



Published in final edited form as:

Psychosom Med. 2008 November ; 70(9): 993–1004. doi:10.1097/PSY.0b013e31818ce4fa.

Measurement Differences in Depression: Chronic Health-Related and Sociodemographic Effects in Older Americans

Frances M. Yang PhD^{1,2} and Richard N. Jones ScD^{2,3}

¹Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

²Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School, Boston, Massachusetts

³Division of Gerontology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts

Abstract

Objective—We evaluated the influence of five chronic health conditions (high blood pressure, heart conditions, stroke, diabetes, and lung diseases) and four sociodemographic characteristics (age, gender, education, and race/ethnicity) on the endorsement patterns of depressive symptoms in a sample of community-dwelling older adults.

Method—Participants were adults aged 65+ from the 2004 Health and Retirement Study (N=9,448). Depressive symptoms were measured with a nine-item Center for Epidemiologic Studies-Depression scale. Measurement differences attributable to health and sociodemographic factors were assessed with a multidimensional model based on item response theory.

Results—Evidence for unidimensionality was equivocal. Therefore, we used a bifactor model to express symptom endorsement patterns as resulting from a general factor and three specific factors (“dysphoria,” “psychosomatic,” and “lack of positive affect”). Even after controlling for the effects of health on the psychosomatic factor, heart conditions, stroke, diabetes, and lung diseases had significant positive effects on the general factor. Significant effects due to gender and educational levels were observed on the “lack of positive affect” factor. Older adults self-identifying as Latinos had higher levels of general depression. On the symptom level, meaningful measurement noninvariance due to race/ethnic differences were found in the following five items: depressed, effort, energy, happy, and enjoy life.

Conclusions—The increased tendency to endorse depressive symptoms among persons with specific health conditions is in part explained by specific associations among symptoms belonging to the psychosomatic domain. Differences attributable to the effects of health conditions may reflect distinct phenomenological features of depression. The bifactor model serves as a vehicle for testing such hypotheses.

Keywords

Center for Epidemiological Studies-Depression (CES-D); cardiovascular disease and risk factors (CVRFs); chronic health conditions, Item Response Theory (IRT); Differential Item Functioning (DIF); Multiple Indicators Multiple Causes (MIMIC)

Introduction

Chronic illness and acute health events are more frequent among older adults, and the comorbidity of somatic illness and psychiatric distress complicates the diagnosis and treatment of depression (1). Understanding how chronic illness affects measurement of depression while controlling for sociodemographic characteristics would help us understand the presentation of depressive symptoms among older adults with somatic illness.

The National Heart, Lung and Blood Institute (NHLBI) (2) reports that 15–20% of heart disease patients suffer from major depression, with even more suffering from subsyndromal forms (3). A Working Group (3) has identified the need for more standardized assessment, diagnosis, and treatment of depression among patients with cardiovascular disease (CVD) to help reduce the progression of depression and other comorbidities.

Older adults are more likely than younger ones to mask depression with somatic rather than psychiatric symptoms (4,5). Depressed older Black or African Americans are more likely than Whites to report somatic symptoms (6). It is hypothesized that such differences in the patterns of depressive symptoms are not purely due to race, ethnic, or cultural differences, but reflect other socio-demographic and chronic health conditions (7–9). For example, Jonas and Lando (10) found a higher risk of hypertension among Black women with anxiety and/or depression, as compared to White women, White men, and Black men.

The measurement of depression is likely affected by sociodemographic factors (11–13). One of the earliest studies examining measurement bias due to gender differences used the widely known Center for Epidemiological Studies-Depression (CES-D) scale. Eliminating five CES-D symptoms with gender bias helped reduce the mean difference in levels of depressive symptomatology between adult males and females (14). Callahan and Wolinsky examined the factor structure of the CES-D due not only to gender, but also due to race/ethnicity in an older population, showing that five items contributed to disparities (15). Cole and colleagues conducted the first study to simultaneously examine age, gender, and race/ethnic bias in the CES-D for an older population (12). They found that older Blacks were more prone to endorse interpersonal symptoms in the CES-D than older Whites; while older females, regardless of race/ethnicity, were likely to endorse the “crying” symptom, which was consistent with earlier findings (14). The findings of Cole and colleagues were confirmed using a multiple indicators, multiple causes (MIMIC) model by Yang and Jones (16). These studies show measurement bias in CES-D items attributable to gender and race/ethnic differences among older adults.

Ross and Mirowsky (17) found the following four-factor structure in the original 20-item CES-D: depressed affect, enervation, lack of positive affect, and interpersonal problems. In this analysis, we used a large nationally representative sample of older adults (the Health and Retirement Study (HRS)) (18) to examine measurement noninvariance in the nine-item Center for Epidemiologic Studies—Depression (CES-D) Scale under the MIMIC model. Based on the nine available items in the HRS/CES-D, which do not include the symptoms characterizing the interpersonal factor in the original CES-D, we predicted the following three specific factors: “dysphoria,” “psychosomatic,” and “lack of positive affect.”

We assess the dimensionality of the HRS/CES-D by considering both a unidimensional and multidimensional model simultaneously by using a bifactor model (e.g., 19). The bifactor model (20) consists of one general factor (unidimensional) and two or more specific factors (multidimensional), which each item loading on both the general factor and one specific factor. The general factor is used to explain the correlation between all items, and the specific factors explain residual covariation among a sub-set of items (20). In the case of depression, the bifactor model shows that symptoms are correlated because they share a common general trait

and one independent source of common variation (e.g., a tendency to endorse somatic symptoms).

We then proceeded to test for measurement noninvariance with respect to health conditions and sociodemographic characteristics on both the latent depressive factors, found in the bifactor model, and individual symptoms of depression. This line of research is consistent with new directions in item bias modeling based on multidimensional test structures (c.f., 20,21). Our evaluation is unique, as we examine simultaneously the effect of sociodemographic and chronic health conditions in a multidimensional structure of depression. Few studies have examined the relationship between the CES-D factor structure and the health and sociodemographic characteristics of older adults in a US population.

Methods

Sample

Data were obtained from the 2004 *Health and Retirement Study* (HRS), whose details have been reported before (18). Briefly, the HRS was designed to inform major policy affecting retirement, health insurance, and economic well-being. The first wave (1992) of the HRS consisted of face-to-face interviews with a representative sample of U.S. adults aged 51 to 61 years and their spouses, with over-samples of Blacks, Hispanics, and Florida residents. It was joined in 1993 by a companion study, Assets and Health Dynamics of the Oldest Old (AHEAD), which took advantage of the initial eligibility screening of the HRS study to identify adults born before 1924 and their spouses (22). In 1998 the HRS and AHEAD studies were combined with two new cohorts (persons born 1924–30 and 1942–47 (23)) using public-use data sets from the Institute for Social Research (ISR) at the University of Michigan (<http://hrsonline.isr.umich.edu>). ISR also provides rich documentation of the protocol, instrumentation, sampling strategy, and statistical weighting procedures. Only respondents who were 65 and older in 2004 with complete CES-D, background, and health data were included (N=9,448).

Measures

Depression—The 2004 HRS protocol included a modified telephone-administered version of the *Centers for Epidemiologic Studies Depression Scale* (CES-D) (24) we call the HRS/ CES-D. Participants were asked, “Now think about the past week and the feelings you have experienced. Please tell me if each of the following was true for you much of the time during the past week.” Then the interviewer asked nine questions, such as, “Much of the time during the week, you felt depressed. (Would you say yes or no?).” We reverse-scored the following three positively worded symptoms: happy, enjoyed life, and full of energy. The standard CES-D has been shown to measure depressive symptomology with very high internal consistency and reliability (25,26), including studies of older Hispanics (27) and older Blacks (28).

Sociodemographics—The HRS survey contained extensive sociodemographic questions, including self-identified race/ethnicity, age, sex, and education. Subjects were first asked if they considered themselves Hispanic or Latino(a). This question was followed by: “Do you consider yourself primarily White or Caucasian, Black or African American, American Indian, or Asian, or something else?” Only Whites, Blacks, and Hispanics were included in this study, because the remaining race categories of “American Indian, or Asian, or something else” were collapsed into an “other” category in the public-use data. Participants were asked their educational level: no formal education, grades 1–11, high school graduate, some college, college graduate, and post-college (17+ years). The effects of age, gender, and education were modeled as linear effects and were mean-centered to make meaningful interpretations of the intercept parameters (29).

Chronic health conditions—A number of chronic conditions were assessed in the health sub-section of the HRS, based upon the respondent's answer to the questions that began with: "Has a doctor ever told you that you had..." high blood pressure or hypertension, heart conditions (including heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems), recent stroke (event or sequelae requiring visit to physician within two years of interview for HRS respondents, any history of stroke for AHEAD respondents), diabetes or high blood sugar, and chronic lung diseases (e.g., chronic bronchitis or emphysema). Answering "yes" to one of these chronic conditions did not exclude the respondent from answering "yes" to another condition.

Statistical Analyses

Statistical analyses included exploratory factor analysis, confirmatory factor analysis, and a bifactor measurement model of depression using model-fitting strategies adapted from Differential Item Functioning (DIF) methods. Our use of confirmatory factor analysis models and bifactor measurement models are based in item response theory (IRT) by analyzing the observed binary response variables using a system of multivariate probit or logit regressions. IRT, also referred to as latent trait theory, is used to estimate the level on the latent trait of an individual responding to a test item, the severity of the test item, and the accuracy with which the item measures the underlying trait (i.e. the item discrimination) (30,31). Yang and Jones' (16) discussion of IRT and its use in mental health and aging research provides the foundation for the models in this study. DIF, also known as item bias or measurement noninvariance, occurs when respondents from different groups with the same latent trait respond differently to the same item (32,33).

Analyses were conducted using Mplus version 5.0. We used a combination of approaches to determine dimensionality of the HRS/CES-D. Based on exploratory factor analysis (EFA), we determined whether the first eigenvalue was >3.5 times the second and subsequent eigenvalues and the magnitude of the factor loadings on the first factor to help determine unidimensionality (34). In addition, we conducted a modified parallel analysis (35) and used a theory-driven approach to compare the simple structure model to the single factor model of the HRS/CES-D using confirmatory factor analysis (CFA).

Based on the dimensionality assessment, we used the bifactor measurement model to compare the factor loadings of HRS/CES-D items on the general depression factor and the three specific factors (dysphoria, psychosomatic, and lack of positive affect) as shown in Figure 2a. The "bi" in the bifactor model refers to each item loading on the general depression factor and one additional specific factor, which are both first-order factors. For example, the depressed item loads on both the general depression factor and the specific dysphoria factor in Figure 2a.

The substantive interpretation of the bifactor model is that the items are correlated because they are caused by a general depressive trait and by one additional source common to at least two other items. The common and specific factors are assumed uncorrelated as a matter of model identification. The bifactor model is different from a simple structure first-order model with multiple factors (multiple common factors that are either correlated or uncorrelated) and a second-order factor model (a first-order factor correlated with a higher-order factor).

One purpose of the bifactor model is to assess the assumption of unidimensionality. If the loadings on the general factor are greater than the loadings on the specific factors, the items are viewed as belonging to a sufficiently unidimensional set. Based upon the few published studies using the bifactor model (20,36), the assessment of unidimensionality is based upon the goodness-of-fit indices, with the lowest coefficients representing the model that best fits the data, regardless of the number of factors.

The bifactor model addresses the problem of multidimensionality by capturing the item covariation for “specific” or “nuisance” factors, independent of the item covariation in the general factor. The specific factors in the bifactor model are conceptually distinct from the specificities or uniqueness of standard factor models. We looked for evidence of DIF effects on the general and specific factors of depression, as well as on the depressive symptom items, with respect to chronic health conditions and sociodemographic characteristics. A significant and unmediated effect of a specific chronic health condition or sociodemographic characteristic on an observed CES-D item indicates some degree of DIF for that item.

We evaluated DIF—the condition in which a randomly selected person from a group defined by a sociodemographic or health condition is predicted to respond differently to any of the nine CES-D symptoms relative to a person from the complement group at the same latent trait level—using the multiple indicators, multiple causes (MIMIC) model. MIMIC is an analytic approach grounded in modern measurement theory (37) and is an item response theory (IRT)-based structural equation model (SEM) (13,38). This was originally presented as a model of unobserved quantities measured by formative causes and reflective indicators (39). Only recently has the MIMIC model been used to investigate issues of measurement noninvariance, when individuals from different groups at the same level of the latent trait have different probabilities of endorsing a given symptom (13).

We examined both group differences and DIF effects. Group differences, which are also referred to as group effects throughout this study, are parameters capturing underlying differences in depression severity: relationships between observed covariates (sociodemographic and health characteristics) and observed dependent variables (individual depressive symptoms) mediated (at least in part) via the general latent trait. DIF effects capture the influence of an independent variable on the observed dependent variables while adjusting for the underlying depression severity traits and covariates (32).

To build the MIMIC models, we begin with a model with all group effects freely estimated and all DIF effects constrained to zero. Using modification indices (MI; fit derivatives scaled as chi-square, χ^2) for the regressions of CES-D symptoms and factors on the health and sociodemographic variables, we iteratively identify DIF effects that would significantly improve fit. The MI reveals the effects of health and sociodemographic variables on the HRS/CES-D symptoms and factors that, if freely estimated, would significantly improve model fit. Iteratively, we relaxed constraints on these parameters, allowing regression effects to be freely estimated one at a time. At each iteration, we used robust chi-square difference testing (DIFFTEST) in a forward-fitting stepwise model-building strategy to determine the significance of modifications (see Yang and Jones (16)). We incorporated complex sampling design weights in the weighted least squares mean and variance (WLSMV) adjusted estimator that implements a multivariate probit model for the DIFFTEST procedure.

Lastly, we re-estimated our final models using a multivariate logit parameterization using robust maximum likelihood (MLR) methods to express DIF effects of independent variables on the likelihood of endorsing a specific item as odds ratios (exponentiated regression coefficients from logistic regression models) (40). As a practical matter, the configuration of the Mplus software forces us to use the MLR estimator for the final multivariate model to estimate odds ratios, which provides an intuitive sense of the magnitude of DIF effects for clinical and public health audiences.

Developing an overall assessment of DIF for the bifactor model requires an examination of both the significance and the amount of DIF effects on the items and the group effects mediated by factors specific to select item clusters. DIF indices, especially in large samples, need magnitude measures to focus inferences away from trivial but statistically significant effects.

With our large sample size, trivial measurement differences will be detected as statistically significant. However, the IRT field is currently without clearly articulated thresholds for the description of “small” or “large” DIF. Thus, we have used Cole and colleagues’ rule (41) as a magnitude criterion: proportional odds ratio >2.0 or <0.5 shows that those in the focal group have more than twice the odds of endorsing the individual item than those in the reference group, after being matched on overall latent depression.

Model fit was assessed with the root mean square error of approximation (RMSEA) (42,43) and the comparative fit index (CFI) (44,45). The RMSEA provides a measure of discrepancy per model degree of freedom and approaches zero as fit improves. Browne and Cudek (42) recommended rejecting models with RMSEA values greater than 0.1; Hu and Bentler (46) suggested that values close to 0.06 or less indicate adequate model fit. The CFI ranges between 0 and 1; values greater than 0.95 generally indicate adequate fit (38,47). Other more commonly used fit indices for structural equation modeling applications are not currently available for models that include ordinal dependent variables (48).

Results

Our sample was elderly (aged 65 to 102, average age 74) and included a slight majority of women. Most had completed high school or higher-level education. The most frequent health condition was high blood pressure, which does not preclude other chronic conditions, such as heart conditions and lung diseases.

Factor Analysis

We evaluated up to a four-factor solution under EFA (Table 2). Based upon the factor correlations and factor loadings in the EFA, a single-factor solution was a reasonable fit to the HRS/CES-D: all items have large and positive loadings on the first factor, and the first Eigenvalue ($=5.17$) is >3.5 times larger than others ($=1.11$). Modified parallel analysis also supports the assumption of unidimensionality (Figure 1) (35).

However, the case for unidimensionality is not entirely clear-cut. Based upon the RMSEA, the results shown in Table 2 suggest that a single factor model does not provide satisfactory fit, in line with previous research demonstrating multidimensionality in the CES-D. Among the two-, three-, and four-factor solutions, a three-factor solution appears the best fit (lowest RMSEA).

We evaluated the three-factor solution in a CFA framework. We estimated both the single-factor ($\chi^2=1518.72$, $df=22$, $p<0.001$; CFI= 0.93; RMSEA=0.09) and three-factor structure ($\chi^2=538.07$, $df=21$, $p<0.001$; CFI= 0.98; RMSEA=0.05). Based on the goodness-of-fit statistics, the three-factor structure meets the criteria of CFI >0.95 and RMSEA ≤ 0.05 . The three dimensions of the CES-D were consistent with proposed theoretical constructs (17): 1) “dysphoria” (*depressed, lonely, and sad*), 2) “psychosomatic” (*trouble getting going, no energy, everything was an effort, and restless sleep*), and 3) “lack of positive affect” (*not happy and not enjoy life*), since we had reverse-scored the positive items.

Since the evidence for unidimensionality was equivocal, we estimated a bifactor model, which fit the data well ($\chi^2=307.90$, $df=17$, $p<0.001$; CFI= 0.99; RMSEA=0.04) (20). The bifactor model was parameterized with a global factor and specific factors paralleling the simple-structure three-factor CFA model. Figure 2a shows the factor loadings for the bifactor model of the CES-D symptoms. Note the low loadings of *depressed, lonely, and sad* items on the “dysphoria” factor, which also showed no significant variance. We decided to drop the dysphoria-specific factor, because its estimated variance was not significantly different from zero and the content of the observed items was most closely aligned with the general depression

factor. Figure 2b shows the factor loadings for the reduced bifactor model with only two specific factors of “psychosomatic” and “lack of positive affect.”

Based on the factor analysis, we tested three different MIMIC models (Models 1a-3b) for measurement noninvariance that might be due to (a) only chronic health conditions and (b) both chronic health conditions and sociodemographic characteristics. We evaluated the effects of only chronic health conditions on the measurement of general depression in Model 1a, while in Model 1b we evaluated the effects of both chronic health conditions and sociodemographic characteristics on the measurement of one general depressive factor. Next, we examined the bifactor model with two specific latent depressive factors (“psychosomatic” and “lack of positive affect”) and a general depressive factor (Models 2a-3b). We examined group effects of covariates on specific depression factors, which reflect associations of covariates and either general or specific latent traits. Covariates were included hierarchically as health conditions in Model 2a, and health conditions with sociodemographic characteristics in Model 2b. The final model (3a) is the MIMIC bifactor model, which includes the effects of the chronic health conditions on the specific depressive factors, and Model 3b includes the DIF effects of both chronic health conditions and sociodemographic characteristics.

Model 1: Unidimensional HRS/CES-D MIMIC Model

The results for the MIMIC measurement model treating the CES-D symptoms as unidimensional are presented in Table 3 as Models 1a (with chronic health conditions) and 1b (chronic health conditions and sociodemographic characteristics), with standardized factor loadings for the common factor in the HRS/CES-D symptoms. The models showed inadequate fit based on the criteria mentioned in the Methods section (Model 1a: $\chi^2= 605.26$, $df= 19$, $p<0.001$; CFI= 0.94; RMSEA=0.06; Model 1b: $\chi^2= 534.35$, $df= 23$, $p<0.001$; CFI= 0.93; RMSEA=0.05). The loadings for Models 1a and 1b showed that all items of the CES-D were similar and significantly related to the underlying general depressive factor. The “depressed” (standardized loading=0.88), “not happy” (standardized loadings for Models 1a and 1b are 0.82 and 0.86, respectively), and “not enjoy life” (standardized loadings for Models 1a and 1b are 0.83 and 0.87, respectively) symptoms were most strongly related to the general depressive factor.

Table 4 displays the structural model for group differences in only chronic health conditions (Model 1a) and both sociodemographic and chronic health conditions (Model 1b) on the general latent depressive factor. The general factor was significantly related to each of the background variables, except among Blacks. Persons with lung diseases had a 0.36 standard deviation (SD) units greater general depression relative to persons without lung diseases, while holding all other covariates constant. Global depression levels increased with having a health condition and advancing age. Among women, compared to men, global depression levels were also higher, as was also the case among Latino participants compared to Whites. But overall depression levels were lower among persons with increasing educational attainment in one year increments.

Model 2: Bifactor MIMIC Model with Two Specific Factors

The second model fit was a bifactor model with “psychosomatic” and “lack of positive affect.” Model 2a was the bifactor model with just chronic health conditions ($\chi^2= 163.77$, $df= 21$, $p<0.001$; CFI= 0.98; RMSEA=0.03), which showed good fit. Model 2b was a bifactor model with additional sociodemographic characteristics, also showing adequate fit ($\chi^2= 298.81$, $df= 27$, $p<0.001$; CFI= 0.96; RMSEA=0.03).

As in the unidimensional MIMIC model (Models 1a-b), the loadings are similar in Models 2a and 2b, also showing that the *depressed*, *lonely*, *sad*, *not happy*, and *not enjoy life* symptoms

were strongly correlated with the underlying general depressive factor. *Everything was an effort, restless sleep, trouble getting going, and no energy* had the lowest correlations with the general depressive factor. These items loaded significantly on Factor 2 (“psychosomatic” factor); with *everything was an effort* and *restless sleep* more highly correlated with Factor 2 than the general depressive factor. Although *not happy* and *not enjoy life* are related to “lack of positive affect” (Factor 3), they loaded higher on the general depressive factor.

Table 4 details the standardized group differences on the latent common factors. All of the differences, even when statistically significantly different from zero, were of trivial to small magnitude using Cohen’s effect size taxonomy (49). In Model 2a, older persons with each chronic health condition experienced significantly higher levels of the general depressive factor than the complement groups. These same differences were also found in Model 2b, along with having higher general depressive levels experienced by Blacks, Latinos, and females compared to their counterpart groups of Whites and males, respectively. There was also a .07 standardized unit increase in general depression level for every increase in one year above the mean age of 74. After controlling for sociodemographic characteristics in Model 2b, compared to those who did not report having high blood pressure during the past year, those with high blood pressure no longer experienced significantly higher levels of the general depressive factor. Thus the association of high blood pressure with general depression is likely confounded by sociodemographic characteristics. The other health conditions were relatively unconfounded in their relationship with general depression by sociodemographic characteristics.

It is interesting to compare the effect of health conditions, while controlling for sociodemographic characteristics, on the general depression factor in the unidimensional model (Table 4, Model 1b) and the bifactor model with two specific factors (Model 2b). Despite larger standardized effects on the “psychosomatic” factor, the effects of chronic health conditions on the general factor remain essentially unchanged across models. That is, in a unidimensional model (Model 1b), persons with heart conditions have only a 0.01 SD unit higher level of depression, on average, than in Model 2b. The implication is that, despite the fact that older persons with heart conditions tend to selectively endorse symptoms loading on the psychosomatic domain, this does not bias the association of heart conditions and the general depressive trait.

An interesting pattern emerges for minority older adults. The level of general depression is greater for Blacks relative to Whites in Model 2b, but not in Model 1b. This implies that the association of race/ethnicity and depressive severity in the unidimensional case is an underestimation, deflated by the inclusion of specific factors that cancel the independent association of race/ethnicity and a general depressive trait, after removing the variance specifically shared by both “psychosomatic” and “lack of positive affect” factors. Both Blacks and Latinos, relative to Whites, had lower levels of the “lack of positive affect” and “psychosomatic” factors.

The “lack of positive affect” factor was influenced by only one chronic health condition: lung diseases (Models 2a-b). All sociodemographic characteristics, except for Latinos, had significantly higher magnitudes of the standardized estimates for the “lack of positive affect” factor (Model 2b). Again, Blacks had lower levels of the “lack of positive affect” factor than Whites. However, this effect was nearly four times greater (-0.47 SD units, $p < .05$) for the “lack of positive affect” factor than the “psychosomatic” factor (-0.12 SD units, $p < .05$). Females have a lower “lack of positive affect,” even though they have a 0.29 SD unit higher level of depression than males. Although persons with higher educational achievement had greater levels of “lack of positive affect” (0.26 SD units, $p < .001$), they experienced lower levels of general depression (-0.23 SD units, $p < .001$) when compared to those with lower educational

achievement (Model 2b). At older ages, individuals tend to report more depression but less of levels of the “lack of positive affect” factor.

Model 3: Final Bi-Factor MIMIC model with DIF Effects

After assessing the associations of health and sociodemographic characteristics on general and two specific (but orthogonal) factors, we asked whether any of these relationships are due to specific symptoms rather than (or in addition) to specific factors. We included DIF effects of health covariates only on individual items identified using a stepwise forward model-building procedure (16), which showed adequate fit ($\chi^2=146.67$, $df=20$, $p<0.001$; CFI= 0.99; RMSEA=0.03). As for Model 3b (DIF effects of health conditions and sociodemographic characteristics), there was adequate fit as well ($\chi^2=114.02$, $df=25$, $p<0.001$; CFI= 0.99; RMSEA=0.02). The loadings of the unidimensional model (Model 1), and Models 3a and 3b were very similar. In fact, all the factor loadings for the general depressive trait in the bifactor model are almost equal to those of the unidimensional model, as predicted by Reise, Morizot, and Hays (20).

Table 3 displays the measurement model of latent depression. The notable differences across the measurement portion of the bifactor models (2a-3b) are that the *restless sleep* item now does satisfy the criterion for essential unidimensionality (loading on the general factor is two-fold greater than the loading on the specific factor), but the *no energy* item does not satisfy this criterion. Ultimately, these two items are perhaps not the best indicators of a general depressive construct, due to their high affinity with an orthogonal dimension that captures shared variance among items assessing symptoms of anergia and sleep disturbance.

In the structural component of the models (Table 4), we found group differences for endorsing the “psychosomatic” factor for each of the chronic health conditions. However, the “lack of positive affect” factor was influenced by only sociodemographic variables: gender and educational differences. Women were more likely to experience lower “lack of positive affect” than men (fully standardized parameter estimate = -0.31 , $p<.05$), which was the reverse of the DIF effect on the general depressive factor (fully standardized parameter estimate = $+0.36$). The effect was reversed for education level; with each additional level of education, older adults experience a 0.31 S.D. increase in “lack of positive affect” but are less likely to admit to having the “psychosomatic” factor. Compared to Model 3a, there were weaker effects for stroke on the general depression factor in Model 3b. This is explained by a significant relationship between stroke and the *no energy* symptom (Table 5). Between models 3a and 3b, there is a 0.23 SD unit lower level of the “psychosomatic” factor among those with heart condition, an effect size reduction of about one-half. This is explained by a significant association of heart conditions with *restless sleep* and *no energy* (Table 5). After controlling for sociodemographic characteristics, there was also a decrease in the “psychosomatic” factor among those with diabetes. However, the overall depression level remained relatively unchanged for the diabetes group between Models 3a and 3b. The most notable difference was that the lung diseases no longer had an effect on the “lack of positive affect” factor in Model 3b, while there were significant and positive effects in Models 2a, 2b, and 3a. This is because lung diseases had a specific association with the no energy symptom (Table 5).

Upon adding in DIF effects from Models 2b to Model 3b, we see stronger effects for female and education level on general and specific factors. There was no difference in the mean level of depression between Blacks and Whites in Models 1 and 3b, implying that the greater level of general depression suggested by Model 2b for Blacks when compared to Whites is accounted for by DIF. Similar differences are noted in the effect between Blacks and Whites on the specific factors. Compared to Models 1 and 2b, the magnitude of the group difference was about the same as Model 3b, which were all between Cohen's 'trivial' and 'small' effects (50).

The specific “psychosomatic” factor, indicated by symptoms capturing *anergia* and *sleep* disturbance, was only related to health conditions. No significant associations with sociodemographic background variables were detected in the iterative model-building procedure. Persons with any of the studied health conditions had higher levels on the “psychosomatic” factor.

Table 5 shows the pattern of DIF effects of each covariate on a specific symptom expressed as odds ratios. Based on the magnitude rule of Cole and colleagues (41) for examining items with meaningful DIF, we found five items with measurement bias attributable to race/ethnic differences. When compared to Whites, Blacks were more likely to endorse the *depressed* item (odds ratio, OR =3.49, 95% confidence interval CI=[2.43, 6.02]) and the *everything was an effort* symptom (OR=4.13, 95% CI=[3.21, 5.30]); but were less likely to endorse the *not enjoy life* item (OR=0.41, 95% CI=[0.23, 0.74]).

Latinos, compared to Whites, were also more likely to endorse the *everything was an effort* item (OR=2.09, CI=[1.16, 3.76]), but had a lower likelihood of endorsing the symptoms of *no energy* (OR=0.44, CI=[0.29, 0.67]) and *not happy* (OR=0.41, CI=[0.20, 0.84]). The MIMIC-model-implied mean level of depression overestimates the proportional odds of endorsing the *everything was an effort* item. We note that throughout the analyses, the *no energy* item consistently misbehaved. For example, in the four-factor EFA, it splits off and defines a singleton fourth factor. In Table 3, it has the lowest loading on the general factor. In Table 5, virtually all the exogenous variables have significant DIF effects on this item.

Discussion

Our goals were to examine the factor structure of the HRS/CES-D in a large US population sample of older adults while examining instances of measurement noninvariance found on both the depressive factors and items due to health conditions and sociodemographic characteristics. There are two aspects of examining measurement noninvariance due to chronic illnesses: (1) examining the DIF effects of chronic illnesses separately and with sociodemographic characteristics on the individual CES-D items, consistent with standard approaches to DIF assessment; and (2) examining DIF due to the chronic illnesses only, as well as with sociodemographic characteristics on the specific factors of the CES-D, which is a change to the bifactor model.

Without significantly meaningful DIF effects of the covariates on the depressive symptom items, DIF is absent. Alternatively, when the covariates have substantial group effects on the specific factors, this implies that DIF is present for the whole item cluster, but a relatively more parsimonious model than a second-order factor model is derived that assumes the DIF effects are relatively constant for all items in the cluster. If substantive attention is focused on the “general” factor, effects on specific factors are of secondary concern from a measurement perspective.

Our hypothesis for the three specific factors of the HRS/CES-D was partially disconfirmed. We found evidence supporting essential unidimensionality for the HRS/CES-D using the parallel analysis and a bifactor model with two factors. Adding sociodemographic characteristics to the chronic health conditions for examining DIF in the bifactor model allowed us to find significant relationships between distinct HRS/CES-D factors and symptoms and specific sociodemographic characteristics of older adults.

Each chronic health condition was significantly associated with the “psychosomatic” factor. However, in most cases, when examining a bifactor model for HRS/CESD factors and symptoms, all of the health conditions positively predicted higher general depression, and the magnitudes of effect were relatively unaffected by relationships of health conditions to specific

factors and specific items. In the bifactor models with only chronic health conditions, the standardized coefficients were all significantly related to general depression and the “psychosomatic” factor. Only the standardized coefficients for stroke increased and high blood pressure decreased for predicting general depression when examining DIF. The bifactor model with DIF effects due to only chronic health conditions showed that heart conditions and diabetes significantly predicted higher levels of the psychosomatic factor, while all the other chronic conditions predicted lower levels of the “psychosomatic” factor.

Upon including sociodemographic effects in the bifactor model with DIF effects, the standardized coefficients increased for the general depressive factor only for lung diseases when compared to the bifactor model with DIF effects and only chronic health conditions. Each of the chronic health conditions, except for lung diseases, had lower standardized coefficients for the “psychosomatic” factor within the bifactor model with DIF effects. There was no longer a significant effect of lung diseases on the “lack of positive affect” factor in the presence of sociodemographic effects, while all other chronic health conditions had no significant DIF with or without the sociodemographic effects.

Among the chronic health conditions, the cardiovascular diseases and risk factors (including high blood pressure, heart condition, stroke, and diabetes) had similar effects on depressive symptoms across models. Based on the vascular depression hypothesis, these results could be interpreted as the psychosomatic depressive factor due to vascular risk factors possibly cause structural alterations in the brain that then lead to persistent mood disorders (51). However, the differences were very small. “Vascular depression” frames the development of depression in older adults as the growth of cerebrovascular lesions and abnormal changes in the subcortical gray matter due to one or more vascular risk factors (51–54). Older adults who suffer from vascular depression have been shown to display deep apathy, social isolation, interest reduction, cognitive impairment, and functional decline (55).

Under the bifactor model, each of the chronic health conditions except for high blood pressure significantly predicted higher general depression. Therefore, failure to recognize and to model distinct and orthogonal components of depressive factor and symptom endorsement patterns may lead to a spurious or overestimated association of health conditions with the severity of depressive factors and symptoms. We suggest that depressive symptom burden is overestimated among those with high blood pressure, heart conditions, stroke, and diabetes when the measurement device includes the “psychosomatic” factor of depression and crude raw score scoring. We note that the magnitude differences between the unidimensional and bifactor models are small between those with and without chronic health problems. Whether the measurement disturbances we find in the specific factors constitute bias is a matter of interpretation, best informed by clinical judgment and/or further analysis with clinically meaningful endpoints.

Under the unidimensional model, each of the sociodemographic characteristics, except for being Black, was significantly related to general depression. When the measurement device includes the “lack of positive affect” factor, depressive symptom burden was underestimated among those who are older and female. However, the inclusion of the “lack of positive affect” depressive item leads to overestimation of general depression among Latinos and those with higher education.

Although clinical implications can be drawn from our findings, this study was based on a non-clinical sample with self-reported measures of chronic health conditions. The major limitation of our study is that the nine-item HRS/CES-D is not a clinical instrument for the diagnosis of depression, but a measure of level of depressive symptomatology (56). Moreover, the HRS/CES-D with dichotomous response options is difficult to express on a scale similar to the full

20-item CES-D with four category response options, posing some challenge to the generalizability of data from the HRS/CES-D.

The cross-sectional analysis is another limitation that restricts our ability to determine whether differences in depressive factors and symptoms are attributable to prior experiences of depression, especially among those with lung diseases. In addition, the cross-sectional design does not establish these sociodemographic conditions and chronic health conditions as risk factors, but their co-occurrence with specific subsets of depressive factors allow us to identify possible risk factors for future longitudinal studies. Since the HRS is a longitudinal study, future studies using prior and future waves of data would eliminate this limitation. In terms of a methodological limitation, the MIMIC model cannot estimate different factor loadings for groups defined by exogenous variables, but multiple group approaches or alternative estimation procedures based on maximum likelihood estimating procedures can be used to address this limitation.

There is also limited information regarding the sub-clinical cognitive impairment of older adults, as similar analyses have shown the significant impact of cognitive functioning on depression in older adults (57). Furthermore, since many chronic health conditions have been linked with reduced cognitive functioning, which is a potential confounder that should be modeled in future analyses.

Despite these limitations, our study does have strengths. First, our use of the HRS and complex sampling weights in analysis renders our statistical results generalizable to older Americans. The advantage of testing measurement noninvariance in bifactor models is that we can examine measurement noninvariance on the general and specific depressive factors, as well as on the item level (58). The approach provides more information than a unidimensional or second-order factor model, because the bifactor model can show whether the source of DIF is differential group mean differences on the specific factors or the general factor. We gained the ability to control for effects of the background variables on general depression and determine whether there are effects on the specific factors not accounted for by the association of the general depression factor regressed on the background variables. This is the key to our interpretation of the DIF effects on the specific factors and the items as evidence of measurement noninvariance. Though this approach is not entirely new (59), it has been underutilized in medical research.

Our approach is relatively innovative and informative with respect to examining the CES-D and keeps with new directions in measurement noninvariance research (60). The importance of this translational research is that the CES-D has been used in both the clinical and research settings to help in early *screening* of non-institutionalized older adults (61). We anticipate building upon this study through future longitudinal analysis of the HRS.

We recommend that researchers who use the CES-D to quantify depressive syndromes attend to the possibility of measurement noninvariance with respect to sociodemographic and health characteristics. We document measurement noninvariance, but additional research is needed to determine whether this constitutes bias. To determine the extent to which group differences affect the underlying construct of depression, we compared the effects of the covariates in a model without measurement noninvariance (Model 2) and a model with measurement noninvariance (Model 3). Table 4 shows similar effects for both the sociodemographic and chronic health conditions on the general factor across each model, with the exception of the effect on Blacks in Model 2b. We conclude that the group differences in the general depressive factor were minimal. Furthermore, reduction in effects seen on general depression, as measured by the HRS/CES-D, is attributable to chronic medical conditions after controlling for the DIF effects on the “psychosomatic” symptom factor, which coincided with a minor change in the

effects on the general depression factor. This is of particular importance among older adults with high co-morbidity, as we show that age is associated with the largest increase in the effects on general depression after controlling for the DIF effects of both “psychosomatic” and “lack of positive affect” factors. The results from this study also confirm prior findings that older adults tend to report the “psychosomatic” factor, rather than other psychiatric factors of depression (10,11). A treatment study of depression in the context of chronic disease, such as heart disease, may show attenuated effects if the outcome measure includes the symptoms of the “psychosomatic” factor and this factor is endorsed preferentially due to sequelae associated with the chronic disease rather than an underlying psychopathology. On the other hand, studies examining non-pharmacologic treatments for chronic disease, such as exercise, may show spuriously high treatment-related gains if the exercise addresses physiologic symptoms (e.g., fatigue) but the underlying level of “dysphoria” remains unchanged among older adults.

Conclusion

The measurement of depression using symptom tools with “psychosomatic” and “lack of positive affect” items may be biased with respect to chronic health conditions and sociodemographic variables. However, the effects of such noninvariance on assessments of differences on the underlying depressive construct attributable to health conditions and sociodemographic factors are not large. Our model evaluates measurement equivalence, specifically in how depressive symptoms are correlated between the general depression and specific factors. Examining DIF using the bifactor model is potentially more parsimonious than the MIMIC model because the former allows for direct regression of multiple first order latent factors and items on covariates, while the latter allows for direct regression on only one latent factor and items simultaneously. Furthermore, the bifactor model may be less prone to type-I error (over-fitting) due to residual correlations among observed depression indicators. We also find the bifactor solution appealing because the effects of background variables on item dimensions have potentially clinical meaning, rather than apparently idiosyncratic effects on specific items.

The findings from this study can help guide research and clinical practice concerning depression etiology and detection of depressive symptomatology in late life through modern measurement methods. Specifically, researchers wishing to develop new depression assessment tools can avoid measurement bias from the outset by conducting IRT and DIF analyses on depression screening items using large nationally representative datasets. Based on the IRT and DIF analyses, researchers have the option of refining the wording of screening questions found with DIF in different groups of individuals, using mixed methods of quantitative and qualitative techniques.

Given the data on hand, we would not recommend dropping items to improve the measurement properties of the CES-D in preference to a model-based approach, as we have demonstrated here. Dropping items without superior items to replace them would reduce the reliability and validity of the CES-D. This study encourages the reader to consider replacing consistently problematic items (e.g., *no energy* item). It is notable that this item was not among Radloff et al.’s original CES-D item set (24) but was added to the HRS/CES-D (56).

The Institute of Medicine (IOM) (62) recommends that the National Institutes of Health (NIH) advance the development, testing, and refinement of appropriate measures for assessing race/ethnic health disparities in large population-based studies. The findings from this study are aligned with the goals of both the IOM and NHLBI and further support the importance not only of examining cardiovascular diseases and depression, but also of exploring the complex relationship among lung diseases and specific depressive symptoms common in older adults. We further recommend considering the issue of health disparities in overall depression

screening due to gender and race/ethnic differences in the endorsement of depressive symptoms. Measurement bias research is one of the first steps toward developing better detection and treatment methods for depression to improve medical, financial, and psychosocial outcomes for a substantial segment of the American population. With a population that is experiencing greater longevity and a growing proportion of minorities, the importance of understanding the etiology of depression is critical to reducing the progression of depressive symptoms and improving treatments for comorbidities with depression.

Acronyms used in text

CES-D, Center for Epidemiological Studies-Depression
 CVD, cardiovascular disease
 NHLBI, National Heart, Lung and Blood Institute
 NIH, National Institutes of Health
 IRT, Item Response Theory
 HRS, Health and Retirement Study
 AHEAD, Assets and Health Dynamics of the Oldest Old
 ISR, Institute for Social Research
 HRS/CES-D, modified CES-D used in HRS
 DIF, Differential Item Functioning
 EFA, exploratory factor analysis
 CFA, confirmatory factor analysis
 SEM, structural equation model
 MIMIC, Multiple Indicators Multiple Causes
 RMSEA, root mean square error of approximation
 CFI, comparative fit index
 SD, standard deviation
 CHF, congestive heart failure
 WLSMV, Weighted Least Squares Means and Variance
 DIFFTEST, Chi-square difference testing

Acknowledgment

This research was made possible through the National Institutes of Health (NIH)/National Institute on Aging (NIA) 5-T32 AG023480 award, NIH/NIA 5 R01 AG025308-02, NIH/NIA 5 R01 AG025308-02 Diversity Supplement, NIH/NIA P60AG008812, Harvard Medical School Livingston Fellowship, and the NARSAD Young Investigator Award. The authors are grateful for the research support of Eileen Crehan, Doug Tommet, editing support of Paul Guttry, and the suggestions of three anonymous reviewers.

References

1. Reynolds CF, Kupfer DJ. Depression and aging: A look to the future. *Psychiatric Services* 1999;50:1167–1172. [PubMed: 10478902]
2. National Heart Lung and Blood Institute Working Group (NHLBI). Bethesda, MD: National Institutes of Health; 2004. Assessment and Treatment of Depression in Patients with Cardiovascular Disease, Working Group Report; p. 1-17.
3. Davidson KW, Kupfer DJ, Bigger JT, Califf RM, Carney RM, Coyne JC, Czajkowski SM, Frank E, Frasure-Smith N, Freedland KE, Froelicher ES, Glassman AH, Katon WJ, Kaufmann PG, Kessler RC, Kraemer HC, Krishnan KR, Lesperance F, Rieckmann N, Sheps DS, Suls JM. Assessment and treatment of depression in patients with cardiovascular disease: National Heart, Lung, and Blood Institute Working Group Report. *Psychosom Med* 2006;68:645–650. [PubMed: 17012516]
4. Kleinman, A. New York: The Free Press; 1988. Rethinking psychiatry.
5. Arean, P.; Alvidrez, J.; Nery, R.; Estes, C.; Linkins, K. Recruitment and Retention of Older Minorities in Mental Health Services Research. In: Curry, L.; Jackson, J., editors. *The Science of Inclusion*:

- Recruiting and Retaining Racial and Ethnic Elders in Health Research. Washington, DC: The Gerontological Society of America; 2003. p. 17-25.
6. Robins, L.; Regier, D. New York: Free Press; 1991. *Psychiatric Disorders in America*.
 7. AHCPR. Rockville, MD: Agency for Health Care, Policy and Research; 1993. *Depression Guideline Panel: Detection and diagnosis. Clinical practice guideline, number 5., Depression in primary care: volume 1. Publication No. 93-0550.*
 8. U.S. Department of Health and Human Services: *Mental Health: Culture, Race, and Ethnicity—A Supplement to Mental Health: A Report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services; 2001.
 9. Bruce ML. Psychosocial Risk Factors for Depressive Disorders in Late Life. *Biological Psychiatry* 2002;52:175–184. [PubMed: 12182924]
 10. Jonas B, Lando J. Negative affect as a prospective risk factor for hypertension. *Psychosomatic Medicine* 2000;62:188–196. [PubMed: 10772396]
 11. Gallo JJ, Cooper-Patrick L, Lesikar S. Depressive symptoms of whites and African Americans aged 60 years and older. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 1998;53:277–286.
 12. Cole SR, Kawachi I, Maller SJ, Berkman LF. Test of item-response bias in the CES-D scale. Experience from the New Haven EPESE Study. *Journal of Clinical Epidemiology* 2000;53:285–289. [PubMed: 10760639]
 13. Gallo JJ, Anthony JC, Muthén BO. Age differences in the symptoms of depression: a latent trait analysis. *Journals of Gerontology, Psychological Sciences* 1994;49:251–264.
 14. Stommel M, Given BA, Given CW, Kalaian HA, Schulz R, McCorkle R. Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). *Psychiatry Research* 1993;49:239–250. [PubMed: 8177918]
 15. Callahan CM, Wolinsky FD. The effect of gender and race on the measurement properties of the CES-D in older adults. *Med Care* 1994;32:341–356. [PubMed: 8139299]
 16. Yang FM, Jones RN. Center for Epidemiologic Studies-Depression scale (CES-D) item response bias found with Mantel-Haenszel method successfully replicated using latent variable modeling. *Journal of Clinical Epidemiology* 2007;60:1195–1200. [PubMed: 17938063]
 17. Ross C, Mirowsky J. Components of depressed mood in married men and women: The Center for Epidemiologic Studies Depression Scale. *Am J Epidemiology* 1984;119:997–1004.
 18. Juster F, Suzman R. An overview of the Health and Retirement Study. *Journal of Human Resources* 1995;30:S7–S56.
 19. Oort F. Using restricted factor analysis to detect item bias. *Methodika* 1992;6:150–166.
 20. Reise S, Morizot J, Hays RD. The Role of the Bifactor Model in Resolving Dimensionality Issues in Health Outcomes Measures. *Quality of Life Research* 2007;16:19–31. [PubMed: 17479357]
 21. Gibbons RD, Bock RD, Hedeker D, Weiss DJ, Segawa E, Bhaumik DK, Kupfer DJ, Frank E, Grochocinski VJ, Stover A. Full-Information Item Bifactor Analysis of Graded Response Data *Applied Psychological Measurement*. 2007;31:4–19.
 22. Soldo B, Hurd M, Rodgers W, Wallace R. Asset and Health Dynamics Among the Oldest Old: an overview of the AHEAD Study. *J Gerontol B Psychol Sci Soc Sci* 1997;52:1–20. [PubMed: 9215354]
 23. HRS THaRS: The Health and Retirement Study. Design History 2003;Vol. 2003
 24. Radloff L. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;1:385–401.
 25. Long Foley K, Reed P, Mutran E, DeVellis R. Measurement adequacy of the CES-D among a sample of older African-Americans. *Psychiatry Research* 2002;109:61–69. [PubMed: 11850052]
 26. Hann D, Winter K, Jacobsen P. Measurement of depressive symptoms in cancer patients: evaluation of the Center for Epidemiological Studies Depression Scale (CES-D). *Journal of Psychosomatic Research* 1999;46:437–443. [PubMed: 10404478]
 27. Danao L, Padilla G, Johnson D. An English and Spanish quality of life measure for rheumatoid arthritis. *Arthritis Rheumatology* 2001;45:167–173.

28. Roberts R. Reliability of the CES-D Scale in different ethnic contexts. *Psychiatry Research* 1980;2:125–134. [PubMed: 6932058]
29. Kraemer HC, Blasey CM. Centring in regression analyses: a strategy to prevent errors in statistical inference. *Int J Methods Psychiatr Res* 2004;13:141–151. [PubMed: 15297898]
30. Lord F. A theory of test scores. *Psychometric monographs* 7:x 1952:84.
31. Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute of Educational Research; 1960.
32. Jones RN, Gallo JJ. Education and sex differences in the Mini Mental State Examination: Effects of differential item functioning. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 2002;57B:P548–P558.
33. Teresi JA, Kleinman M, Ocepek-Welikson K. Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Statistics in Medicine* 2000;19:1651–1683. [PubMed: 10844726]
34. Teresi J, Golden R, Cross P, Gurland B, Kleinman M, Wilder D. Item bias in cognitive screening measures: Comparisons of elderly white, Afro-American, Hispanic and high and low education subgroups. *Journal of Clinical Epidemiology* 1995;48:473–483. [PubMed: 7722601]
35. Drasgow F, Lissak RI. Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology* 1983;68:363–373.
36. Simms LJ, Gros DF, Watson D, O'Hara MW. Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depress Anxiety*. 2007
37. Embretson, SE.; Reise, SP. *Item Response Theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2000.
38. Muthén, BO. Latent variable modeling in heterogeneous populations; (1989, Los Angeles, California and Leuven, Belgium). *Meetings of Psychometric Society; Psychometrika*; 1989. p. 557-585.
39. Hauser, RM.; Goldberger, AS. *The Treatment of Unobservable Variables in Path Analysis*. In: Costner, HL., editor. *Sociological Methodology*. San Francisco: Jossey-Bass; 1971. p. 81-117.
40. Muthén, LK.; Muthén, BO. *Mplus Version 5.0*. Los Angeles: Muthén & Muthén; 1998–2008
41. Cole SR, Kawachi I, Maller SR, Berkman LF. Test of item-response bias in the CES-D scale: Experience from the New Haven EPESE Study. *Journal of Clinical Epidemiology* 2000;53:285–289. [PubMed: 10760639]
42. Browne, M.; Cudek, R. Alternative ways of assessing model fit. In: Bollen, K.; Long, J., editors. *Testing structural equation models*. Thousand Oaks, CA: Sage; 1993. p. 136-162.
43. Muthén, B.; Khoo, S-T.; Francis, D. Los Angeles: UCLA Graduate School of Education and Information Studies; 1998. Multi-stage analysis of sequential developmental processes to study reading progress: New methodological developments using general growth mixture modeling.
44. Bentler, P.; Chou, C. Practical issues in structural equation modeling. In: Long, J., editor. *Common problems/proper solutions: Avoiding error in quantitative research*. Newbury Park, CA: Sage; 1988.
45. Muthén, B. Los Angeles: Graduate School of Education and Information Studies; 1998. The development of heavy drinking and alcohol related problems from ages 18 to 37 in a U.S. national sample.
46. Hu L, Bentler P. Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecifications. *Psychological Methods* 1998;4:424–453.
47. Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990;107:238–246. [PubMed: 2320703]
48. Muthén B. Dichotomous factor analysis of symptom data. *Sociological Methods and Research* 1989;18:19–65.
49. Cohen, J. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988. *Statistical power analysis for the behavioral sciences*.
50. Cohen, J. New York: Academic Press; 1969. *Statistical power analysis for the behavioral sciences*.
51. Provinciali L, Coccia M. Post-stroke and vascular depression: a critical review. *Neurol Sci* 2002;22:417–428. [PubMed: 11976972]

52. Alexopoulos GS, Meyers BS, Young RC, Campbell S, Silbersweig D, Charlson M. 'Vascular depression' hypothesis. *Arch Gen Psychiatry* 1997;54:915–922. [PubMed: 9337771]
53. Krishnan KR, Hays JC, Blazer DG. MRI-defined vascular depression. *Am J Psychiatry* 1997;154:497–501. [PubMed: 9090336]
54. Bots ML, van Swieten JC, Breteler MM, de Jong PT, van Gijn J, Hofman A, Grobbee DE. Cerebral white matter lesions and atherosclerosis in the Rotterdam Study. *Lancet* 1993;341:1232–1237. [PubMed: 8098390]
55. Steffens DC, Krishnan KR. Structural neuroimaging and mood disorders: recent findings, implications for classification, and future directions. *Biol Psychiatry* 1998;43:705–712. [PubMed: 9606523]
56. Steffick, D. Ann Arbor, MI: Survey Research Center, University of Michigan; 2000. Documentation of affective functioning measures in the Health and Retirement Study. (available at <http://www.umich.edu/~hrswww/docs/userg/index.html>).
57. Mast BT. The impact of cognitive impairment on the phenomenology of geriatric depression. *American Journal of Geriatric Psychiatry* 2005;13:694–700. [PubMed: 16085785]
58. Chen F, Sousa K, West S. Testing Measurement Invariance of Second-Order Factor Models. *Structural Equation Modeling* 2005;12:471–492.
59. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods* 2004;9:466–491. [PubMed: 15598100]
60. Zumbo BD. Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly* 2007;4:223–233.
61. Himmelfarb S, Murrell SA. Reliability and validity of five mental health scales in older persons. *J Gerontol* 1983;38:333–339. [PubMed: 6841929]
62. Institute of Medicine: Examining the Health Disparities Research Plan of the National Institutes of Health. *Unfinished Business*. Washington, DC: The National Academies; 2006.

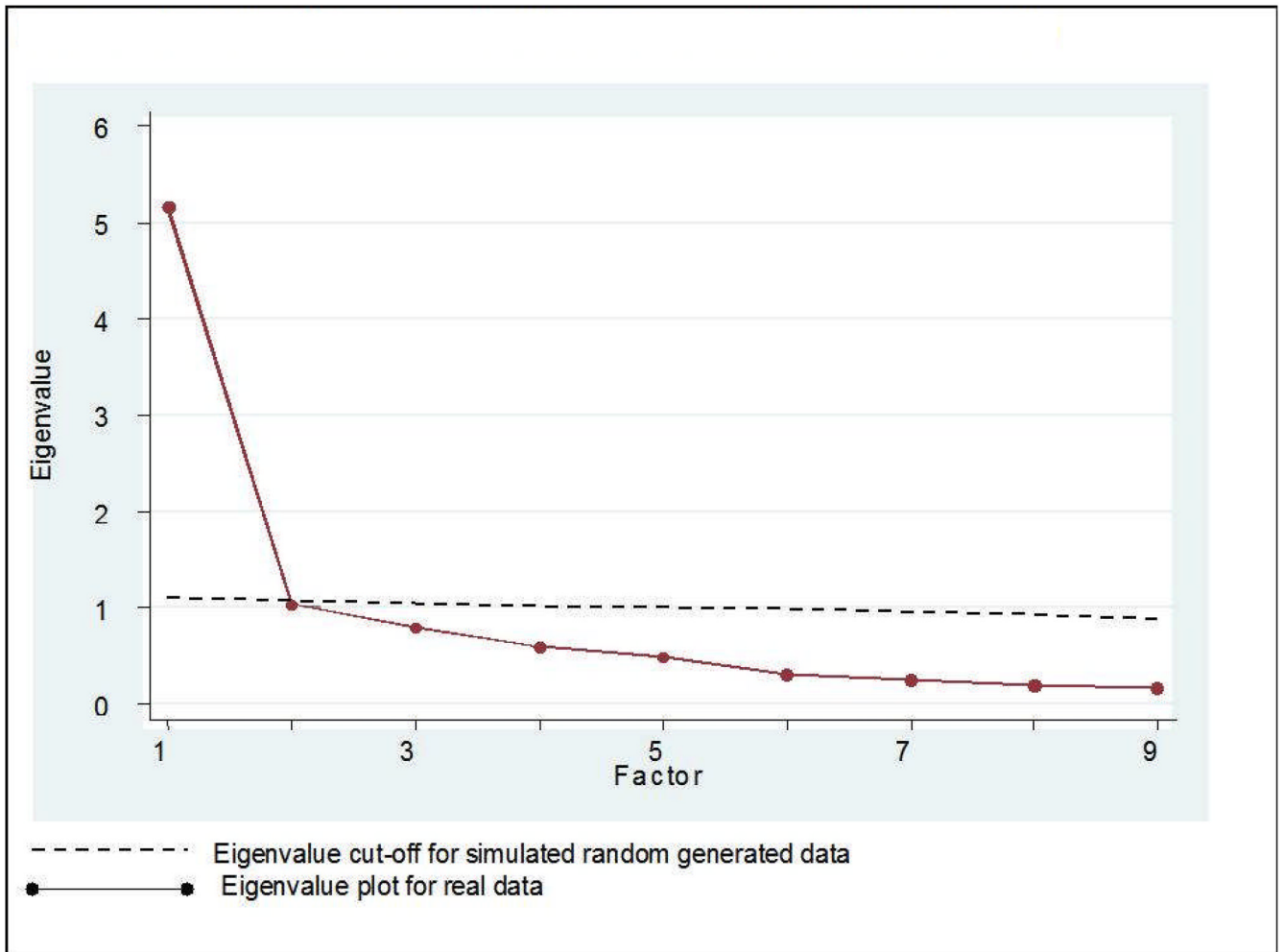


Figure 1. Profile plot of Real and Simulated Random Eigenvalues for CES-D (N=9,448)

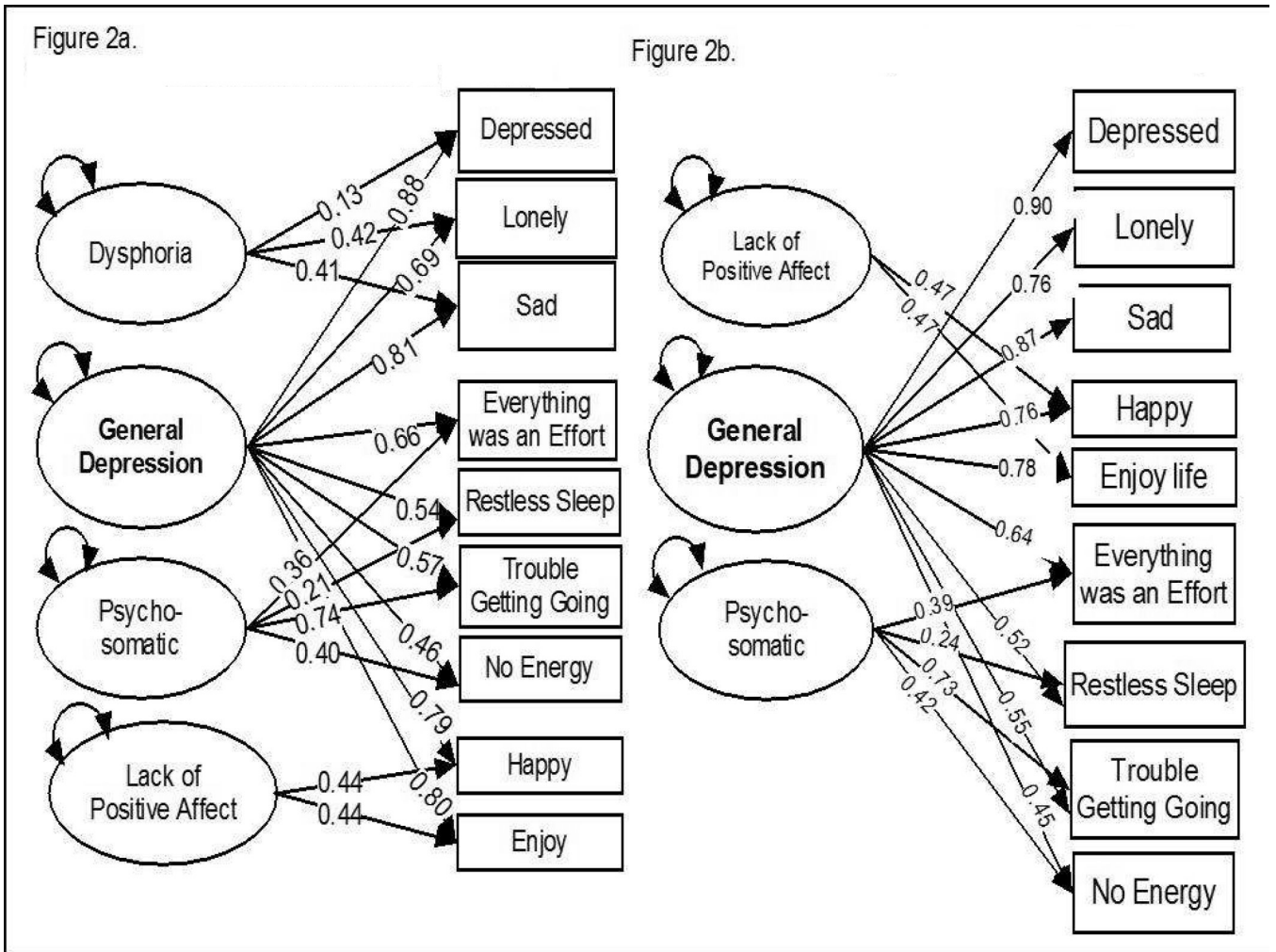


Figure 2.
 Figure 2a. Factor loadings for bifactor model with three specific factors from HRS/CES-D (n=9,448)
 Figure 2b. Factor loadings for bifactor model with two specific

Table 1
 Respondents Characteristics from the Health and Retirement Study (2004) (N=9,448)

Socio-demographic	N	%	Mean	SD
Age	9,448	100.0	74.4	7.2
Gender				
Male	3,820	40.5		
Female	5,628	59.5		
Highest Education Completed				
No degree	2,490	26.4		
Completed General Educational Development (GED)	417	4.4		
High school diploma	4,741	50.2		
Two year college degree	287	3.0		
Four year college degree	1,016	10.8		
Master's degree and above	497	5.3		
Race/Ethnicity				
White	7,958	84.2		
Black or African American	1,209	12.8		
Hispanic	281	3.0		
Chronic Health Conditions				
High Blood Pressure	5,797	61.4		
Heart Conditions	2,952	31.3		
Stroke	807	8.5		
Diabetes	1,889	20.0		
Lung Diseases	1,071	11.3		

Table 2
Exploratory factor analysis solutions for the Health and Retirement Study Center for Epidemiological Studies-Depression items
(N=9,448)

Factor-solutions for depressive symptoms	1-Factor	2-Factor	3-Factor	4-Factor
Depressed	0.86	0.77	0.14	0.67
Lonely	0.74	0.92	0.04	0.78
Sad	0.84	0.26	0.15	0.81
Trouble getting going	0.71	0.01	-0.04	0.27
No Energy (reversed scored)	0.56	0.75	-0.02	0.05
Everything was an effort	0.69	0.77	-0.10	0.07
Restless sleep	0.54	0.92	0.35	-0.22
Happy (reverse scored)	0.81	0.26	0.72	0.27
Enjoy Life (reverse scored)	0.84	0.01	0.77	0.21
Factor Correlations		F1	F2	F3
F1		1.00	0.56	1.00
F2		0.65	1.00	0.57
F3		1.00	0.56	1.00
F4		1.00	1.00	0.56
χ^2		1518.72	652.65	158.39
df		22	16	11
p-value		<.001	<.001	<.001
RMSEA		0.09	0.07	0.04
			169.84	6
			<.001	<.001
			0.05	0.05
			0.17	0.11
			0.01	0.07
			0.14	-0.04
			0.01	0.54
			-0.03	0.01
			-0.12	0.96
			0.37	0.50
			0.76	-0.10
			0.78	0.02
			F1	F2
			F1	F3
			0.54	0.59
			1.00	0.41
			1.00	1.00
			1.00	0.49
				1.00

Table 3
Standardized measurement slopes and mean differences for the latent depression measurement model for older adults in the Health and Retirement Study (2004) (N=9,448)

Factors	Unidimensional			Bifactor Model			Bifactor Model with Direct Effects			
	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b	Model 2a	Model 2b	Model 3a	Model 3b
General										
Depressed	0.88 *** [†]	0.88 *** [†]	0.90 *** [†]	0.91 *** [†]	0.90 *** [†]	0.92 *** [†]	0.88 *** [†]	0.86 *** [†]	0.86 *** [†]	0.88 *** [†]
Lonely	0.72 *** [†]	0.70 *** [†]	0.75 *** [†]	0.75 *** [†]	0.75 *** [†]	0.74 *** [†]	0.72 *** [†]	0.70 *** [†]	0.72 *** [†]	0.74 *** [†]
Sad	0.83 *** [†]	0.84 *** [†]	0.86 *** [†]	0.86 *** [†]	0.86 *** [†]	0.88 *** [†]	0.83 *** [†]	0.81 *** [†]	0.83 *** [†]	0.88 *** [†]
Everything was an effort	0.67 *** [†]	0.66 *** [†]	0.60 *** [†]	0.62 *** [†]	0.61 *** [†]	0.59 *** [†]	0.67 *** [†]	0.66 *** [†]	0.66 *** [†]	0.59 *** [†]
Restless sleep	0.54 *** [†]	0.54 *** [†]	0.50 *** [†]	0.50 *** [†]	0.50 *** [†]	0.50 *** [†]	0.54 *** [†]	0.53 *** [†]	0.53 *** [†]	0.50 *** [†]
Trouble getting going	0.65 *** [†]	0.66 *** [†]	0.52 *** [†]	0.53 *** [†]	0.52 *** [†]	0.53 *** [†]	0.65 *** [†]	0.64 *** [†]	0.64 *** [†]	0.53 *** [†]
No energy	0.49 *** [†]	0.54 *** [†]	0.38 *** [†]	0.38 *** [†]	0.38 *** [†]	0.41 *** [†]	0.49 *** [†]	0.48 *** [†]	0.48 *** [†]	0.41 *** [†]
Not happy	0.82 *** [†]	0.86 *** [†]	0.76 *** [†]	0.81 *** [†]	0.76 *** [†]	0.83 *** [†]	0.82 *** [†]	0.81 *** [†]	0.81 *** [†]	0.83 *** [†]
Not enjoy Life	0.83 *** [†]	0.87 *** [†]	0.77 *** [†]	0.82 *** [†]	0.77 *** [†]	0.84 *** [†]	0.83 *** [†]	0.82 *** [†]	0.82 *** [†]	0.84 *** [†]
Factor 2: Psychosomatic										
Everything was an effort			0.45 *** [†]	0.44 *** [†]	0.46 *** [†]	0.46 *** [†]	0.45 *** [†]	0.44 *** [†]	0.46 *** [†]	0.46 *** [†]
Restless sleep			0.26 *** [†]	0.26 *** [†]	0.26 *** [†]	0.26 *** [†]	0.26 *** [†]	0.26 *** [†]	0.26 *** [†]	0.26 *** [†]
Trouble getting going			0.62 *** [†]	0.59 *** [†]	0.62 *** [†]	0.65 *** [†]	0.62 *** [†]	0.59 *** [†]	0.71 *** [†]	0.65 *** [†]
No Energy			0.52 *** [†]	0.59 *** [†]	0.52 *** [†]	0.45 *** [†]	0.52 *** [†]	0.59 *** [†]	0.45 *** [†]	0.45 *** [†]
Factor 3: Lack of Positive affect										
Not Happy			0.48 *** [†]	0.49 *** [†]	0.48 *** [†]	0.49 *** [†]	0.48 *** [†]	0.49 *** [†]	0.48 *** [†]	0.49 *** [†]
Not Enjoy Life			0.48 *** [†]	0.49 *** [†]	0.48 *** [†]	0.49 *** [†]	0.48 *** [†]	0.49 *** [†]	0.48 *** [†]	0.49 *** [†]

* p<.05

[†] Fully standardized measurement slopes (factor loadings) with regard to mean and variance depressive symptoms regressed on latent depressive factors.

Table 4

Standardized estimates for chronic health and sociodemographic group differences in the latent depression structural model for older adults in the Health and Retirement Study (2004) (N=9,448).

Structural Model	Unidimensional			Bifactor Model			Bifactor Model with Direct Effects		
	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b			
General									
Depression									
Blood Pressure	0.11 **** [†]	0.06 * [†]	0.11 **** [†]	0.05 [†]	0.10 **** [†]	0.05 [†]			
Heart conditions	0.18 **** [†]	0.17 **** [†]	0.17 **** [†]	0.16 **** [†]	0.17 **** [†]	0.15 **** [†]			
Stroke	0.20 **** [†]	0.16 ** [†]	0.19 ** [†]	0.14 * [†]	0.21 **** [†]	0.15 ** [†]			
Diabetes	0.13 **** [†]	0.12 ** [†]	0.11 ** [†]	0.10 * [†]	0.11 ** [†]	0.11 ** [†]			
Lung diseases	0.39 **** [†]	0.36 **** [†]	0.35 **** [†]	0.32 **** [†]	0.35 **** [†]	0.36 **** [†]			
Age (centered)	0 //	0.09 **** [§]	0 //	0.07 **** [§]	0 //	0.10 **** [§]			
Female	0 //	0.23 **** [†]	0 //	0.29 **** [†]	0 //	0.36 **** [†]			
Educational level (centered)	0 //	-0.24 **** [§]	0 //	-0.23 **** [§]	0 //	-0.25 **** [§]			
Black	0 //	0.01	0 //	0.23 **** [†]	0 //	0.01			
Latino	0 //	0.20 * [†]	0 //	0.26 ** [†]	0 //	0.19 * [†]			
Factor 2: Psychosomatic									
Blood Pressure			0.27 **** [†]	0.26 **** [†]	0.22 **** [†]	0.21 **** [†]			
Heart conditions			0.30 **** [†]	0.30 **** [†]	0.47 **** [†]	0.24 **** [†]			
Stroke			0.40 **** [†]	0.40 **** [†]	0.32 **** [†]	0.31 **** [†]			
Diabetes			0.25 **** [†]	0.26 **** [†]	0.32 **** [†]	0.19 **** [†]			
Lung diseases			0.47 **** [†]	0.46 **** [†]	0.31 **** [†]	0.31 **** [†]			
Age (centered)			0 //	0.06 **** [†]	0 //	0.00			
Female			0 //	0.07	0 //	0.00			
Educational level (centered)			0 //	0.02	0 //	0.00			
Black			0 //	-0.12 * [†]	0 //	0.00			
Latino			0 //	-0.29 * [†]	0 //	0.00			
Factor 3: Lack of Positive Affect									
Blood Pressure			-0.04	0.02	0.00	0.00			
Heart conditions			-0.06	-0.05	0.00	0.00			
Stroke			0.02	0.08	0.00	0.00			

Structural Model	Unidimensional			Bifactor Model			Bifactor Model with Direct Effects		
	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b			
General									
Depression									
Diabetes			0.01	0.03	0.00	0.00			
Lung diseases			0.20 [‡]	0.18 [‡]	0.19 [‡]	0.00			
Age (centered)			0 ^{//}	-0.10 ^{***§}	0 ^{//}	-0.15 ^{***§}			
Female			0 ^{//}	-0.19 [‡]	0 ^{//}	-0.31 [‡]			
Educational level (centered)			0 ^{//}	0.26 ^{***§}	0 ^{//}	0.31 ^{***§}			
Black			0 ^{//}	-0.47 [‡]	0 ^{//}	0.00			
Latino			0 ^{//}	-0.40	0 ^{//}	0.00			

* p<.05

** p<.01

*** p<.001

[‡] Standardized regression coefficients with respect to the mean and variance of latent variables.

[§] Regression coefficients are fully standardized with regard to mean and variance of latent depressive factors; regressed on age and education variables.

^{//} Parameter constrained to zero.

Table 5

Results of DIF detection using the bifactor MIMIC model (Model 3): Odds Ratios between CES-D items and health and sociodemographic characteristics conditioned on latent depression in Health and Retirement Study (2004) (N=9,448).

Covariates	Depressed	Lonely	Sad	CES-D items Odds Ratio ^a (95% Confidence Interval)					
				Everything was an effort	Restless sleep	Trouble getting going	No energy	Not happy	Not enjoy life
Blood Pressure	1	1	1	1	1	1	1.19 (1.02, 1.39)	1	1
Heart conditions	1	1	1	1.23 (1.08, 1.40)	1	1	1.21 (1.05, 1.38)	1	1
Stroke	1	1	1	1	1	1	1.44 (1.09, 1.90)	1	1
Diabetes	1	1	1	1	1	1	1.29 (1.09, 1.51)	1	1
Lung diseases	1	1	1	1	1	1	1.44 (1.19, 1.74)	1	1
Age (centered)	0.98 (0.96, 1.00)	1.02 (1.01, 1.03)	1	0.98 (0.97, 0.98)	1	1	1.02 (1.01, 1.02)	1	1
Female	0.63 (0.49, 0.80)	1	1	0.80 (0.69, 0.94)	1.15 (0.99, 1.33)	1	1	1	1
Education level (centered)	1	1	1.28 (1.19, 1.38)	0.85 (0.80, 0.90)	1	1	1.20 (1.15, 1.26)	1	1
Black	3.49 (2.43, 6.02)	1.63 (1.34, 1.98)	1	4.13 (3.21, 5.30)	1	1	0.73 (0.61, 0.87)	1	0.41 (0.23, 0.74)
Latino	1	1	1	2.09 (1.16, 3.76)	1	0.54 (0.29, 1.02)	0.44 (0.29, 0.67)	0.41 (0.20, 0.84)	1

^aMIMIC direct effect estimates scaled as ORs that are exactly 1.00 imply direct effect estimates not estimated in the final fitted model.