Demography (2021) 58(4):1423–1443 Published online: 10 May 2021
DOI 10.1215/00703370-9304855 © 2021 The Authors
This is an open access article distributed under the terms of a Creative Commons license (CC BY-NC-ND 4.0).

The Comparability of Perceived Physical and Mental Health Measures Across Immigrants and Natives in the United States

Dina Maskileyson, Daniel Seddig, and Eldad Davidov

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

ABSTRACT The comparative study of perceived physical and mental health in general—and the comparative study of health between the native-born and immigrants, in particular—requires that the groups understand survey questions inquiring about their health in the same way and display similar response patterns. After all, observed differences in perceived health may not reflect true differences but rather cultural bias in the health measures. Research on cross-country measurement equivalence between immigrants and natives on self-reported health measures has received very limited attention to date, resulting in a growing demand for the validation of existing perceived health measures using samples of natives and immigrants and establishing measurement equivalence of health-related assessment tools. This study, therefore, aims to examine measurement equivalence of self-reported physical and mental health indicators between immigrants and natives in the United States. Using pooled data from the 2015–2017 IPUMS Health Surveys, we examine the cross-group measurement equivalence properties of five concepts that are measured by multiple indicators: (1) perceived limitations in activities of daily life; (2) self-reported disability; (3) perceived functional limitations; (4) perceived financial stress; and (5) nonspecific psychological distress. Furthermore, we examine the comparability of these data among respondents of different ethnoracial origins and from different regions of birth, who report few versus many years since migration, their age, gender, and the language used to respond to the interview (e.g., English vs. Spanish). We test for measurement equivalence using multigroup confirmatory factor analysis. The results reveal that health scales are comparable across the examined groups. This finding allows drawing meaningful conclusions about similarities and differences among natives and immigrants on measures of perceived health in these data.

KEYWORDS Immigration • Measurement equivalence • Multigroup confirmatory factor analysis (MGCFA) • Alignment optimization • IPUMS Health Surveys

Introduction

Differences in the prevalence of physical and mental health between immigrants and native-born populations are well established (Singh and Siahpush 2002). The litera-

ELECTRONIC SUPPLEMENTARY MATERIAL The online version of this article (https://doi.org/10.1215/00703370-9304855) contains supplementary material.

ture suggests that upon arrival in a host country, immigrants are likely to be healthier than the native-born population but that this health advantage dissipates with time (Boulogne et al. 2012). This phenomenon has been termed the *healthy immigrant effect*. Empirical evidence of the healthy immigrant effect has been documented for a wide variety of health indicators (e.g., obesity, hypertension, functional disabilities, chronic diseases, and mortality from almost all causes of death) (Akresh 2007) and in major immigrant-receiving countries (see Cunningham et al. [2008] for the United States, McDonald and Kennedy [2004] for Canada, and Martini et al. [2010] for Europe).

A review of the literature on the subject reveals that the use of the term *healthy immigrant effect* is quite vague. On the one hand, some studies refer to immigrants as a homogeneous population, arguing that upon arrival at their destination country, immigrants are in relatively better health, on average, than the members of the native population. On the other hand, empirical evidence has documented this advantage for only certain ethnoracial subgroups of immigrants (Huh et al. 2008). For example, findings on the initial health advantage of immigrants compared with their native-born counterparts (of the same ethnic origin) are less consistent for Asian immigrants than for Hispanic immigrants (Huh et al. 2008). Some studies found that Asian immigrants are likely to report poorer health status relative to their U.S.-born counterparts or the general U.S.-born population (e.g., Huh et al. 2008). Conversely, other studies found that Asians and Pacific Islanders are, on average, healthier than the general U.S. population (e.g., Frisbie et al. 2001).

Furthermore, it is unclear whether these observed differences reflect true differences or merely reflect cultural bias in the measures (Snowden 2003). Indeed, most health indicators were initially developed and tested on culturally homogeneous groups (e.g., White native English-speaking populations) (Snowden 2003). Consequently, concepts and constructs may not be understood in the same way when applied to other cultures (Teresi 2006). Therefore, measurement nonequivalence might result in biased conclusions about differences and similarities in health indicators across different population groups (Hardy et al. 2014). For example, Burgard and Chen (2014) showed how nonequivalence can lead to an underestimation or overestimation of health inequality, and they argued that researchers must take into account the varying levels of presence, quality, comparability, validity, and reliability of health indicators. However, to the best of our knowledge, only a few studies have examined the comparability of specific self-reported health measures between immigrants and natives (see Perreira et al. [2005] and Klokgieters et al. [2021] for a depression scale; see Buchcik et al. [2017] for health-related quality of life). Moreover, no one has systematically examined the comparability of several multidimensional self-reported health measures between immigrants and natives in a single study.

Using data from the 2015–2017 U.S. IPUMS Health Surveys (Lynn et al. 2019), this study addresses this gap by analyzing the comparability of some commonly used physical and mental health indicators between immigrants and U.S.-born natives. Specifically, we aim to evaluate the cross-group comparability of measures of nonspecific psychological distress, perceived disability, perceived limitations in activities of daily living (ADLs), and perceived functional limitations between immigrants and U.S. natives among the adult population aged 18 to 85 years. Furthermore, we examine the comparability of these data among respondents (1) deriving from dif-

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

ferent ethnoracial origins and (2) different regions of birth, who report (3) different rates of elapsed time (years) since migration, as well as their (4) age, (5) gender, and (6) language of interview (English vs. Spanish). By addressing measurement issues in the assessment of health status, this study contributes to a better and more reliable understanding of health disparities in a comparative framework.

Theoretical Background

The Problem of Nonequivalent Measures in Health Inequality Research

Questions relating to the healthy immigrant effect are centrally positioned in the migration literature. However, the reasons underlying this phenomenon remain a topic for vigorous research and debate. The most common explanations for the immigrant health advantage are positive health selection (immigrant self-selection as well as selective immigration policies; Dean and Wilson 2010) and negative health selection of return migrants (the so-called salmon bias effect; Goldman et al. 2014). The prevailing explanations for changes in immigrant health status over time are related to acculturation (i.e., immigrants adopt the health-related behaviors of the native-born population) (Abraído-Lanza et al. 2005), inadequate access to healthcare services (Bustamante et al., 2012), and socioeconomic inequality (Viruell-Fuentes et al. 2012). Although researchers have repeatedly used these explanations to account for health disparities between natives and immigrants, measurement nonequivalence between immigrants and natives has remained largely understudied (Morris 2018). Indeed, individuals from different countries may perceive their health in different ways and may understand questions inquiring about their health differently. As a result, questionnaire items measuring their perceived health might not have the same meaning for immigrants and natives. Individuals from different countries may also differ in how they use subjective response categories because of their different cultural backgrounds (e.g., Grol-Prokopczyk et al. 2015). Thus, their self-reported health responses to survey items may not be comparable. Therefore, cross-group equivalence of measurements of a latent construct that measures physical or mental health is necessary to draw meaningful conclusions in research comparing immigrants and natives. Lack of measurement equivalence hampers comparisons because differences in mean scores or in regression coefficients across groups may be a result of systematic biases in responses across groups or a different understanding of the questions rather than reflect real differences. At the same time, real differences may exist even when findings show no differences in mean scores or regression coefficients across groups (Davidov, Cieciuch et al. 2015; Davidov, Meuleman et al. 2014).

To date, only a few studies have attempted to examine measurement equivalence in health inequality research. These studies systematically tested for comparability and reported errors stemming from nonequivalence (Burgard and Chen 2014). For example, using data on 11 European countries from the Survey of Health, Ageing and Retirement in Europe, Hardy and colleagues (2014) examined how respondents translate morbidity and disability into self-rated health (SRH), how national populations differ in SRH, and how normative and person-specific reporting styles shape SRH. They found that observed country differences in SRH reflect compositional

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423masklevyson.pdf by guest on 30 November 2023

differences, cultural differences in reporting styles, and differing perceptions of how health restricts typical activities. Grol-Prokopczyk et al. (2015) examined vignette equivalence and response consistency in several widely used vignettes: the health vignettes that are implemented in the World Health Organization (WHO) Study on Global Ageing and Adult Health (SAGE) and World Health Survey, as well as similar vignettes in the Health and Retirement Study (HRS). Their results revealed substantial violations of vignette equivalence both cross-nationally and across socioeconomic groups: members of different sociocultural groups appeared to interpret vignettes as depicting fundamentally different levels of health. The authors concluded that the evaluated anchoring vignettes do not fulfill their promise of providing interpersonally comparable measures of health.

Notably, most of these health studies that addressed issues of measurement cross-group equivalence focused on cross-country comparisons of health measures. Single-country studies presumed social or cultural homogeneity in terms of response patterns and, therefore, neglected the question of whether health measures used in these studies were equivalent across groups within the country (e.g., Morris 2018). However, bias related to nonequivalence may "apply to all forms of comparative research, including comparisons of groups within the same single-country data set" (Morris 2018:559). Measurement nonequivalence within countries might lead to biased estimates of the incidence of perceived health problems and, therefore, affect the development of healthcare policies and delivery of healthcare to particular groups (Skinner et al. 2001). Thus, the comparability of health measures must be considered not only in cross-country comparative studies but also within single countries (Morris 2018).

Comparability of Health Measures Between Immigrants and Natives

Studies have repeatedly shown that, on average, recent immigrant population groups (compared with native-born population groups) manifest substantially lower rates of smoking, drinking, obesity, and hypertension; have a lower proportion of disabled persons; report lower levels of chronic diseases; and have a lower risk of mortality from almost all causes of death (Akresh 2007). Furthermore, both aggregate (Singh and Siahpush 2002) and within-group (Finch and Vega 2003; Muennig and Fahs 2002) studies have found that upon arrival, immigrants enjoy better health than their native-born counterparts in the host country. Despite such consistent findings, several converging lines of evidence demonstrate that the experience and conceptualization of health as reported in perceived health measures may differ across immigrants and natives.

Previous studies have suggested that mental health, for example, is contextually based and culturally embedded (e.g., Kleinman 1986). Therefore, it is plausible that the scales used for identifying depression, for instance, may tap into a different construct across the groups being compared. Immigrant groups may report higher/lower depression levels not only because of the actual higher/lower incidence of depression but also because the group expresses psychopathology in a way not captured by measures developed from a conceptualization of depression applicable to majority group members of the host country (e.g., White native-born) (Vega and Rumbaut 1991). Specifically, Hispanic and Black people are more likely to somatize mental health

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

problems, reporting more physical symptoms of distress than European Americans (Guarnaccia et al. 1989). Asians do not differentiate between mental and physical well-being and identify a strong interpersonal element to depression (Kim 2002). Fillenbaum and colleagues (1990) examined seven cognitive screening or neuropsychological tests and compared them with clinical diagnoses. The authors reported that most measures, when adjusted for race and education, had lower specificities for Blacks than for Whites. They suggested that most measures were culturally or educationally biased. Similarly, Teresi et al. (2001) reviewed studies using differential item functioning and examining bias in direct cognitive assessment measures with respect to race/ethnicity and education. They found that item performance varied across groups that differed in terms of education, ethnicity, and race.

Despite the potential prevalence of measurement nonequivalence across groups within countries—in particular, across natives and immigrants—relatively little research has addressed the question of whether immigrants and natives perceive and respond to health measures equivalently (Buchcik et al. 2017; Klokgieters et al. 2021). For example, Perreira et al. (2005) tested the Center for Epidemiologic Studies Depression Scale (CES-D) for use in the multiethnic and foreign-born populations living in the United States. Their results demonstrated that the CES-D is not psychometrically equivalent across either race/ethnicity or immigrant generations. Perreira et al. argued that research using the CES-D might lead to erroneous conclusions about mental health disparities in the U.S. adolescent population. Klokgieters et al. (2021) also examined measurement invariance of the CES-D among people of Dutch, Moroccan, and Turkish origin in the Netherlands. In contrast to Perreira et al.'s (2005) study, they found that the four subscales of the CES-D were measurement-invariant among older adults of Dutch, Turkish, and Moroccan origin. Therefore, it was possible to compare the prevalence of depressive symptoms across these three ethnic groups. Buchcik et al. (2017) tested the reliability and construct validity of health-related quality of life across elderly Polish migrants, Turkish migrants, and German natives. They did not find rigorous and complete invariance and observed some differences between the groups. Therefore, comparing subscale scores of the health-related quality-of-life measure between different ethnoracial groups may be problematic.

Furthermore, important within-group differences or similarities among immigrants may be biased because of properties of the measurement scales rather than reflecting true differences or similarities. To the best of our knowledge, no one has yet tested whether and to what extent measurement equivalence exists across different subgroups of immigrants. For example, it is not clear whether recent immigrants share a similar understanding of the questions with immigrants who spent more time in the United States or U.S.-born natives, whether older and younger immigrants display measurement equivalence in their perceived health measures, or whether measurement equivalence applies for those who responded to the survey in English or Spanish. Indeed, studies providing a systematic analysis of comparability of the various health measures between immigrants and natives are lacking. Standard measures of health, developed largely for majority population samples, have been applied to immigrants without close attention to the potential impact of cultural differences on their response patterns. Therefore, further studies are needed to determine the validity of the health indicators for cross-cultural comparisons.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

The Current Study

Data

The data for the analysis were obtained from the 2015–2017 IPUMS Health Surveys (formerly Integrated Health Interview Surveys) (Lynn et al. 2019). These surveys produce a harmonized data and documentation set based on material originally included in the public use files of the U.S. National Health Interview Survey (NHIS), the leading source of information on the health of the U.S. population. The IPUMS Health Surveys include many variable features not available in the original NHIS public-use files (Davern et al. 2012). The IPUMS Health Surveys data set provides detailed information on health conditions, health status, health behaviors, healthcare utilization, and insurance coverage; the data set also contains information across a rich set of individual and household characteristics, including immigrant status, sex, age, and education. We restrict our sample to individuals age 18 and older who had records in the sample adult core in 2015–2017 and who were randomly selected to respond to the Adult Functioning and Disability (AFD) section of the questionnaire. The final analytical pooled sample comprises 206,202 individuals (169,658 U.S.-born natives and 36,544 first-generation immigrants). We distinguish among four ethnoracial groups: (1) non-Hispanic White (130,691 natives and 5,644 immigrants); (2) non-Hispanic Black (21,185 natives and 2,967 immigrants); (3) non-Hispanic Asian (3,286 natives and 9,564 immigrants); and (4) Hispanic (14,496 natives and 18,369 immigrants) respondents. We thus have eight groups for the analysis: four native and four immigrant groups. We also distinguish among the following additional categories of immigrants and natives: region of birth, years since migration, language of interview, gender, and age.

Variables

We focus on the following self-perceived physical and mental health measures: (1) limitation in ADLs; (2) functional limitations; (3) self-reported disability; (4) nonspecific psychological distress (Kessler Psychological Distress Scale [K6]); and (5) perceived financial stress. These scales represent commonly used measures of self-perceived physical and mental health (Wang and Kaushal 2019). Notably, ADLs, functional limitations, self-reported disability, and perceived financial stress are part of the International Classification of Functioning, Disability and Health (ICF), which provides a standard language and framework for the description of health status (World Health Organization 2001).

Limitations in ADLs and Functional Limitations

Studies of health status and quality of life frequently assess functional disability. The concept of functional disability distinguishes basic daily activities that are necessary to be able to function daily both as an individual and in the community from other major social roles, such as work disability or social interactions (Nagi 1976). We rely on five questions measuring ADLs on a 4-point scale ranging from 1 = no

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

difficulty to 4 = cannot do it at all, and on nine questions measuring functional limitations using a 5-point scale ranging from 1 = not at all difficult to 5 = cannot do it at all. Table 1 summarizes the five scales we examine, the questions used to measure each scale, response categories to each question, and means and standard deviations of the responses for U.S. natives and immigrants, respectively.

Self-reported Disability

The self-reported disability questions presented in the IPUMS Health Surveys use the World Health Survey's ICF as a conceptual framework. The questions focus on perceived disability defined as difficulties in functioning in any of the following domains: (1) hearing (deafness or severe hearing loss); (2) vision (blindness or severe visual impairment even with corrective lenses); (3) cognitive difficulty (cognitive impairment in memory, concentration, or decision-making); (4) physical difficulty (physical impairments causing difficulty in walking or climbing stairs); (5) mobility limitations (physical, mental, or emotional conditions lasting at least six months and inhibiting the unassisted performance of basic activities outside the home); and (6) personal care limitations (physical, mental, or emotional conditions lasting at least six months and inhibiting unassisted attending to own personal needs, such as bathing, dressing, or moving around inside the home). Information on perceived disability is an important component of health information because it shows how well an individual is able to function in daily areas of life in general. Along with traditional indicators of a population's health status, such as mortality and morbidity rates, disability has become important in measuring disease burden, evaluating the effectiveness of health interventions, and planning health policy (Üstün et al. 2010). The six questions in the scale are measured using a dichotomous scale with the response options 1 = no difficulty and 2 = difficulty in performing the behavior.

Nonspecific Psychological Distress

The K6 was developed by Kessler and colleagues (Kessler et al. 2002) as a measure of nonspecific psychological distress, using a 30-day reference period (Fleishman and Zuvekas 2007). It has been used in numerous studies of mental health, focusing on particular races or ethnic groups (Wang and Kaushal 2019). The sociological and broader social science literature has relied extensively on the K6 as a reliable measure of depressive symptoms (Kessler et al. 2002). The responses are measured on a 5-point scale ranging from 0 = none of the time to 4 = all the time.

Perceived Financial Stress

Among the array of chronic stressors that people may confront in their daily lives, perceived financial stress is one of the most pivotal (Pearlin et al. 1997). We rely on eight questions measuring financial stress on a 4-point scale ranging from 1 = not worried at all and 4 = very worried.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423masklieyson.pdf by guest on 30 November 2023

Immigrant Status and Ethnoracial Origin

Immigrant status pertains to individuals who were not born in the United States. To distinguish between ethnoracial groups, we include dummy variables indicating whether the respondent was non-Hispanic Black, non-Hispanic Asian, or Hispanic, with non-Hispanic White as the reference category.

Variables for Within-Group Comparisons

We include 10 regions of birth: (1) Mexico, Central America, Caribbean Islands; (2) South America; (3) Europe; (4) Russia (and former USSR areas); (5) Africa; (6) the Middle East; (7) the Indian subcontinent; (8) Asia; (9) Southeastern Asia; and (10) the United States (for native U.S.-born residents). We also divide immigrants into groups by the number of years since migration: (1) recent immigrants (up to 5 years in the United States); (2) immigrants living in the United States for 5–15 years; (3) immigrants living in the United States 15 or more years; and (4) U.S. native-born residents. In addition, we include language of interview (Spanish vs. English), respondent's age (50 or younger vs. older than 50), and gender (men vs. women) as categories across which we test for measurement invariance. (See section 1 of the online appendix for the descriptive statistics by group.)

Methods

Measurement invariance and *measurement equivalence* indicate "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn and McArdle 1992:117). In other words, measurement invariance guarantees that the underlying latent variables measure the same phenomenon in different groups or at different times. Measurement equivalence may be tested by different methods, but the most common is multigroup confirmatory factor analysis (MGCFA; Reise et al. 1993). MGCFA analyzes unobserved (latent) constructs that are reflected by multiple observed (manifest) indicators. The relation between each indicator and its corresponding latent construct is expressed as a factor loading parameter. Furthermore, intercept parameters indicate the expected value of an observed variable when the latent variable is 0. With MGCFA, statistical analyses can compare the properties of a measure in two or more groups.

The cross-cultural methodological literature considers several levels of measurement equivalence (Davidov et al. 2014). *Configural invariance* refers to equivalence of the measurement structure and indicates that each latent variable (factor) has the same set of indicators in each group, the model fits the data well in each group, all factor loadings are substantial, and correlations among factors are less than 1. Configural invariance does not imply, however, that the meaning of a latent variable is the same across groups. *Metric invariance* additionally requires factor loadings to be equal across groups and is considered necessary to be able to compare factor covariances and unstandardized regression coefficients meaningfully. Finally, *scalar measurement invariance* additionally requires that intercepts of the indicators are equal

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

across groups. When the analyses demonstrate scalar measurement invariance, we can confidently compare the means of the latent variables across groups.

The literature suggests that when measurement indicators have fewer than four or five categories, researchers should test for measurement invariance while treating the measures as categorical rather than continuous (e.g., Davidov et al. 2018). Scalar measurement invariance in this case requires that the thresholds linking the underlying latent response variables to the observed categorical variables be held equal across groups in addition to the intercepts of the latent response variables (see Millsap and Yun-Tein 2004). When this specification is supported empirically, we can confidently compare the means of the substantive latent variables across groups.

Software packages such as Mplus (Muthén and Muthén 1998–2017) provide various measures of fit that can be used to evaluate the model. These measures of fit include the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean residual (SRMR). We consider models with a CFI value higher than 0.90 and RMSEA and SRMR values lower than 0.08 as acceptable (West et al. 2012). We adopt a bottom-up analytical strategy that begins with the least-restrictive (no invariance) model. We then introduce restrictions on the measurement parameters (factor loadings, intercepts, and thresholds). If the model fit does not significantly decrease by imposing additional restrictions, the more-restrictive invariant model could be supported by the data. Traditional difference tests based on chi-square values are sensitive to sample size (Saris et al. 1987). We follow the recommendations of Rutkowski and Svetina (2017) for cutoff values of fit indices when testing measurement invariance with categorical indicators. Accordingly, differences between a metric and a configural model are considered irrelevant when the change in the CFI is smaller than 0.004 and the change in the RMSEA is smaller than 0.05. Differences between a scalar and a metric model are considered irrelevant when the difference in CFI is smaller than 0.004 and the difference in RMSEA is smaller than 0.01. However, we do not consider the recommendations for chi-square differences because of the very large sample size in this study.

We treat all question items as categorical because the number of categories is always five or fewer and because the univariate distributions of the responses to the question items are highly skewed in many cases (Rhemtulla et al. 2012). Indeed, a large proportion of the respondents reported that they did not suffer from any physical or mental health issues. We use the mean- and variance-adjusted weighted least squares (WLSMV) estimator (Muthén et al. 1997), which is implemented in the software package Mplus (Muthén and Muthén 1998–2017) to address the categorical character of the data. With weighted least squares, the model in question is fitted to a vector of variable thresholds and a matrix of polychoric correlations assuming multivariate normality of the underlying (latent) continuous response variables. The fit function minimizes the weighted sum of differences between the observed polychoric and model-implied correlations (i.e., the model residuals). WLSMV estimation does not require large sample sizes and produces unbiased and efficient estimates under a number of conditions (e.g., Li 2016). With no covariates in the estimated models, the treatment of missing data is equivalent to pairwise deletion (Asparouhov and Muthén 2010b).[1]

[1] Because of the skewness of the data, empty categories may pose a problem when a majority of respondents are not affected by particular health issues. This was the case for various items measuring limitations

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Results

Descriptive Results

Table 1 reveals significant variations across immigrants and natives in the self-reported health levels. Immigrants reported better physical health (i.e., perceived disability, ADLs, functional limitations) than native-born individuals, in line with the healthy immigrant effect hypothesis. However, with regard to mental health (i.e., nonspecific psychological distress and perceived financial stress), the findings are somewhat less consistent. Finally, immigrants tended to feel worse than natives on only two parameters of psychological distress (sadness and hopelessness), whereas they felt much more worried, on average, than natives on all items assessing financial distress. This pattern is similar across all ethnic groups of immigrants and natives.

Testing for Measurement Equivalence Across Native and Immigrant Ethnoracial Groups: MGCFA

Next, we present the results of the measurement equivalence testing using the MGCFA models performed for each of the five latent variables measuring perceived physical and mental health. First, we perform our analyses both in the overall sample and across the eight population groups described earlier. The models for each of the five latent variables—limitations in ADLs, functional limitations, self-reported disability, perceived financial stress, and nonspecific psychological distress—fit the data well both for the total sample as well as for the four immigrant (non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, Hispanic) and four native-born (non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, Hispanic) groups (i.e., CFI $\geq$ 0.90, RMSEA $\leq$ 0.08, SRMR $\leq$ 0.08). The standardized factor loadings are high and exceed 0.5 in almost all models and samples.[2]

Given the good fit of the models to all samples, measurement invariance testing is performed in the next step. Specifically, we impose equality constraints across eight samples on the measures consecutively (moving from a less-restrictive to a more-restrictive model, as described in the Methods section). At each step, we inspect whether the deterioration in model fit is too large to accept the more constrained model. Table 2 summarizes the results of the measurement invariance tests. As mentioned earlier, CFI $\geq$ 0.90, RMSEA $\leq$ 0.08, and SRMR $\leq$ 0.08 are considered acceptable (West et al. 2012).

The *configural* model for all constructs fits the data well: the fit indices for this model are acceptable across eight sample groups for the five health measures. For example, the fit indices for limitations in ADLs are as follows: CFI = 0.993 ($\geq$0.90);

in ADLs. Thus, we collapse response categories 3 and 4 (high difficulty in the activities) in this case. Moreover, empty cells appear in the bivariate table of the items, "Do you have serious difficulty dressing or bathing?" and "Do you have serious difficulty walking or climbing stairs?", measuring self-reported disability in the Black immigrants group. We thus exclude the former item.

[2] In five group-specific models, either the RMSEA or SRMR exceeds the threshold value (0.08). However, the remaining fit statistics indicate good fit in all models.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Table 1 Mean responses to the perceived health scales, by immigrant status and ethnoracial group

	White		Black		Asian		Hispanic		All	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
Limitations in Activities of Daily Living (1=no difficulty to 4=cannot do it at all)										
(1) Do you have difficulty walking or climbing steps?	1.28	1.21	1.36	1.16	1.14	1.14	1.22	1.22	1.29	1.19
	(0.65)	(0.59)	(0.73)	(0.50)	(0.46)	(0.46)	(0.62)	(0.59)	(0.66)	(0.55)
(2) Do you have difficulty walking 100 yards on level ground that would be about the length of one football field or one city block?	1.26	1.18	1.39	1.16	1.12	1.13	1.22	1.22	1.27	1.18
	(0.70)	(0.59)	(0.84)	(0.59)	(0.47)	(0.51)	(0.66)	(0.63)	(0.72)	(0.59)
(3) Do you have difficulty walking a third of a mile on level ground that would be the length of five football fields or five city blocks?	1.27	1.19	1.37	1.16	1.12	1.14	1.18	1.19	1.28	1.18
	(0.68)	(0.54)	(0.75)	(0.53)	(0.43)	(0.49)	(0.56)	(0.55)	(0.68)	(0.53)
(4) Do you have difficulty walking up or down 12 steps?	1.26	1.18	1.37	1.17	1.13	1.14	1.22	1.23	1.26	1.19
	(0.65)	(0.54)	(0.76)	(0.54)	(0.45)	(0.50)	(0.63)	(0.63)	(0.66)	(0.58)
(5) Do you have difficulty raising a 2-liter bottle of water or soda from waist to eye level?	1.08	1.10	1.10	1.07	1.04	1.07	1.08	1.10	1.08	1.09
	(0.37)	(0.43)	(0.41)	(0.36)	(0.26)	(0.37)	(0.39)	(0.44)	(0.38)	(0.41)
Self-reported Disability (1=no to 2=yes)										
(1) Do you have serious difficulty walking or climbing stairs?	1.10	1.09	1.15	1.06	1.06	1.06	1.06	1.07	1.10	1.07
	(0.31)	(0.29)	(0.35)	(0.24)	(0.24)	(0.23)	(0.25)	(0.26)	(0.31)	(0.26)
(2) Do you have serious difficulty dressing or bathing?	1.03	1.03	1.05	1.02	1.02	1.02	1.02	1.03	1.03	1.03
	(0.17)	(0.18)	(0.21)	(0.13)	(0.15)	(0.15)	(0.14)	(0.16)	(0.17)	(0.16)
(3) Do you have serious difficulty doing errands?	1.06	1.07	1.09	1.04	1.05	1.05	1.05	1.05	1.07	1.05
	(0.24)	(0.26)	(0.28)	(0.20)	(0.21)	(0.22)	(0.22)	(0.21)	(0.25)	(0.22)

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Table 1 (continued)

	White		Black		Asian		Hispanic		All	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
(4) Do you have serious difficulty concentrating or remembering?	1.06	1.06	1.08	1.03	1.04	1.04	1.05	1.04	1.06	1.04
	(0.24)	(0.23)	(0.27)	(0.17)	(0.20)	(0.19)	(0.23)	(0.21)	(0.24)	(0.20)
(5) Are you deaf or have serious difficulty hearing?	1.08	1.06	1.04	1.02	1.05	1.03	1.04	1.04	1.07	1.04
	(0.27)	(0.23)	(0.19)	(0.12)	(0.22)	(0.18)	(0.18)	(0.19)	(0.26)	(0.19)
(6) Are you blind or have serious difficulty seeing even when wearing glasses?	1.04	1.04	1.06	1.02	1.03	1.03	1.04	1.04	1.04	1.04
	(0.19)	(0.19)	(0.23)	(0.14)	(0.18)	(0.16)	(0.18)	(0.20)	(0.19)	(0.19)
Functional Limitations (1 = not at all difficult to 5 = cannot do it at all)										
By yourself, and without using any special equipment, how difficult is it for you to:										
(1) Lift or carry something as heavy as 10 pounds such as a full bag of groceries?	1.29	1.31	1.45	1.20	1.19	1.23	1.24	1.31	1.30	1.28
	(0.86)	(0.91)	(1.05)	(0.71)	(0.71)	(0.80)	(0.80)	(0.92)	(0.88)	(0.88)
(2) Walk up 10 steps without resting?	1.35	1.30	1.56	1.25	1.20	1.17	1.29	1.32	1.37	1.27
	(0.95)	(0.87)	(1.15)	(0.78)	(0.72)	(0.68)	(0.87)	(0.91)	(0.97)	(0.84)
(3) Sit for about 2 hours?	1.27	1.22	1.35	1.14	1.17	1.15	1.27	1.26	1.28	1.21
	(0.78)	(0.72)	(0.89)	(0.54)	(0.62)	(0.58)	(0.79)	(0.78)	(0.79)	(0.70)
(4) Stand or be on your feet for about 2 hours?	1.65	1.53	1.79	1.33	1.38	1.31	1.44	1.45	1.65	1.42
	(1.26)	(1.16)	(1.36)	(0.93)	(1.00)	(0.90)	(1.07)	(1.09)	(1.26)	(1.05)
(5) Stoop, bend, or kneel?	1.72	1.53	1.79	1.39	1.39	1.32	1.49	1.49	1.71	1.45
	(1.20)	(1.07)	(1.27)	(0.95)	(0.95)	(0.89)	(1.06)	(1.08)	(1.20)	(1.02)
(6) To walk 1/4 mile?	1.51	1.38	1.69	1.24	1.30	1.23	1.36	1.34	1.52	1.31
	(1.13)	(0.99)	(1.28)	(0.78)	(0.90)	(0.79)	(0.97)	(0.93)	(1.14)	(0.90)
(7) Go out to things like shopping, movies, or sporting events?	1.27	1.24	1.36	1.14	1.17	1.16	1.21	1.21	1.28	1.19
	(0.82)	(0.79)	(0.92)	(0.60)	(0.68)	(0.66)	(0.75)	(0.74)	(0.82)	(0.72)

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Table 1 (continued)

	White		Black		Asian		Hispanic		All	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
(8) Push or pull large objects like a living room chair?	1.43 (1.05)	1.37 (1.00)	1.60 (1.22)	1.28 (0.86)	1.25 (0.82)	1.28 (0.88)	1.31 (0.93)	1.37 (1.02)	1.44 (1.06)	1.34 (0.97)
(9) Participate in social activities such as visiting friends, attending clubs and meetings, going to parties?	1.22 (0.74)	1.21 (0.74)	1.30 (0.85)	1.12 (0.55)	1.16 (0.65)	1.13 (0.60)	1.18 (0.69)	1.18 (0.68)	1.23 (0.75)	1.17 (0.67)
Perceived Financial Stress (1 = not worried at all to 4 = very much worried)										
(1) How worried are you right now about not being able to maintain the standard of living you enjoy?	2.06 (1.01)	2.19 (1.03)	2.25 (1.08)	2.52 (1.08)	2.01 (0.95)	2.09 (0.95)	2.24 (1.06)	2.58 (1.08)	2.10 (1.03)	2.38 (1.06)
(2) How worried are you right now about not being able to pay medical costs of a serious illness or accident?	2.15 (1.09)	2.33 (1.12)	2.28 (1.14)	2.64 (1.14)	2.10 (1.00)	2.23 (1.07)	2.36 (1.14)	2.76 (1.15)	2.18 (1.10)	2.54 (1.15)
(3) How worried are you right now about not being able to pay your rent, mortgage, or other housing costs?	1.61 (0.88)	1.79 (0.98)	1.93 (1.04)	2.27 (1.09)	1.66 (0.85)	1.81 (0.93)	1.89 (1.01)	2.36 (1.12)	1.67 (0.92)	2.12 (1.08)
(4) How worried are you right now about not being able to make the minimum payments on your credit cards?	1.45 (0.78)	1.59 (0.89)	1.75 (0.99)	2.04 (1.09)	1.53 (0.77)	1.63 (0.86)	1.74 (0.98)	2.13 (1.11)	1.50 (0.83)	1.86 (1.02)
(5) How worried are you right now about not being able to pay medical costs for normal healthcare?	1.80 (0.98)	1.97 (1.03)	2.03 (1.08)	2.31 (1.11)	1.79 (0.91)	1.93 (0.97)	2.03 (1.07)	2.50 (1.15)	1.85 (1.00)	2.25 (1.11)
(6) How worried are you right now about not having enough money for retirement?	2.23 (1.07)	2.34 (1.09)	2.36 (1.14)	2.65 (1.13)	2.10 (0.99)	2.22 (1.03)	2.37 (1.11)	2.67 (1.12)	2.26 (1.08)	2.50 (1.11)

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskilyson.pdf by guest on 30 November 2023

Table 1 (continued)

	White		Black		Asian		Hispanic		All	
	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants	Natives	Immigrants
(7) How worried are you right now about not having enough money to pay for your children's college?	2.43 (1.13)	2.49 (1.16)	2.50 (1.22)	2.75 (1.20)	2.53 (1.05)	2.45 (1.13)	2.65 (1.16)	2.97 (1.11)	2.46 (1.15)	2.77 (1.15)
(8) How worried are you right now about not having enough to pay your normal monthly bills?	1.76 (0.95)	1.88 (0.99)	2.12 (1.08)	2.35 (1.08)	1.76 (0.89)	1.86 (0.93)	2.05 (1.04)	2.46 (1.10)	1.83 (0.98)	2.20 (1.08)
Nonspecific Psychological Distress (0=none of the time to 4=all the time)										
(1) During the past 30 days, how often did you feel that everything was an effort?	0.56 (1.01)	0.54 (0.99)	0.69 (1.17)	0.54 (1.06)	0.51 (0.94)	0.43 (0.90)	0.60 (1.08)	0.50 (1.01)	0.58 (1.04)	0.49 (0.98)
(2) During the past 30 days, how often did you feel hopeless?	0.24 (0.68)	0.28 (0.72)	0.26 (0.71)	0.20 (0.65)	0.20 (0.60)	0.18 (0.59)	0.29 (0.76)	0.32 (0.79)	0.24 (0.69)	0.27 (0.72)
(3) During the past 30 days, how often did you feel nervous?	0.68 (1.00)	0.72 (1.02)	0.54 (0.95)	0.44 (0.84)	0.59 (0.91)	0.45 (0.81)	0.68 (1.04)	0.54 (0.94)	0.66 (1.00)	0.54 (0.92)
(4) During the past 30 days, how often did you feel so sad that nothing could cheer you up?	0.39 (0.79)	0.48 (0.88)	0.49 (0.90)	0.46 (0.86)	0.33 (0.73)	0.42 (0.79)	0.48 (0.88)	0.50 (0.90)	0.41 (0.81)	0.48 (0.87)
(5) During the past 30 days, how often did you feel worthless?	0.20 (0.66)	0.18 (0.61)	0.19 (0.64)	0.12 (0.50)	0.16 (0.54)	0.15 (0.53)	0.23 (0.70)	0.21 (0.67)	0.20 (0.66)	0.18 (0.62)
Number of Cases	130,691	5,644	21,185	2,967	3,286	9,564	14,496	18,369	169,658	36,544

Note: Standard deviations are shown in parentheses.

Source: IPUMS Health Surveys, 2015–2017.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Table 2 Measurement invariance tests, by immigrant status and ethnoracial group

Latent Construct	Model	Chi-sq. Diff.	*df* Diff.	CFI	RMSEA	SRMR
Limitations in Activities of Daily Living	Configural			0.993	0.072	0.073
	Metric	155.675	56	0.994	0.060	0.074
	Scalar	191.845	56	0.994	0.053	0.075
Self-reported Disability	Configural			0.993	0.038	0.040
	Scalar	90.515	21	0.994	0.029	0.042
Functional Limitations	Configural			0.991	0.072	0.032
	Metric	753.477	77	0.991	0.066	0.033
	Scalar	1,052.350	245	0.993	0.047	0.033
Perceived Financial Stress	Configural			0.997	0.067	0.017
	Metric	1,749.309	49	0.997	0.063	0.019
	Scalar	2,324.327	105	0.996	0.052	0.021
Nonspecific Psychological Distress	Configural			0.997	0.049	0.012
	Metric	1,448.636	35	0.994	0.054	0.015
	Scalar	725.120	119	0.995	0.034	0.016

Notes: Chi-sq. Diff. = chi-square difference, *df* Diff. = difference of degrees of freedom, CFI = comparative fit index, RMSEA = root mean square error of approximation, and SRMR = standardized root mean residual. Chi-square differences are based on second-order adjusted test statistics (Asparouhov and Muthén 2010a), in which one model is tested against a less restrictive model (i.e., metric against configural and scalar against metric).

RMSEA = 0.072 (≤0.08); SRMR = 0.073 (≤0.08). These results indicate that the same model structure is given across the eight sample groups. Thus, configural invariance as a baseline model for each scale is established.

Given that configural invariance is supported, we can test for higher levels of invariance by first imposing equality constraints on the factor loadings across groups.[3] Differences between a configural and a *metric* model are considered irrelevant when the change in the CFI is smaller than 0.004 and the change in the RMSEA is smaller than 0.05 (Rutkowski and Svetina 2017). Table 2 indicates that for limitations in ADLs, for example, the change in CFI is only 0.001 (0.994 – 0.993), lower than 0.004; and the RMSEA is 0.012 (0.072 – 0.060), lower than 0.05. Similar results are obtained for the other four constructs. Next, we impose additional equality constraints on the item thresholds across groups. Differences between a metric and a *scalar* model are considered irrelevant when the difference in CFI is smaller than 0.004 and the difference in RMSEA is smaller than 0.01. Table 2 indicates that the differences are irrelevant for all constructs, suggesting that the health items have both equal semantic load and comparable latent means for immigrants and natives from four ethnoracial backgrounds. Thus, we conclude that our perceived physical and

[3] It is sometimes recommended that the thresholds and factor loadings be dealt with in tandem because both of them influence the respective item probability curves. We retain the metric invariance test to illustrate the procedure, which is a more fine-tuned way of discovering sources of noninvariance (e.g., Lubke and Muthén 2004). However, for binary items and WLSMV estimation, we skip the test of metric invariance because the metric model is not identified (Muthén and Muthén 1998–2017).

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

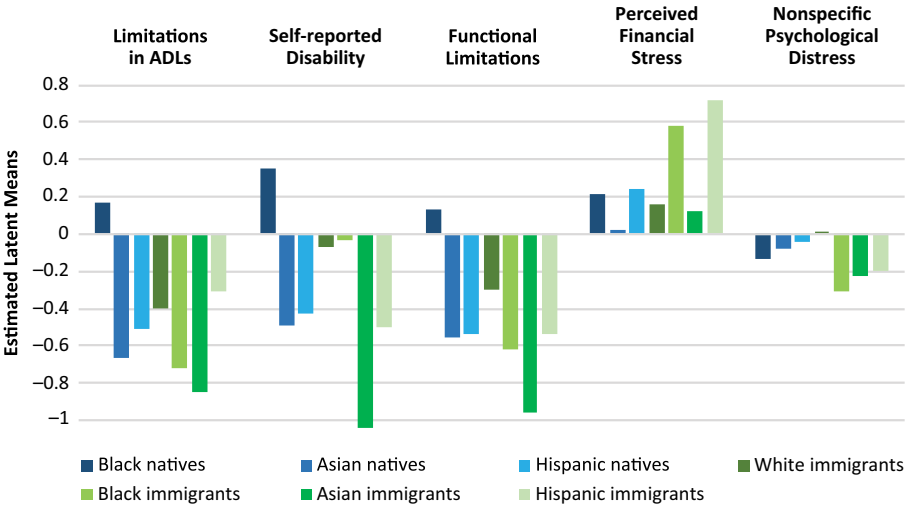

Fig. 1 Latent means of health variables by immigrant status and ethnoracial group. Scaling of the estimated means is arbitrary, and only mean differences can be interpreted meaningfully. The reference group, with a mean of zero, is White natives. Higher scores refer to higher degrees of reported limitations in activities of daily living (ADLs), higher degrees of self-reported disability, higher degrees of functional limitations, higher degrees of perceived financial stress, and higher degrees of psychological distress.

psychological health measures are comparable across the eight groups of immigrants and natives in the U.S. sample and that comparisons of latent mean scores across groups are valid for all constructs.

Next, we reexamine the measurement invariance properties of the scale, controlling for potential differences in the basic demographic composition of the subgroups by regressing the latent variables on age, gender, education, and family income while retesting for measurement invariance. We also control for the year of the survey. The measurement invariance results remain the same, indicating that differences in the demographic composition of the subgroups are no threat to the comparability of the measurements of physical and mental health across native and immigrant ethnoracial subgroups.

Figure 1 illustrates the differences in the latent variables measuring perceived physical and psychological health across the eight (four immigrant and four native) groups. The estimated latent means are based on the MGCFA scalar invariance models. The latent means for the non-Hispanic White natives group are set to 0; this is the reference group. Latent means of the other groups are estimated as a difference score between their means and the mean of the reference group. Higher scores on the bars refer to higher levels of reported limitations in ADLs, self-reported disability, functional limitations, perceived financial stress, and psychological distress.

Overall, Figure 1 demonstrates that all immigrant and native groups (with the exception of Black natives) displayed better physical and psychological health scores for four of the five measured constructs. Perceived financial stress was the only health measure for which the groups under study reported worse outcomes in comparison to White natives. Interestingly, Black natives reported worse health compared with White natives across all health indicators except psychological distress.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Testing for Measurement Invariance Within Native and Immigrant Ethnoracial Groups

We conduct additional measurement invariance tests to evaluate whether our measurement invariance findings hold within native and immigrant ethnoracial groups. We assess the comparability of the perceived health measures within the immigrant and native subgroups separately across the following categories: region of birth, language of interview, years since migration, gender, and age (see sections 2–6 of the online appendix). The measures again show the highest (scalar) level of invariance, thus indicating that comparisons within these groups are meaningful.[4] Latent mean differences across the tested groups are presented in Figures A1–A5 in the online appendix.

Discussion and Conclusions

Immigration and health researchers have long reported that at the time of their arrival in a host country, immigrants are likely to be healthier than native-born populations (Akresh 2007). However, the reasons underlying the initial health differences between immigrants and natives remain a topic for vigorous research and debate. Previous studies have focused on selectivity explanations and have not examined whether the utilized health indicators are comparable across population groups of different ethnoracial origin and immigrant groups. That is, no coherent evidence exists regarding whether the healthy immigrant effect is true or a result of culturally different perceptions of health and variability in the way survey questions are understood by respondents. Testing for measurement invariance of the scale before comparing scores across immigrant and ethnoracial groups is a prerequisite, however, to allow researchers to draw valid conclusions about similarities and differences. Indeed, differences in health may result from differences in response styles of various immigrant and ethnoracial groups or from their different understanding of the concept. At the same time, differences in response characteristics or in the understanding of the concept may conceal actual differences. Thus, in the current study, we aim to fill this gap and investigate measurement equivalence of five commonly used health indicators across different immigrant and ethnoracial groups in the United States. We utilize data from the 2015–2017 IPUMS Health Surveys (Lynn et al. 2019), which include multiple-item measures of five health indicators: nonspecific psychological distress, perceived financial stress, perceived disability, limitation in ADLs, and functional limitations. Our analyses involve two analytical methods: MGCFA and alignment optimization.

The results of the MGCFA analysis reveal support for measurement equivalence in the data. Thus, the findings imply that the health scales used in the IPUMS Health Surveys are comparable across most of the immigrant and ethnoracial groups participating in the surveys and that their latent means may be compared by immigrant status and across ethnoracial origin with confidence. Therefore, the differences across ethnoracial and immigrant groups are due to the healthy immigrant effect or some additional factors but not due to lack of invariance.

[4] Finally, we reanalyze all models using an unweighted least squares estimator (Forero et al. 2009) to test that the results do not rely on one specific estimation method. The global fit measures again suggest scalar measurement invariance across groups. See the online appendix for further details.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Indeed, when comparing our invariant latent scores, we reveal considerable differences in the level of mental and physical health. Specifically, our analysis affirms the robustness of the key finding of the literature: namely, that immigrants are healthier than U.S. natives and that this advantage declines with years of stay in the United States. We also find that Blacks and Hispanics are, on average, healthier than their non-Hispanic White peers.

Although this research has succeeded in establishing measurement invariance across the groups explored, it is not without limitations. First, our study examines both perceived physical and psychological health measures, but the literature discusses several other self-reported measures of health that may be subject to non-invariance across the compared groups. Second, our study is limited to the U.S. context. Although we established measurement equivalence for the data at hand, it may not necessarily hold for other time points, across more specific ethnic groups (e.g., from specific geographical areas in the world), or in other countries. Third, because the data do not contain information as to whether the native-born respondents are second-generation immigrants, we cannot examine whether measurement invariance holds across the first and second generation of immigrants compared with the native-born. Finally, the data are restricted to those aged 85 or younger and exclude institutionalized adults, and the interviews were conducted in English and Spanish only. Furthermore, we cannot distinguish among immigrant groups by reason for migration (e.g., to distinguish between refugees and economic immigrants). These factors may have affected the results and potential comparability with other data sets. Future studies could address these important issues by analyzing data for other countries, covering a diverse range of immigrant groups at different points in time, and using additional scales measuring respondents' health.

Notwithstanding these limitations, the current study is one of the first to assess the measurement comparability of various measures of health across several ethnoracial and immigrant groups systematically using state-of-the-art empirical methods. Researchers interested in comparing the health measures across ethnoracial and immigrant groups in different countries may apply similar procedures to examine whether their concepts of interest display a sufficient level of equivalence across these groups. We hope that the current study will thus help researchers in their endeavor to conduct a meaningful comparative study of immigrants' health. ■

Acknowledgments The authors would like to thank Lisa Trierweiler for the English proof of the manuscript prior to its acceptance, and Laura Tesch, Teresa Artman, and Bethany Sage Curtis for copy-editing after its acceptance. We also extend our gratitude to anonymous reviewers and to the editors of *Demography* for helpful comments. Eldad Davidov would like to thank the University of Zurich Research Priority Program Social Networks for support during work on this study.

References

Abraído-Lanza, A. F., Chao, M. T., & Gates, C. Y. (2005). Acculturation and cancer screening among Latinas: Results from the National Health Interview Survey. *Annals of Behavioral Medicine, 29,* 22–28.

Akresh, I. R. (2007). Dietary assimilation and health among Hispanic immigrants to the United States. *Journal of Health and Social Behavior, 48,* 404–417.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Asparouhov, T., & Muthén, B. (2010a). *Simple second order chi-square correction* (Mplus Version 6, technical appendix). Retrieved from https://www.statmodel.com/download/WLSMV_new_chi21.pdf

Asparouhov, T., & Muthén, B. (2010b). *Weighted least squares estimation with missing data* (Mplus Technical Appendix, 2010). Retrieved from https://www.statmodel.com/download/GstrucMissingRevision.pdf

Boulogne, R., Jougla, E., Breem, Y., Kunst, A. E., & Rey, G. (2012). Mortality differences between the foreign-born and locally-born population in France (2004–2007). *Social Science & Medicine, 74,* 1213–1223.

Buchcik, J., Westenhöfer, J., Fleming, M., & Martin, C. R. (2017). Health-related quality of life (HRQoL) among elderly Turkish and Polish migrants and German natives: The role of age, gender, income, discrimination and social support. In I. Muenstermann (Ed.), *People's movements in the 21st century—Risks, challenges and benefits* (pp. 55–75). Rijeka, Croatia: InTech.

Burgard, S. A., & Chen, P. V. (2014). Challenges of health measurement in studies of health disparities. *Social Science & Medicine, 106,* 143–150.

Bustamante, A. V., Fang, H., Garza, J., Carter-Pokras, O., Wallace, S. P., Rizzo, J. A., & Ortega, A. N. (2012). Variations in healthcare access and utilization among Mexican immigrants: The role of documentation status. *Journal of Immigrant and Minority Health, 14,* 146–155.

Cunningham, A. S., Ruben, J. D., & Venkat Narayan, K. M. (2008). Health of foreign-born people in the United States: A review. *Health & Place, 14,* 623–635.

Davern, M., Blewett, L. A., Lee, B., Boudreaux, M., & King, M. L. (2012). Use of the integrated health interview series: Trends in medical provider utilization (1972–2008). *Epidemiologic Perspectives & Innovations, 9,* 2.

Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., & Hausherr, M. (2015). The comparability of measurements of attitudes toward immigration in the European Social Survey: Exact versus approximate measurement equivalence. *Public Opinion Quarterly, 79*(S1), 244–266.

Davidov, E., Datler, G., Schmidt, P., & Schwartz, S. H. (2018). Testing the invariance of values in the Benelux countries with the European Social Survey: Accounting for ordinality. In E. Davidov, P. Schmidt, J. Billiet, & B. Meuleman (Eds.), *Cross-cultural analysis: Methods and applications* (2nd ed., pp. 157–179). New York, NY: Routledge.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40,* 55–75.

Dean, J. A., & Wilson, K. (2010). "My health has improved because I always have everything I need here . . .": A qualitative exploration of health improvement and decline among immigrants. *Social Science & Medicine, 70,* 1219–1228.

Fillenbaum, G., Heyman, A., Williams, K., Prosnitz, B., & Burchett, B. (1990). Sensitivity and specificity of standardized screens of cognitive impairment and dementia among elderly Black and White community residents. *Journal of Clinical Epidemiology, 43,* 651–660.

Finch, B. K., & Vega, W. A. (2003). Acculturation stress, social support, and self-rated health among Latinos in California. *Journal of Immigrant Health, 5,* 109–117.

Fleishman, J. A., & Zuvekas, S. H. (2007). Global self-rated mental health: Associations with other mental health measures and with role functioning. *Medical Care, 45,* 602–609.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16,* 625–641.

Frisbie, W. P., Cho, Y., & Hummer, R. A. (2001). Immigration and the health of Asian and Pacific Islander adults in the United States. *American Journal of Epidemiology, 153,* 372–380.

Goldman, N., Pebley, A. R., Creighton, M. J., Teruel, G. M., Rubalcava, L. N., & Chung, C. (2014). The consequences of migration to the United States for short-term changes in the health of Mexican immigrants. *Demography, 51,* 1159–1173.

Grol-Prokopczyk, H., Verdes-Tennant, E., McEniry, M., & Ispány, M. (2015). Promises and pitfalls of anchoring vignettes in health survey research. *Demography, 52,* 1703–1728.

Guarnaccia, P. J., DeLaCancela, V., & Carrillo, E. (1989). The multiple meanings of *ataques de nervios* in the Latino community. *Medical Anthropology, 11,* 47–62.

Hardy, M. A., Acciai, F., & Reyes, A. M. (2014). How health conditions translate into self-ratings: A comparative study of older adults across Europe. *Journal of Health and Social Behavior, 55,* 320–341.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18,* 117–144.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Huh, J., Prause, J. A., & Dooley, C. D. (2008). The impact of nativity on chronic diseases, self-rated health and comorbidity status of Asian and Hispanic immigrants. *Journal of Immigrant and Minority Health, 10,* 103–118.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., . . . Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32,* 959–976.

Kim, M. T. (2002). Measuring depression in Korean Americans: Development of the Kim depression scale for Korean Americans. *Journal of Transcultural Nursing, 13,* 109–117.

Kleinman, A. (1986). *Social origins of distress and disease*. New Haven, CT: Yale University Press.

Klokgieters, S., Mokkink, L., Galenkamp, H., Beekman, A., & Comijs, H. (2021). Use of CES-D among 56–66 year old people of Dutch, Moroccan and Turkish origin: Measurement invariance and mean differences between the groups. *Current Psychology, 40,* 711–718.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48,* 936–949.

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling, 11,* 514–534.

Lynn, A. B., Rivera Drew, J. A., King, M. L., & Williams, K. C. W. (2019). *IPUMS Health Surveys: National Health Interview Survey, Version 6.4* [Data set]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D070.V6.4

Martini, A., Chellini, E., & Sala, A. (2010). Mortality in immigrants in Tuscany. *Epidemiologia e Prevenzione*, *35*(5–6), 275–281.

McDonald, J. T., & Kennedy, S. (2004). Insights into the 'healthy immigrant effect': Health status and health service use of immigrants to Canada. *Social Science & Medicine, 59,* 1613–1627.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39,* 479–515.

Morris, K. A. (2018). Measurement equivalence: A glossary for comparative population health research. *Journal of Epidemiological Community Health, 72,* 559–563.

Muennig, P., & Fahs, M. C. (2002). Health status and hospital utilization of recent immigrants to New York City. *Preventive Medicine, 35,* 225–231.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (Unpublished technical report). Retrieved from https://www.statmodel.com/download/Article_075.pdf

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Nagi, S. Z. (1976). An epidemiology of disability among adults in the United States. *Milbank Memorial Fund Quarterly: Health and Society, 54,* 439–467.

Pearlin, L. I., Aneshensel, C. S., & LeBlanc, A. J. (1997). The forms and mechanisms of stress proliferation: The case of AIDS caregivers. *Journal of Health and Social Behavior, 38,* 223–236.

Perreira, K. M., Deeb-Sossa, N., Harris, K. M., & Bollen, K. (2005). What are we measuring? An evaluation of the CES-D across race/ethnicity and immigrant generation. *Social Forces, 83,* 1567–1601.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17,* 354–373.

Rutkowski, L., & Svetina, D. (2017). Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Applied Measurement in Education, 30,* 39–51.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology, 17,* 105–129.

Singh, G. K., & Siahpush, M. (2002). Ethnic-immigrant differentials in health, behaviors, morbidity, and cause-specific mortality in the United States: An analysis of two national data bases. *Human Biology, 74,* 83–109.

Skinner, J. H., Teresi, J. A., Holmes, D., Stahl, S. M., & Stewart, A. L. (2001). Measurement in older ethnically diverse populations. *Journal of Mental Health and Aging, 7*(1), 5–8.

Snowden, L. R. (2003). Bias in mental health assessment and intervention: Theory and evidence. *American Journal of Public Health, 93,* 239–243.

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023

Teresi, J., Abrams, R., Holmes, D., Ramirez, M., & Eimicke, J. (2001). Prevalence of depression and depression recognition in nursing homes. *Social Psychiatry and Psychiatric Epidemiology, 36,* 613–620.

Teresi, J. A. (2006). Different approaches to differential item functioning in health applications: Advantages, disadvantages and some neglected topics. *Medical Care, 44*(Suppl. 3-11), S152–S170.

Üstün, T. B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., . . . Pull, C. (2010). Developing the World Health Organization disability assessment schedule 2.0. *Bulletin of the World Health Organization, 88,* 815–823.

Vega, W. A., & Rumbaut, R. G. (1991). Ethnic minorities and mental health. *Annual Review of Sociology, 17,* 351–383.

Viruell-Fuentes, E. A., Miranda, P. Y., & Abdulrahim, S. (2012). More than culture: Structural racism, intersectionality theory, and immigrant health. *Social Science & Medicine, 75,* 2099–2106.

Wang, J. S. H., & Kaushal, N. (2019). Health and mental health effects of local immigration enforcement. *International Migration Review, 53,* 970–1001.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York, NY: The Guilford Press.

World Health Organization. (2001). *International classification of functioning, disability and health (ICF)*. Geneva, Switzerland: WHO.

Dina Maskileyson (corresponding author)
dmaskile@uni-koeln.de

Maskileyson • University of Cologne, Faculty of Management, Economics and Social Sciences, Institute of Sociology and Social Psychology, Cologne, Germany; https://orcid.org/0000-0002-5687-8336

Seddig • University of Cologne, Faculty of Management, Economics and Social Sciences, Institute of Sociology and Social Psychology, Cologne, Germany; https://orcid.org/0000-0003-1595-6948

Davidov • University of Cologne, Faculty of Management, Economics and Social Sciences, Institute of Sociology and Social Psychology, Cologne, Germany; Department of Sociology, and URPP Social Networks, University of Zurich, Zurich, Switzerland; https://orcid.org/0000-0002-3396-969X

Downloaded from http://read.dukeupress.edu/demography/article-pdf/58/4/1423/934165/1423maskileyson.pdf by guest on 30 November 2023