06, $\chi^2 = 253$, and df = 18 ($P < 0.001$)). Additionally, there was a Heywood case for the "people act as if you are dishonest" item under the second factor in the 3-factor model, suggesting that there were too many factors estimated (60). Because the case for unidimensionality was not clearly established following the exploratory factor analyses, the 1- and 2-factor models were further fitted using confirmatory factor analysis.

Results from the confirmatory factor analysis models revealed an excellent fit for the single-factor solution with correlated factor indicators (CFI = 0.996, RMSEA = 0.05, $\chi^2 = 190$, and df = 24 ($P < 0.001$)). The 2-factor model improved (from CFI = 0.966, RMSEA = 0.118, $\chi^2 = 1,597$, and df = 34 ($P < 0.001$) to CFI = 0.995, RMSEA = 0.05, $\chi^2 = 263$, and df = 29 ($P < 0.001$)) when correlated factor indicators were included (additional data available upon request). The fit indices for confirmatory factor analysis models indicated that the 1-factor model was slightly better than the 2-factor model with the correlated factor indicators. Additionally, the first and second factors in the 2-factor model were highly correlated at 0.89, suggesting that 2 separate factors might not be empirically meaningful. Given this, the size of the first factor relative to the second, and findings from prior studies supporting the unidimensionality of the scale (37, 45), we retained the single-factor solution.

DIF analyses

Results from the MIMIC models are presented in Table 2. The structural component of the model presents the standardized regression estimates, while the measurement component of the model presents the measurement slopes (i.e., standardized factor loadings). Initial estimates from MIMIC

Table 2.   Standardized Parameter Estimates[a] From Multiple-Indicator, Multiple-Cause Models for Differences in the Latent "Everyday Discrimination" Construct by Race/Ethnicity, With and Without Adjustment for Direct Effects (Differential Item Functioning) and Covariates, Study of Women's Health Across the Nation, 1996–1997

	No Adjustment for Direct Effects		Adjustment for Direct Effects	
	Model 1a[b]	Model 1b[c]	Model 2a[b]	Model 2b[c]
Structural component				
Race/ethnicity (1 vs. 0)				
Caucasian	−0.23***	−0.22***	−0.14***	−0.18***
Chinese	−0.03	0.04*	−0.03	0.05*
Hispanic	−0.46***	−0.36***	−0.45***	−0.37***
Japanese	−0.18**	−0.11***	−0.14*	−0.07***
African-American (referent)				
Covariates				
Age (per year)		−0.04***		−0.05***
Education (per year)		−0.04*		−0.09***
Language[d]		−0.20***		−0.20***
Measurement component				
Treated with less courtesy	0.91	0.91	0.93	0.93
Treated with less respect	0.92	0.92	0.92	0.92
Receive poorer service	0.77	0.77	0.71	0.71
People act as if you are not smart	0.77	0.78	0.77	0.78
People act as if they are afraid of you	0.59	0.59	0.56	0.57
People act as if you are dishonest	0.75	0.76	0.73	0.74
People act as if they are better than you	0.77	0.77	0.76	0.77
You are insulted	0.73	0.73	0.73	0.72
You are threatened/harassed	0.73	0.73	0.73	0.75
People ignore you	0.77	0.78	0.77	0.76
Model fit				
χ^2	2,657.49	2,570.25	2,446.66	2,313.56
Degrees of freedom	71.00	98.00	52.00	75.00
P value	<0.001	<0.001	<0.001	<0.001

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

[a] The structural component of each model presents the standardized regression estimates, while the measurement component presents the measurement slopes (i.e., standardized factor loadings).

[b] Models 1a and 2a were race/ethnicity-only models.

[c] Models 1b and 2b included adjustment for age, education, and interview language.

[d] Non-English vs. English (referent).

models examining the association between race/ethnicity and the everyday discrimination construct do not account for direct effects, or DIF (models 1a and 1b). As detailed in Table 2, in initial models (model 1a, unadjusted), Caucasian, Hispanic, and Japanese women all had significantly lower scores on the everyday discrimination latent construct than African-American women. The scores of Chinese women were comparable to those of African-American women. After adjusting for age, education, and interview language (model 1b), Caucasian, Hispanic, and Japanese women remained lower on the everyday discrimination latent construct, while Chinese women were slightly (but significantly) higher, compared with African-American women. After accounting for direct effects/DIF (models 2a and 2b), estimates were

slightly lower. Additionally, in comparing models 1b and 2b, DIF accounted for 18% of the difference between African-American and Caucasian women (small effect size (61)) and 36% of the difference between African-American and Japanese women (small-to-moderate effect size (61)). However, the overall pattern of racial/ethnic differences in the everyday discrimination construct remained unchanged and statistically significant.

Direct effects, or DIF, for all final models are shown in Table 3. Items statistically less likely to be endorsed by Caucasian women compared with African-American women were: "You receive poorer service in restaurants or stores," "People act as if they are afraid of you," "You are treated as if you are dishonest," and "You are insulted or harassed."

Table 3. Odds Ratios for Statistically Significant and "Meaningful" Differential Item Functioning on the Everyday Discrimination Scale by Race/Ethnicity After Adjustment for Age, Education, and Language, Study of Women's Health Across the Nation, 1996–1997

	Race/Ethnicity							
	Caucasian		Chinese		Hispanic		Japanese	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Treated with less courtesy					2.67[a]	1.51, 4.72		
Treated with less respect	1.36	1.14, 1.62						
Receive poorer service	0.42[a]	0.35, 0.50			0.20	0.10, 0.38	0.51	0.38, 0.70
People act as if you are not smart							0.60	0.45, 0.76
People act as if they are afraid of you	0.83	0.71, 0.96					0.62	0.48, 0.81
People act as if they think you are dishonest	0.44[a]	0.36, 0.53	0.73	0.54, 0.99			0.61	0.43, 0.86
People act as if they're better than you								
You are insulted	0.63	0.53, 0.74						
You are threatened/harassed					1.12	0.58, 2.15		
People ignore you								

Abbreviations: CI, confidence interval; OR, odds ratio.

[a] Items with odds ratios greater than 2 or less than 0.5 display "meaningful" differential item functioning.

The "treated with less respect" item was more likely to be endorsed by Caucasian women than by African-American women.

Chinese women were statistically less likely to endorse "You are treated as if you are dishonest" compared with African-American women, and Hispanic women were statistically more likely to endorse "being treated with less courtesy" but less likely to endorse "receiving poorer service" compared with African-American women. Among Japanese women, statistically significant DIF was observed for "You receive poorer service in restaurants or stores," "People act as if you aren't smart," "People act as if they are afraid of you," and "You are treated as if you are dishonest"—all less likely to be endorsed in comparison with African-American women. However, using Cole's criterion (odds ratio >2 or odds ratio <0.5) (52), there were only 3 items with meaningful DIF: "you are treated with less courtesy," which Hispanic women were more likely to endorse than African-American women; "you receive poorer service," which Caucasian, Hispanic, and for the most part Japanese women (odds ratio = 0.51) were less likely to endorse than African-American women; and "you are dishonest," which Caucasians were less likely to endorse than African-American women.

DISCUSSION

Findings from the current study revealed that items on the EDS functioned similarly for women from 5 different racial/ethnic groups, with 3 notable exceptions. After adjustment for age, educational status, and interview language, one item—"receiving poorer service in restaurants or stores"—functioned differently for African-American women than for most other groups. A second item—"being treated as if you are dishonest"—functioned similarly for African-American women and women of other racial/ethnic minority groups (Chinese, Japanese, and Hispanic) but differed for Caucasian women. A third item—"being treated with less courtesy than other people"—functioned differently for Hispanic women only compared with African-American women.

In examining differences between African Americans and Caucasians on the scale (two of the most commonly compared groups in this area of study), we found that African-American women differed from Caucasian women on 2 separate items. The 2 items, "poorer service in restaurants or stores" and "being treated as if you are dishonest," parallel anecdotal accounts in popular media (62, 63) and previous findings from qualitative research studies in which African-American women have reported feeling as if they were "being watched" in stores because others perceived that they might steal something (23, 64). In a recent study of discrimination in African-American women, Nuru-Jeter et al. (64) reported, "In the women's everyday lives, shopping was a frequently mentioned context for [discriminatory] experiences. Participants reported being followed in stores, ignored by clerks, and treated disrespectfully or with suspicion or disdain in public settings" (64, p. 34). Similarly, in an examination of the Experiences of Discrimination Scale, Krieger et al. (45) found that the item referencing discriminatory treatment in "getting service in a store or restaurant" functioned differently for African Americans compared with Caucasians and Hispanics. Taken together, these findings suggest that experiences of everyday discrimination may differ slightly for African-American women compared with Caucasian women, with certain "public" encounters having more relevance for African-American women than their Caucasian counterparts.

Interestingly, items on the EDS functioned similarly for African-American and Chinese women, and in fully adjusted MIMIC models, Chinese women actually had higher levels of everyday discrimination than their African-American counterparts. Because there is a relative dearth of research documenting the unique experiences of everyday discrimination for Chinese women, it is unclear whether anecdotal

accounts and prior research findings from African-American women generalize to this group. However, current findings indicate that the experience of everyday discrimination for Chinese women may parallel that of African-American women.

Only 1 item functioned similarly for racial/ethnic minority women overall compared with Caucasian women. "Being treated as if you are dishonest" appears to be a significant component of everyday discrimination for African-American, Chinese, Japanese, and Hispanic women but less relevant for Caucasian women. As noted above, for African-American women, this experience may also be linked to public encounters; however, it is possible that for racial/ethnic minority women in general, being perceived as dishonest occurs with everyday frequency across a variety of settings.

It is unclear why the "courtesy" item was more likely to be endorsed by Hispanic women than by African-American women. It is possible that being treated with less courtesy than others is a particularly salient aspect of the day-to-day discriminatory experiences of Hispanic women. However, because this was the only item demonstrating DIF, this experience does not seem to be linked to other types of exposures that would appear to be consistent with discourteous treatment (i.e., poor service in restaurants/stores, being treated with less respect) in this subgroup. Because most of the qualitative and descriptive studies on day-to-day experiences of discrimination have been conducted among African-American women (22, 23, 64), very little is known about everyday experiences of discrimination among Hispanic women and how their experiences may or may not differ from those of other racial/ethnic groups. Consequently, it is difficult to draw conclusions about these results. Additional research is warranted.

As previously reported in this cohort (26), there were significant mean-level differences in reports of everyday discrimination by race/ethnicity, with African-American and Chinese women reporting the highest levels of everyday discrimination, Hispanic women reporting the lowest levels of everyday discrimination, and Japanese and Caucasian women reporting levels in between. Although the factors underlying this patterning of results remain undetermined, it is important to note that these group-level differences in everyday discrimination persisted even after we adjusted for differences in item functioning. This indicates that the observed racial/ethnic differences in reports of everyday discrimination were not attributable to DIF. Indeed, the magnitude of the coefficient for racial/ethnic differences in reports of everyday discrimination remained relatively unchanged even after adjustment for DIF.

Overall, findings suggest that the EDS can be used across racial/ethnic groups as originally intended. However, some caution should be used. Meaningful DIF was observed in 3 out of 10 items on the SWAN scale. Because there was no DIF on the additional item, "People ignore you," this actually corresponds to 3 out of 9 items on the standard EDS. Thus, one-third of the items on the scale demonstrated DIF. When comparing results across groups for which there was observed DIF (e.g., African Americans compared with Caucasians), investigators may wish to examine associations for each racial/ethnic group sepa-

rately or conduct sensitivity analyses with and without items that demonstrated DIF.

In standard DIF testing, items that demonstrate meaningful DIF are eliminated (48, 49). The advantage of this approach is that it creates a scale that is relatively free of DIF and permits analysis across diverse groups (48, 49). However, this approach also has disadvantages. Eliminating items may lead to changes in the measurement of everyday discrimination experienced by one or more groups. For example, in the case of everyday discrimination, it appears as if the experience, or profile, of everyday discrimination differs for women of different racial/ethnic groups. Thus, for example, eliminating the item "receiving poorer service in restaurants and stores" may improve functioning of the scale across different racial/ethnic groups but might not adequately capture the experience of everyday discrimination for African-American women, given the potential salience of certain "public" encounters for this group. Similarly, eliminating the "dishonest" item when comparing racial/ethnic minority groups with Caucasians might also prove problematic, as would removing the "courtesy" item when comparing Hispanics with African Americans. Researchers using this scale will need to make decisions with these tradeoffs in mind.

Findings from this study should be interpreted in the context of study limitations. First, the current sample was comprised of women only. Although Essed's original work on everyday discrimination was conducted with women (22, 23), the scale is currently used in samples of both women and men. However, little is known about whether item functioning on the scale differs by gender. Thus, it is unclear whether these findings would generalize to men. Second, the majority of women in SWAN are middle-class. Some have argued that middle-class women are more likely to experience discrimination than their lower socioeconomic status counterparts because they often work, live, and socialize in more integrated environments. There is empirical support for this notion (35, 65); nonetheless, findings from the current study may not generalize to women from lower socioeconomic status backgrounds. Third, there are known regional differences within the United States in the concentration of various racial/ethnic groups and their respective histories of discrimination (e.g., Jim Crow laws in the South and Japanese internment in the West). The SWAN study population is concentrated in select areas throughout the United States (the Northeast, Midwest, and West), and thus findings from this population may not generalize to women in other areas (the South).

Further, it is important to note that there may be limitations to the EDS itself. Because items on the scale are based on experiences of a particular racial/ethnic group (i.e., African-American women), it is possible that there are questions more relevant to other racial/ethnic groups (i.e., questions about language proficiency) that were not included (19). It is also noteworthy that, in contrast to some prior studies using this scale (37, 45), the unidimensionality assumption was not fully met in the current sample. Although the 2- and 3-factor models did not fit our data better, it is possible that there are other dimensions of everyday discrimination (for women from all racial/ethnic backgrounds) that were not represented by items on the current scale. It could be that

the current version of the EDS does not include enough items or that the existing items are simply not broad enough to adequately capture the range of experiences that comprise everyday discrimination across racial/ethnic groups. It is also possible that a single measurement model does not sufficiently explain the covariation among the items on the scale across multiple racial/ethnic groups. Additional quantitative (multiple group measurement invariance analyses) and qualitative (cognitive interviewing) research is needed to comprehensively address this issue (66, 67).

Despite the limitations noted above, this study had several strengths. To our knowledge, this was one of the first DIF analyses of the EDS, which is currently one of the most widely used discrimination scales in epidemiologic research. The cohort used in this analysis, SWAN, is large and community-based and includes participants from 5 different racial/ethnic groups, which greatly increases the generalizability of our results. We were also able to control for a number of potential confounders that might have influenced our findings, such as age, education, and interview language.

To conclude, given the growing body of research examining the association between discrimination and health, there is an increasing need for empirical instruments that can be utilized in a number of different populations (20). Results from this analysis provide some preliminary support for the use of one particular scale—the EDS—across several racial/ethnic groups. However, findings suggest that some caution should be used when making cross-racial/ethnic comparisons. Further research on the measurement of day-to-day experiences of discrimination for persons from a variety of sociodemographic backgrounds is needed. Finally, because the current findings are limited to a single scale, additional research examining measurement bias by race/ethnicity across discrimination scales more broadly may be warranted.

ACKNOWLEDGMENTS

REFERENCES

1. Gee GC, Spencer MS, Chen J, et al. A nationwide study of discrimination and chronic health conditions among Asian Americans. *Am J Public Health*. 2007;97(7):1275–1282.
2. Barnes LL, Mendes De Leon CF, Wilson RS, et al. Racial differences in perceived discrimination in a community population of older blacks and whites. *J Aging Health*. 2004; 16(3):315–337.
3. Brondolo E, Brady N, Thompson S, et al. Perceived racism and negative affect: analyses of trait and state measures of affect in a community sample. *J Soc Clin Psychol*. 2008;27(2): 150–173.
4. Krieger N, Sidney S. Racial discrimination and blood pressure: the CARDIA study of young black and white adults. *Am J Public Health*. 1996;86(10):1370–1378.
5. Lewis TT, Everson-Rose SA, Powell LH, et al. Chronic exposure to everyday discrimination and coronary artery calcification in African-American women: the SWAN Heart Study. *Psychosom Med*. 2006;68(3):362–368.

6. Paradies Y. A systematic review of empirical research on self-reported racism and health. *Int J Epidemiol*. 2006;35(4): 888–901.

7. Williams DR, Mohammed SA. Discrimination and racial disparities in health: evidence and needed research. *J Behav Med*. 2009;32(1):20–47.

8. Dominguez TP, Dunkel-Schetter C, Glynn LM, et al. Racial differences in birth outcomes: the role of general, pregnancy, and racism stress. *Health Psychol*. 2008;27(2):194–203.

9. Barnes LL, de Leon CF, Lewis TT, et al. Perceived discrimination and mortality in a population-based study of older adults. *Am J Public Health*. 2008;98(7):1241–1247.

10. Pascoe EA, Smart Richman L. Perceived discrimination and health: a meta-analytic review. *Psychol Bull*. 2009;135(4): 531–554.

11. Taylor TR, Williams CD, Makambi KH, et al. Racial discrimination and breast cancer incidence in US black women: the Black Women's Health Study. *Am J Epidemiol*. 2007; 166(1):46–54.

12. Lewis TT, Kravitz HM, Janssen I, et al. Self-reported experiences of discrimination and visceral fat in middle-aged African-American and Caucasian women. *Am J Epidemiol*. 2011;173(11):1223–1231.

13. Lewis TT, Barnes LL, Bienias JL, et al. Perceived discrimination and blood pressure in older African American and white adults. *J Gerontol A Biol Sci Med Sci*. 2009;64(9):1002–1008.

14. Troxel WM, Matthews KA, Bromberger JT, et al. Chronic stress burden, discrimination, and subclinical carotid artery disease in African American and Caucasian women. *Health Psychol*. 2003;22(3):300–309.

15. Beatty DL, Matthews KA. Unfair treatment and trait anger in relation to nighttime ambulatory blood pressure in African American and white adolescents. *Psychosom Med*. 2009; 71(8):813–820.

16. Friedman EM, Williams DR, Singer BH, et al. Chronic discrimination predicts higher circulating levels of E-selectin in a national sample: the MIDUS study. *Brain Behav Immun*. 2009;23(5):684–692.

17. Hunte HE, Williams DR. The association between perceived discrimination and obesity in a population-based multiracial and multiethnic adult sample. *Am J Public Health*. 2009; 99(7):1285–1292.

18. Hunte HE. Association between perceived interpersonal everyday discrimination and waist circumference over a 9-year period in the midlife development in the United States Cohort Study. *Am J Epidemiol*. 2011;173(11):1232–1239.

19. Gee GC, Ro A, Shariff-Marco S, et al. Racial discrimination and health among Asian Americans: evidence, assessment, and directions for future research. *Epidemiol Rev*. 2009;31(1): 130–151.

20. Bastos JL, Celeste RK, Faerstein E, et al. Racial discrimination and health: a systematic review of scales with a focus on their psychometric properties. *Soc Sci Med*. 2010;70(7):1091–1099.

21. Williams DR, Yu Y, Jackson JS, et al. Racial differences in physical and mental health. *J Health Psychol*. 1997;2(3): 335–351.

22. Essed P. *Everyday Racism: Reports From Women of Two Cultures*. Alameda, CA: Hunter House Publishers; 1990.

23. Essed P. Knowledge and resistance: black women talk about racism in the Netherlands and the USA. *Fem Psychol*. 1991; 1(2):201–219.

24. Gee GC, Ro A, Gavin A, et al. Disentangling the effects of racial and weight discrimination on body mass index and obesity among Asian Americans. *Am J Public Health*. 2008; 98(3):493–500.

25. Gee GC, Chen J, Spencer MS, et al. Social support as a buffer for perceived unfair treatment among Filipino Americans: differences between San Francisco and Honolulu. *Am J Public Health*. 2006;96(4):677–684.

26. Brown C, Matthews KA, Bromberger JT, et al. The relation between perceived unfair treatment and blood pressure in a racially/ethnically diverse sample of women. *Am J Epidemiol*. 2006;164(3):257–262.

27. Smart Richman L, Pek J, Pascoe E, et al. The effects of perceived discrimination on ambulatory blood pressure and affective responses to interpersonal stress modeled over 24 hours. *Health Psychol*. 2010;29(4):403–411.

28. Lewis TT, Aiello AE, Leurgans S, et al. Self-reported experiences of everyday discrimination are associated with elevated C-reactive protein levels in older African-American adults. *Brain Behav Immun*. 2010;24(3):438–443.

29. Guyll M, Matthews KA, Bromberger JT. Discrimination and unfair treatment: relationship to cardiovascular reactivity among African American and European American women. *Health Psychol*. 2001;20(5):315–325.

30. Osterlind SJ, Everson HT. *Differential Item Functioning*. Thousand Oaks, CA: Sage Publications, Inc; 2009.

31. Holland P, Wainer H. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1993.

32. Embretson S, Reise S. *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2000.

33. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000;38(suppl 9):II28–II42.

34. De Ayala RJ. *The Theory and Practice of Item Response Theory*. New York, NY: The Guilford Press; 2009.

35. Dailey AB, Kasl SV, Holford TR, et al. Neighborhood- and individual-level socioeconomic variation in perceptions of racial discrimination. *Ethn Health*. 2010;15(2):145–163.

36. Borrell LN, Jacobs DR Jr, Williams DR, et al. Self-reported racial discrimination and substance use in the Coronary Artery Risk Development in Adults Study. *Am J Epidemiol*. 2007; 166(9):1068–1079.

37. Kessler RC, Mickelson KD, Williams DR. The prevalence, distribution, and mental health correlates of perceived discrimination in the United States. *J Health Soc Behav*. 1999; 40(3):208–230.

38. Pérez DJ, Fortuna L, Alegria M. Prevalence and correlates of everyday discrimination among U.S. Latinos. *J Community Psychol*. 2008;36(4):421–433.

39. Sowers M, Crawford S, Sternfeld B, et al. SWAN: a multicenter, multiethnic, community-based cohort study of women and the menopausal transition. In: Lobo RA, Kelsey J, Marcus R, eds. *Menopause: Biology and Pathobiology*. New York, NY: Academic Press, Inc; 2000:175–188.

40. Harrell SP. A multidimensional conceptualization of racism-related stress: implications for the well-being of people of color. *Am J Orthopsychiatry*. 2000;70(1):42–57.

41. Harrell SP, Merchant MA, Young SA. Psychometric properties of the Racism and Life Experiences Scales (RaLES). Presented at the 105th Annual Convention of the American Psychological Association, Chicago, Illinois, August 15–19, 1997.

42. Sims M, Wyatt SB, Gutierrez ML, et al. Development and psychometric testing of a multidimensional instrument of perceived discrimination among African Americans in the Jackson Heart Study. *Ethn Dis*. 2009;19(1):56–64.

43. Chae DH, Takeuchi DT, Barbeau EM, et al. Alcohol disorders among Asian Americans: associations with unfair treatment, racial/ethnic discrimination, and ethnic

identification (the National Latino and Asian Americans Study, 2002–2003). *J Epidemiol Community Health.* 2008;62(11):973–979.

44. Taylor TR, Kamarck TW, Shiffman S. Validation of the Detroit Area Study Discrimination Scale in a community sample of older African American adults: the Pittsburgh Healthy Heart Project. *Int J Behav Med.* 2004;11(2):88–94.

45. Krieger N, Smith K, Naishadham D, et al. Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health. *Soc Sci Med.* 2005;61(7):1576–1596.

46. Teresi JA, Golden RR, Cross P, et al. Item bias in cognitive screening measures: comparisons of elderly white, Afro-American, Hispanic and high and low education subgroups. *J Clin Epidemiol.* 1995;48(4):473–483.

47. Sorbom D. Model modification. *Psychometrika.* 1989;54: 371–384.

48. Angoff W. Perspective on differential item functioning methodology. In: Holland P, Wainer H, eds. *Differential Item Functioning.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1993:3–24.

49. Burton E, Burton NW. The effect of item screening on test scores and test characteristics. In: Holland P, Wainer H, eds. *Differential Item Functioning.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1993:321–336.

50. Yang FM, Jones RN. Center for Epidemiologic Studies-Depression Scale (CES-D) item response bias found with Mantel-Haenszel method was successfully replicated using latent variable modeling. *J Clin Epidemiol.* 2007;60(11):1195–1200.

51. Yang FM, Jones RN. Measurement differences in depression: chronic health-related and sociodemographic effects in older Americans. *Psychosom Med.* 2008;70(9):993–1004.

52. Cole SR, Kawachi I, Maller SJ, et al. Test of item-response bias in the CES-D scale. Experience from the New Haven EPESE study. *J Clin Epidemiol.* 2000;53(3):285–289.

53. Browne M, Cudek R. Alternative ways of assessing model fit. In: Bollen K, Long J, eds. *Testing Structural Equation Models.* Thousand Oaks, CA: Sage Publications, Inc; 1993:136–162.

54. Muthén B, Khoo S-T, Francis D. *Multi-Stage Analysis of Sequential Developmental Processes to Study Reading Progress: New Methodological Developments Using General Growth Mixture Modeling.* Los Angeles, CA: Graduate School of Education and Information Studies, University of California, Los Angeles; 1998.

55. Bentler P, Chou C. Practical issues in structural equation modeling. In: Long J, ed. *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research.* Newbury Park, CA: Sage Publications, Inc; 1988.

56. Muthén B. *The Development of Heavy Drinking and Alcohol Related Problems From Ages 18 to 37 in a U.S. National Sample.* Los Angeles, CA: Graduate School of Education and Information Studies, University of California, Los Angeles; 1998.

57. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107(2):238–246.

58. Muthén BO. Latent variable modeling in heterogeneous populations. *Psychometrika.* 1989;54(4):557–585.

59. Muthén LK, Muthén BO. *Mplus, Version 6.11.* Los Angeles, CA: Muthén & Muthén; 1998–2011.

60. StatSoft, Inc. *Electronic Statistics Textbook.* Tulsa, OK: StatSoft, Inc; 2011. (http://www.statsoft.com/textbook/). (Accessed June 18, 2011).

61. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1988.

62. Fifield A, O'Shaughnessy E. Shopping while black. *Good Housekeeping.* 2001;233(5):129–136.

63. Stein R. Study links discrimination, blacks' health. *Washington Post.* 2005;May 1:A17.

64. Nuru-Jeter A, Dominguez TP, Hammond WP, et al. ''It's the skin you're in'': African-American women talk about their experiences of racism. An exploratory study to develop measures of racism for birth outcome studies. *Matern Child Health J.* 2009;13(1):29–39.

65. Hunt MO, Wise LA, Jipguep MC, et al. Neighborhood racial composition and perceptions of racial discrimination: evidence from the Black Women's Health Study. *Soc Psychol Q.* 2007;70(3):272.

66. Nápoles-Springer AM, Santoyo-Olsson J, O'Brien H, et al. Using cognitive interviews to develop surveys in diverse populations. *Med Care.* 2006;44(11):S21–S30.

67. McHorney CA, Fleishman JA. Assessing and understanding measurement equivalence in health outcome measures: issues for further quantitative and qualitative inquiry. *Med Care.* 2006;44(11):S205–S210.

Appendix Table 1. Response Frequencies and Mean Scores for Items on the Everyday Discrimination Scale, Study of Women's Health Across the Nation, 1996–1997

Item	Response Frequency, %				Mean Score[a] (Standard Deviation)
	Never	Rarely	Sometimes	Often	
Treated with less courtesy	23.4	50.4	24.1	2.1	2.05 (0.75)
Treated with less respect	25.7	51.9	20.7	1.8	1.99 (0.73)
Receive poorer service at restaurants and in stores	33.9	51.2	13.9	1.0	1.82 (0.70)
People act as if they think you are not smart	38.0	44.0	15.8	2.2	1.82 (0.77)
People act as if they are afraid of you	52.6	33.2	12.7	1.5	1.63 (0.76)
People act as if they think you are dishonest	67.8	26.8	4.7	0.8	1.38 (0.61)
People act as if they're better than you	26.4	45.2	25.9	2.5	2.04 (0.79)
You are called names or insulted	66.6	26.4	6.3	0.7	1.41 (0.64)
You are threatened/harassed	72.2	22.5	4.4	0.9	1.34 (0.61)
People ignore you or act as if you aren't there	44.3	40.2	14.6	1.0	1.72 (0.74)

[a] Possible range: 1 to 4; higher scores indicate higher reports of everyday discrimination.

Appendix Table 2. Polychoric Correlation Matrix for Items on the Everyday Discrimination Scale, Study of Women's Health Across the Nation, 1996–1997

Item	Item									
	Courtesy	Respect	Poor Service	Not Smart	Afraid	Dishonest	Better	Insulted	Threatened	Ignore
Courtesy	1									
Respect	0.89	1								
Poor service	0.68	0.69	1							
Not smart	0.65	0.66	0.61	1						
Afraid	0.42	0.42	0.46	0.47	1					
Dishonest	0.52	0.53	0.60	0.61	0.60	1				
Better than you	0.62	0.64	0.61	0.67	0.44	0.49	1			
Insulted	0.49	0.54	0.52	0.47	0.41	0.56	0.46	1		
Threatened/ harassed	0.46	0.49	0.45	0.44	0.44	0.50	0.44	0.72	1	
Ignore	0.60	0.63	0.59	0.61	0.45	0.48	0.59	0.49	0.60	1