Published in final edited form as:

Qual Life Res. 2018 September ; 27(9): 2431–2441. doi:10.1007/s11136-018-1878-4.

Cross-national health comparisons using the Rasch model: findings from the 2012 US Health and Retirement Study and the 2012 Mexican Health and Aging Study

Ickpyo Hong¹, Timothy A. Reistetter¹, Carlos Díaz-Venegas², Alejandra Michaels-Obregon³, and Rebeca Wong³

¹Department of Occupational Therapy, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-1142, USA

²Max Planck Institute for Demographic Research, Rostock, Germany

³Sealy Center on Aging, University of Texas Medical Branch, Galveston, TX, USA

Abstract

Purpose—Cross-national comparisons of patterns of population aging have emerged as comparable national micro-data have become available. This study creates a metric using Rasch analysis and determines the health of American and Mexican older adult populations.

Methods—Secondary data analysis using representative samples aged 50 and older from 2012 U.S. Health and Retirement Study ($n = 20,554$); 2012 Mexican Health and Aging Study ($n = 14,448$). We developed a function measurement scale using Rasch analysis of 22 daily tasks and physical function questions. We tested psychometrics of the scale including factor analysis, fit statistics, internal consistency, and item difficulty. We investigated differences in function using multiple linear regression controlling for demographics. Lastly, we conducted subgroup analyses for chronic conditions.

Results—The created common metric demonstrated a unidimensional structure with good item fit, an acceptable precision (person reliability = 0.78), and an item difficulty hierarchy. The American adults appeared less functional than adults in Mexico ($\beta = -0.26$, $p < 0.0001$) and across two chronic conditions (arthritis, $\beta = -0.36$; lung problems, $\beta = -0.62$; all $p < 0.05$). However, American adults with stroke were more functional than Mexican adults ($\beta = 0.46$, $p = 0.047$).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11136-018-1878-4>) contains supplementary material, which is available to authorized users.

Compliance with ethical standards

Conflict of interest All author declares that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was exempted by the Institutional review boards (IRB) of the University of Texas Medical Branch because the research is a study of an existing data sets, the US Health and Retirement Study and the Mexican Health and Aging Study which are publicly available. The study data set is de-identified, such that subjects cannot be identified directly, or through identifiers linked to the subjects.

Conclusions—The Rasch model indicates that Mexican adults were more functional than Americans at the population level and across two chronic conditions (arthritis and lung problems). Future studies would need to elucidate other factors affecting the function differences between the two countries.

Keywords

Cross-cultural comparison; Rasch model; Disability; Health; Outcome measure

Introduction

Cross-national comparisons of patterns of population aging have emerged recently as comparable national micro-data have become available [1–3]. For instance, the World Health Organization Study on global AGEing and adult health [3] and the United States (U.S.) National Institute on Aging [4] have made efforts to harmonize population-based longitudinal aging studies to provide rigorous research opportunities for cross-national comparisons on aging. Currently, 14 national aging studies have been harmonized in 22 European countries, six Asian countries, and three American countries [1–7]. The significance of comparative physical function research lies in the ability to understand how populations exposed to different social and policy contexts are likely to age. Previous studies reported that gender [8] and age [9] influence disability progressions between Americans and Mexicans. For instance, Gerst-Emerson et al. [9] reported that U.S. adults are more likely than Mexican adults to be disabled or have an increased number of disabilities within 2-year transition. In addition, Díaz-Venegas et al. [8] reported that Mexican adults have different patterns of disability progression across genders compared to adults in developed countries, such as the US. However, few studies have been designed to compare the health of the two populations across various chronic conditions and socioeconomic factors.

A major challenge in international health comparisons is incompatible outcome measures across countries. For this reason, previous U.S. and Mexico disability comparison studies relied on estimations of the percentage or likelihood of having difficulties performing daily activities [9, 10]. While this approach provided an indication of disability differences between the populations, these disability comparison approaches have been criticized for using two different measurement scales. Recently, the 1-parameter Rasch model of item response theory (IRT) has been utilized to develop a common metric across 24 European countries [5], as well as the U.S. and England [6, 11]. While the research methodology is highly applicable for comparing the health of U.S. and Mexican populations, no attempts have been made to employ this common metric approach. Therefore, the objectives of this study are (1) to create a metric based on IRT approaches using function questions from the 2012 Health and Retirement Study in the U.S. (HRS) and the 2012 Mexican Health and Aging Study (MHAS) and (2) to determine whether the American adults are more functional than the Mexican adults.

Methods

Study data

We retrieved the American sample from the most recent wave 11 of the 2012 HRS and the Mexican sample from wave 3 of the 2012 MHAS [7, 12] because, while the HRS has been conducted every 2 years since 1992, the MHAS was conducted in 2001, 2003, and 2012. Both HRS and MHAS targeted nationally representative, longitudinal, population-based samples for the U.S. and Mexico comprised health information for individuals over age 50, including socioeconomic, disability, and aging information [1, 2]. The HRS has harmonized portions of the main survey with 30 other international aging surveys to provide opportunities for cross-national comparisons [1]. The harmonized contents between surveys include demographics, activities of daily living (ADL), instrumental ADL (IADL), physical function, cognition, depression, chronic conditions, and other comprehensive socioeconomic indicators [2].

We retrieved a total of 20,554 and 14,448 respondents aged 50 or older from the HRS and MHAS, respectively, who complete the ADL, IADL, and physical function questions. We utilized these self-report survey questions to create a common functionality metric between the two surveys [5, 6]. The Institutional Review Board of the University of Texas Medical Branch approved this study.

Variables

Outcome measure—We used six ADL (dressing, walking across a room, bathing, eating, getting in/out of bed, toileting), four IADL (preparing meals, shopping, taking medications, managing money), seven mobility (jogging 1 mile/1 km, walking 1 block, walking blocks, sitting for 2 h, getting up from a chair, climbing several flights of stairs, climbing one flight of stairs), and five physical function questions (stooping, reaching arms, pulling/pushing large objects, lifting weights of 10 lb/5 kg, picking up a dime) to create a common metric and estimate the functionality status of the U.S. and Mexico adult populations. The harmonized HRS and MHAS have dichotomized the response categories for the 22 function questions, including 1 (having difficulty) and 0 (having no difficulty). For ADLs and mobility tasks, those who said “cannot do” or “does not do” were assigned a value of 1 if they were unable to perform the activity and received help from someone. For IADLs, those who said “cannot do” or “does not do” were assigned a value of 1 (“yes, difficulty”) if they were unable to perform these activities because of a health-related problem [7, 13]. We reversed the response categories into 1 (having no difficulty) and 0 (having difficulties) so that a higher score would indicate more functional capacity.

Socioeconomic and health covariates—Since the HRS and MHAS have different response categories, we made them comparable by collapsing adjacent response categories. The rescaling and description of each variable of the HRS and MHAS were described in a study by Díaz-Venegas et al. [8] We dichotomized the female, single (living alone), insurance (having any health insurances), and employed (currently working) variables into 1 = yes or 0 = no. Age remained continuous. Monthly income was collapsed into four categories in the MHAS: class 1—having no income or debts, class 2—monthly income

between \$1 and \$359 (1–4999 Mexican Pesos), class 3—monthly income between \$360 and \$719 (5000–9999 Mexican Pesos), and class 4—monthly income over \$720 (over 10,000 Mexican Pesos at the average 2012 current exchange rates). In the HRS, the monthly income consists of the same four categories with different amounts of income, including class 1—having no income or debts, class 2—monthly income less than \$10,000, class 3—monthly income between \$10,000 and \$19,999, and class 4—monthly income over \$20,000. Body mass index (BMI) was used as an indication of obesity and it was calculated as weight in kilograms divided by height in meters squared [14]. Depression symptoms were measured by the Center for Epidemiological Studies Depression (CES-D) which consists of nine questions, including feeling depressed, feeling that everything is an effort, restless sleep, feeling happy, feeling lonely, feeling that life is enjoyable, feeling sad, feeling tired, and feeling energetic [15]. We reversed three questions (feeling happy, life is enjoyable, and feeling energetic) of the CES-D so that high scores reflect more severe depressive symptoms. Similarly, the self-rated health question consists of a 5-point rating and was reversed to 1 = poor to 5 = excellent. For cognition, we used the immediate and delayed verbal recall questions. Because the HRS has a 10-word recall item and the MHAS has an eight-word recall item, we converted the responses of the two questions into ratio of success for disability comparisons. Finally, self-reported chronic conditions were classified as 1 = yes or 0 = no for stroke, arthritis, diabetes, cancer, lung problems, heart problems, and hearing problems.

Statistical analysis

Statistical procedures consist of two phases: (1) testing and creating an ideal common metric between the two surveys and (2) comparing the level of functional status across samples by using the ideal item calibration (ruler) from the first step. We randomly separated each data set into two parts: (1) 2000 random respondents (1000 respondents in each data set) to test the unidimensionality assumption among the 22 test items and the full sample ($n = 35,002$) to create a common metric between two surveys; and (2) all 35,002 respondents for comparing the health status between the two populations (Fig. 1). Step I included evaluation of unidimensionality (factor analyses), Rasch analysis (fit statistics, precision, item difficulty hierarchy), and item co-calibrations (a common ruler) of the HRS and MHAS data. Step II includes (1) estimating separate person measures (level of function) for the HRS and MHAS samples using the co-calibrated item measure (e.g., the common ruler) and (2) estimating differences in the level of function between the two populations across chronic conditions while adjusting for covariates.

Descriptive statistics were conducted to examine differences in demographics and socioeconomic variables between the HRS and MHAS. SAS 9.4 software was used to manage the survey data and generate descriptive statistics [16].

Unidimensionality

Unidimensionality is a core assumption of the Rasch model when building a measurement [17, 18]. We examined the unidimensionality of the 22 function questions with the random sample ($n = 2,000$) using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) [17, 18]. We used the first 1000 random sample (HRS = 497, MHAS = 503) to

explore the measurement structure of the 22 function questions using EFA with Weighted Least Squares with Adjustments for the Mean and Variance (WLSMV) estimations [17, 19]. The unidimensionality assumption of the 22 function questions was determined by factor loadings and the goodness of fit statistics including comparative fit index ($CFI > 0.95$), Tucker–Lewis Index ($TLI > 0.95$), and root mean square error of approximation ($RMSEA < 0.06$) [17]. Using the remaining 1000 random sample ($HRS = 503$, $MHAS = 497$), we conducted CFA with a one-factor solution to confirm the identified factor structures obtained in the EFA. Mplus software version 7.4 was used to perform the factor analyses [19].

Rasch analysis

Once we obtained a set of items supporting the unidimensional assumption, we utilized a parametric Rasch model to create a common metric between the HRS and MHAS with the full respondents ($n = 35,002$). The purpose of this Rasch analysis was to create the optimal metric for estimating precise item calibrations for the 22 function questions to compare health status between the U.S. and Mexico populations. First, the item fit to the Rasch model was examined by infit (weighted), and outfit (un-weighted) mean square residual (MnSq). The ideal value of MnSq is 1.0, indicating that the observed and expected score are the same. In this study, the criteria for both infit and outfit statistics were a range of 0.40–1.60 for surveys [20, 21].

Next, we examined differential item functioning (DIF) for the 22 function status questions by survey (HRS vs. MHAS) to test if there were different response patterns between the two populations [22–24]. We utilized the 1000 random sample ($HRS = 503$, $MHAS = 497$) for DIF because all DIF results become significant with the large sample size, which makes it impossible to identify meaningful DIF items. The DIF criteria were 0.64 logits as a moderate DIF and 0.43 logits as a single to moderate in the DIF contrast at an alpha value of 0.05 with 2-sided tests of Rasch Welch t-tests [23, 24]. If items demonstrated DIF, we removed them and re-ran the Rasch analysis on the remaining items.

The precision of the common metric was examined by Rasch-based person reliability, which is similar to the traditional statistics of Cronbach's alpha [24]. In addition, the number of person strata for the Rasch model was examined. Person strata indicates distinct person ability groups that are three standard errors apart from one another [25]. We used at least three person strata as a precision criterion, which is equivalent to a reliability of 0.80 [26]. Next, we examined the item difficulty hierarchy among the final functional questions.

After examining the psychometrics of the common metric, we generated item co-calibrations. We utilized this information to separately anchor the full HRS ($n = 20,554$) and MHAS ($n = 14,448$) samples onto the generated item co-calibrations to estimate health status (person measures) for each sample. The estimated person measures were calibrated along the same linear interval scale (log equivalent units or logits), which allowed us to perform multiple linear regression models [27]. Winsteps® Rasch Measurement, Version 3.91.2 was used to conduct Rasch analysis [24].

Disability comparison

We built multiple linear regression models to compare functional status between the U.S. and Mexico. The dependent variable in the regression models was the functional status person measures (logits) estimated from the HRS and MHAS sample. The independent variables were country (U.S. vs. Mexico), demographics (female, single, health insurance, current employment, monthly income, BMI, depression, and cognition), and chronic conditions (stroke, arthritis, diabetes, cancer, lung problems, heart problems, and hearing problems). First, we estimated functional status differences at the population level (Model I) and then we conducted subgroup analyses across chronic conditions (stroke, arthritis, diabetes, cancer, lung problems, heart problems, and hearing problems) by controlling for all covariates included in Model I. Functional status differences were presented as regression coefficients with corresponding 95% confidence intervals (CI) by country. SAS 9.4 software was used to perform multiple linear regressions [16].

Results

Table 1 presents descriptive characteristics of the 2012 HRS and MHAS sample. The American and Mexican samples were statistically different with the exception of female, memory (immediate word recall and delayed recall), and diabetes. While the Mexican adults were younger (mean = 63.5 years) and had less health insurance (46.5%) than the American adults (mean = 67.2 years and 55.6%, respectively), more Mexican adults were working and had fewer chronic conditions (stroke, arthritis, cancer, lung problem, heart problem, hearing problem). However, the Mexican adults had more depressive symptoms (mean of 3.4 in CES-D) and lower self-rated health (mean of 2.3, indicating “Fair”) scores than the American adults (mean of 2.0 in CES-D and mean of 3.1 self-rated health, indicating “Good”).

Unidimensionality

The EFA with one-factor model demonstrated a dominant measurement construct among the 22 function questions with excellent model fit (CFI = 0.979, TLI = 0.977, RMSEA = 0.045) and high factor loadings on the first factor structure ($\lambda > 0.667$). The eigenvalues of the first four factors were 14.36, 1.84, 0.96, and 0.81, respectively, and the ratio of first to second eigenvalues was 7.79 (> 4 supporting the unidimensionality assumption) [17]. The CFA with one-factor solution also supported the identified factor structure with excellent model fit (CFI = 0.971, TLI = 0.969, RMSEA = 0.059) and high factor loadings on the dominant measurement structure ($\lambda > 0.620$). The dominant factor explained 66.4% of the variance.

Rasch analysis

There were six DIF items across the two surveys, including bathing (DIF contrast = 1.47), managing money (DIF contrast = 1.56), preparing meals (DIF contrast = 0.69), jogging 1 mile/km (DIF contrast = 1.40), sitting for 2 h (DIF contrast = 0.82), and climbing 1 flight of stairs (DIF contrast = 0.73), all $p < 0.05$ (Supplementary Table S1). We removed the six DIF items from the 22 function items.

Table 2 presents the items fit to the Rasch model with no DIF items. After removing the six DIF items, all items were a good fit to the Rasch model, except for reaching arm (Infit MnSq = 1.26, Outfit MnSq = 1.68), picking up a dime (Infit MnSq = 1.29, Outfit MnSq = 2.12), and taking medications (Infit MnSq = 1.11, Outfit MnSq = 2.62). We excluded the misfit items and estimated item parameters for the disability comparisons.

The remaining function items demonstrated a conceptual item difficulty hierarchy (Fig. 2). The most challenging items were climbing several flights of stairs (3.31 logits) and stooping (3.14 logits) and the easiest items were toileting (-1.81 logits) and eating (-2.73 logits). The co-calibrated item measures were employed to anchor the full HRS ($n = 20,554$) and MHAS ($n = 14,448$) samples to estimate the health status for each population using the same measurement scale. The 13 function items demonstrated acceptable precision levels (person reliability = 0.78). The Rasch model separated the sample into approximately 3 distinct groups (person strata = 2.7). However, there were high ceiling effects, with 36.6% ($n = 12,801$) of the sample demonstrating maximum scores. The minimum and maximum person measures estimated by the Rasch model were -5.26 logits and 5.32 logits, respectively.

Disability comparison

Figure 2 presents unadjusted person ability estimates (functional status) from the Rasch model for the HRS and MHAS samples, as well as the correspondence between person measures and the 13 function items. The average person ability estimations for the HRS and MHAS were 2.72 logits (SD = 2.60) and 3.07 logits (SD = 2.35), respectively. Table 3 presents the regression coefficients for functional status by country and covariates in the multiple linear regression model ($R^2 = 0.42$, $p < 0.0001$). The American adults showed significantly lower functional status (more disability) than the Mexican adults when controlling for demographics and chronic conditions ($\beta = -0.258$, $p < 0.0001$). Specifically, aging, female, single, health insurance, high BMI, depressive symptoms, and chronic conditions (stroke, arthritis, diabetes, lung problems, and heart problems) were negatively associated with functional status among the American and Mexican adult populations. However, current employment, high self-rated health, and good cognition (immediate and delayed word memory) were positively associated with functional status for the two populations.

Additionally, we ran subgroup analyses across seven chronic conditions (stroke, arthritis, cancer, diabetes, lung problems, heart problems, and hearing problems) using multiple linear regression models, adjusting for demographics and other chronic conditions. Similar to the health comparison at the total population level, the American adults demonstrated lower function (more disability) than the Mexican adults who reported having arthritis ($\beta = -0.36$, $p < 0.0001$) or lung problems ($\beta = -0.62$, $p < 0.0001$). However, among those with a history of stroke, the American adults had higher functional status than the Mexican adults ($\beta = 0.46$, $p = 0.0476$), and there were no differences among those with diabetes ($\beta = -0.05$, $p = 0.5183$), cancer ($\beta = 0.27$, $p = 0.0601$), hearing problems ($\beta = -0.22$, $p = 0.2942$), or heart problems ($\beta = -0.01$, $p = 0.9097$).

Discussion

This study created a common linear metric for assessing the functional status between the American and Mexican older adult populations using a Rasch model. The resulting item set included 13 function items from the HRS and MHAS to construct a common metric that demonstrated good psychometrics, including a unidimensional measurement structure, conceptual item difficulty hierarchy, good fit statistics, and acceptable precision. Based on the Rasch common metric, older Mexican adults demonstrated better functional status (less disability) than older American adults.

The Rasch model indicates that Mexican adults report better physical function than Americans at the population level and across various chronic conditions (arthritis and lung problems); however, American adults with a history of stroke reported higher functional status than their Mexican counterparts. We speculate that some of the differences found across the two countries may be due to different survival rates between the current cohorts of older adults, where sick adults may be more likely to survive to old age in the U.S. than in Mexico [28]. This pattern may be evident in other comparisons between developed and developing countries. However, we speculate that these cross-national differences may narrow in the future due to rapidly changing socioeconomic factors within developing countries [9].

Our findings are consistent with previous disability comparisons between the U.S. and Mexico. Gerst-Emerson et al. [9] estimated 2-year transitions of disability between the U.S. and Mexican older populations using 2000–2002 HRS and 2001–2003 MHAS data. In their study, the U.S. adults demonstrated a more rapid rate of disability transition compared to the Mexican adults. In addition, a recent study reported that the life expectancy (e.g., free of disability) of Mexican adults was higher than that of American adults [29]. While these previous studies provide an indication of disability differences between the two countries, their estimations were not based on a common metric and did not measure the actual amount of disability of American and Mexican adults. In contrast, our Rasch modelled metric estimates the amount of the latent trait (functional status or disability levels) based on a common scale [6, 25]. In addition, the Rasch model indicates what people can or cannot do. For instance, on average, American adults had difficulty walking blocks, climbing several flights of stairs, and stooping (mean = 2.72 logits); however, Mexican adults (mean = 3.07 logits) had less difficulty in performing the same tasks. Similarly, the two populations had less difficulty in ADL tasks (i.e., eating, shopping, walking, getting out of bed, toileting, and eating).

These disability estimations and specific item difficulty levels provide more compressive information for health policy establishments for community-dwelling adults. For instance, health care services for the easy ADL tasks may not be necessary for the community-dwelling adult populations because they are likely independent in performing these ADL tasks. Instead, upper extremity strengthening programs or adaptive equipment services would be appropriate for these adult populations. In addition, in contrast to previous research which compared disability by gender and age between the U.S. and Mexico [8, 9], our study provided more detailed information about which socioeconomic factors were

associated with disability levels. These findings also serve as descriptive information for quasi-experiential designs (e.g., propensity score matching) that can reflect the cause of health differences between the two countries.

The created common metric demonstrated a similar item difficulty hierarchy as the 2010 and 2011 Survey of Health, Ageing and Retirement in Europe (SHARE) [5]. In our study, the most difficult items were stooping and climbing several flights of stairs and the easiest items were toileting and eating; our results match those found by the Rasch modelled SHARE scale. In addition, the two metrics supported the invariant measurement construct with similar measurement precision: HRS-MHAS (person reliability = 0.78, person strata = 2.7) and SHARE (person reliability = 0.74, person strata = 2.6). These findings indicate that the created common metric between the HRS and MHAS could be co-calibrated with the SHARE metric to compare disability severity among the US, Mexico, and 16 European countries.

In addition, the DIF findings provided critical information for international disability comparisons. While there were no DIFs on most basic ADLs like bathing or preparing meals, and IADLs like managing money, other items demonstrated DIFs across the two countries. These differences may be due to the unique cultural aspects of the two countries, such as different home environments, social roles, and/or technologies in the communities. These were new findings expand the previous literature, which focused only on DIFs across gender and age [5, 6]. For instance, the different home environment or access to technologies between the U.S. and Mexico may result in different item difficulties on the same items. In addition, while the HRS sister studies have attempted to include harmonized measures and procedures [1, 2], some items still reflect differences that pose measure challenges. For instance, the jogging 1 mile item in the HRS was translated into the jogging 1 km item in the MHAS. However, the DIF results indicated that the item difficulty estimation in the HRS (5.93 logits, SE = 0.24) was significantly greater than the item difficulty estimation in the MHAS (4.53 logits, SE = 0.19), which is conceivable given the distance differences. The MHAS administrators might want to use km unit instead of mile unit because km unit is the official unit in Mexico, and it is more understandable for Mexican adults. However, the actual distance of 1 mile is 1.6 km, meaning that the same question could represent a more challenging level for Americans than Mexicans. Therefore, researchers should be aware of distinct cultural differences when conducting healthy comparison studies using the HRS sister studies. In short, the presence of DIFs on these items could affect the item difficulty calibrations and its psychometrics, potentially influencing disability estimations across countries.

Our study has several limitations. First, since the U.S. adults appeared to have lower functional status than the Mexican adults, future studies would need to further expand research on the major factors that affect functional status and cause differences between the two countries, including socioeconomic or contextual factors which were not administered by the national surveys. Second, the disability levels were estimated using a cross-section design, and do not represent causality. Future longitudinal studies would need to identify the actual causes of health differences for the two populations who survived only between two time-points. In addition, we did not control for behavioral factors known to influence both

disability and comorbid conditions, such as smoking, drinking, dietary patterns, sedentary lifestyle, and/or drug use, due to difficulties in creating comparable response categories in those items [7, 9]. Since these factors can also affect functional status, future studies are needed to control for these covariates when estimating and comparing functional status across countries. In addition, the DIF results were based on a randomly selected sample (HRS = 503, MHAS = 497) since the full sample inflated all DIF results which prevents the ability to identify meaningful DIF items. Similarly, the fit standardized mean square results (ZSTD) were also inflated by the full sample [30]. Lastly, while we used the ratio of success as well as a Z score scaling in the immediate and delayed recall items to account for the different number of recall items in the HRS and MHAS, the recall items in the HRS are still harder than those in the MHAS as the American adults had to memorize two more words than the Mexican adults.

Conclusions

A Rasch modelled metric demonstrated that American adults were less functional than Mexican adults. This was the first attempt to compare the functional status between the two populations using a common scale. In addition, some ADL and IADL questions of the HRS and MHAS demonstrated different response patterns on the Rasch model across the U.S. and Mexico, which could result in a biased functional status estimation. Using the Rasch modelled functional status scale, future studies will be able to identify the causes of the health differences between the U.S. and Mexico revealed in this study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Health and Retirement Study (HRS) is sponsored by the National Institute on Aging (NIA U01AG009740) and the Social Security Administration. The Mexican Health and Aging Study (MHAS) is partly sponsored by the National Institutes of Health/ National Institute on Aging (Grant Number NIH R01AG018016) and the Instituto Nacional de Estadística y Geografía (INEGI) in Mexico. Data files and documentation for the HRS and MHAS are public use and available at <https://hrs.isr.umich.edu> and <http://www.MHASweb.org>. We thank Dr. Sarah Toombs Smith at the University of Texas Medical Branch for copyediting the manuscript.

References

1. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: The Health and Retirement Study (HRS). *International Journal of Epidemiology*. 2014; 43(2):576–585. DOI: 10.1093/ije/dyu067 [PubMed: 24671021]
2. Wong R, Michaels-Obregon A, Palloni A. Cohort profile: The Mexican Health and Aging Study (MHAS). *International Journal of Epidemiology*. 2015; 46(2):e2.doi: 10.1093/ije/dyu263
3. Minicuci N, Naidoo N, Chatterji S, Kowal P. Data resource profile: Cross-national and cross-study sociodemographic and health-related harmonized domains from SAGE plus ELSA, HRS and SHARE (SAGE+, Wave 1). *International Journal of Epidemiology*. 2016; 45(5):1403–1403j. DOI: 10.1093/ije/dyw181 [PubMed: 27794522]
4. Shih RA, Jinkook L, Lopamudra D. Harmonization of cross-national studies of aging to the Health and Retirement Study: Cognition. Santa Monica, CA: RAND Corporation; 2012.

5. Buz J, Cortés-Rodríguez M. Measurement of the severity of disability in community-dwelling adults and older adults: Interval-level measures for accurate comparisons in large survey data sets. *British Medical Journal Open*. 2016; 6(9):e011842.doi: 10.1136/bmjopen-2016-011842
6. Cieza A, Oberhauser C, Bickenbach J, et al. The English are healthier than the Americans really? *International Journal of Epidemiology*. 2015; 44(1):229–238. DOI: 10.1093/ije/dyu182 [PubMed: 25231371]
7. MHAS Mexican Health and Aging Study. Data files and documentation (public use): Mexican Health and Aging Study. 2012. Retrieved January 1, 2018, from <https://www.MHASweb.org>
8. Díaz-Venegas C, Reistetter TA, Wong R. Differences in the progression of disability: A US–Mexico comparison. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2016.
9. Gerst-Emerson K, Wong R, Michaels-Obregon A, Palloni A. Cross-national differences in disability among elders: Transitions in disability in Mexico and the United States. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2015; 70(5):759–768. DOI: 10.1093/geronb/gbu185
10. Diaz-Venegas C, Reistetter TA, Wang CY, Wong R. The progression of disability among older adults in Mexico. *Disability and Rehabilitation*. 2016; 38(20):2016–2027. DOI: 10.3109/09638288.2015.1111435 [PubMed: 26729017]
11. Chan KS, Kasper JD, Brandt J, Pezzin LE. Measurement equivalence in ADL and IADL difficulty across international surveys of aging: Findings from the HRS, SHARE, and ELSA. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences*. 2012; 67(1):121–132. DOI: 10.1093/geronb/gbr133
12. Juster FT, Suzman R. An overview of the Health and Retirement Study. *Journal of Human Resources*. 1995.
13. Chien S, Campbell N, Hayden O. , et al. RAND HRS data documentation. Version M. 2013. Retrieved January 1, 2018, from <http://hrsonline.isr.umich.edu/>
14. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011–2012. *JAMA*. 2014; 311(8):806–814. DOI: 10.1001/jama.2014.732 [PubMed: 24570244]
15. Radloff LS. The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1997; 1(3):385–401.
16. SAS for windows. [Computer program]. Version 9.4. Cary, NC: SAS Institute Inc; 2017.
17. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*. 2007; 45(5 Suppl 1):22–31. DOI: 10.1097/01.mlr.0000250483.85507.04
18. Strout WF. A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*. 1990; 55(2):293–325.
19. Muthén LK, Muthén BO. *Mplus User's Guide*. 7. Los Angeles, CA: Muthén & Muthén; 2017.
20. Linacre JM. What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*. 2002; 16(2):878.
21. Wright BD, Linacre JM, Gustafson J, Martin-Lof P. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994; 8(3):370.
22. Montanari GE, Ranalli MG, Eusebi P. Latent variable modeling of disability in people aged 65 or more. *Statistical Methods & Applications*. 2011; 20(1):49–63. DOI: 10.1007/s10260-010-0148-6
23. Zwick R, Thayer DT, Lewis C. An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*. 1999; 36(1):1–28.
24. Linacre JM. A user's guide to WINSTEPS® 3.91.2. [Computer program]. 2017.
25. Wright BD, Masters GN. *Rating scale analysis*. 1. Chicago: Mesa Press, Pluribus Pr; 1982.
26. Fisher WP Jr. Reliability, separation, strata statistics. *Rasch Measurement Transactions*. 1992; 6(3): 238.
27. Wright BD. Logits? *Rasch Measurement Transactions*. 1993; 7(2):288.

28. Durazo-Arvizu RA, Barquera S, Lazo-Elizondo M, Franco M, Cooper RS. Cardiovascular disease surveillance in Mexicans and Mexican Americans: A tale of two countries. *Revista Panamericana de Salud Pública*. 2008; 23(2):119–124. [PubMed: 18371283]
29. Payne CF. Aging in the Americas: Disability-free life expectancy among adults aged 65 and older in the United States, Costa Rica, Mexico, and Puerto Rico. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences*. 2018; 73(2):337–348. DOI: 10.1093/geronb/gbv076
30. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*. 2008; 8(1):33.doi: 10.1186/1471-2288-8-33 [PubMed: 18510722]

Fig. 1.
Study flow diagram

Fig. 2.

Person-item map presents the relationships among person ability of the HRS and MHAS sample and item difficulty for the 13 function questions. The numbers on the left side indicate the measures of person ability and item difficulty in logits. Each # on the left side of the map indicates 543 subjects in the HRS and 443 subjects in the MHAS. Each # on the left side of the map indicates 1–542 subjects in the HRS and 1–442 subjects in the MHAS. The 13 function questions are located on the right side of the map. On the map, the people located on the top indicate high person ability (healthy) and those located at the bottom indicate low person ability (less healthy). The M to the right of the vertical line is the mean of the items. “Ss” indicates 1 standard deviation and “Ts” indicates 2 standard deviations for item difficulty and person ability, respectively

Table 1

Demographics of the Health and Retirement Study (HRS) and the Mexican Health and Aging Study (MHAS)

Variables	Total <i>n</i> = 35,002	HRS <i>n</i> = 20,554	MHAS <i>n</i> = 14,448
Age, mean (SD)	65.6 (11.4)	67.2 (10.7)	63.5 (11.6)
Female	20,481 (58.5%)	12,003 (58.4%)	8478 (58.6%)
Single	12,563 (36.0%)	8298 (40.7%)	4265 (29.5%)
Health insurance	18,066 (51.8%)	11,351 (55.6%)	6715 (46.5%)
Employed	15,092 (43.2%)	7064 (34.5%)	8028 (55.5%)
Race			
Non-Hispanic White	13,801 (39.4%)	13,801 (67.2%)	–
Non-Hispanic Black	3547 (10.1%)	3547 (17.3%)	–
Hispanic	15,355 (43.4%)	907 (4.4%)	14,448 (100%)
Other	2271 (6.5%)	2271 (11.1%)	–
Monthly income ^a			
Class 1	17,482 (50.1%)	332 (1.6%)	4259 (29.6%)
Class 2	10,572 (30.3%)	17,574 (85.5%)	9937 (69.2%)
Class 3	608 (1.7%)	1970 (9.6%)	101 (0.7%)
Class 4	6254 (17.9%)	688 (3.3%)	65 (0.5%)
Body mass index (kg/m ²), mean (SD)	28.1 (5.8)	28.4 (6.2)	27.6 (4.9)
Depression (CES-D), mean (SD) ^b	2.6 (2.5)	2.0 (2.3)	3.4 (2.6)
Self-rated health, mean (SD) ^c	2.8 (1.1)	3.1 (1.1)	2.3 (0.8)
Memory, mean (SD) ^d			
Immediate recall	–	5.2 (1.7)	4.7 (1.2)
Delayed recall	–	4.2 (2.0)	4.4 (2.0)
Chronic conditions			
Stroke	1564 (4.5%)	1280 (6.3%)	284 (2.0%)
Arthritis	12,941 (37.4%)	10,996 (54.5%)	1945 (13.4%)
Diabetes	7683 (22.2%)	4442 (22.0%)	3241 (22.4%)
Cancer	2946 (8.5%)	2642 (13.0%)	304 (2.1%)
Lung problems	2746 (7.9%)	1903 (9.4%)	843 (5.8%)
Heart problems	4920 (14.2%)	4417 (21.9%)	503 (3.5%)
Hearing problems	1902 (5.5%)	1737 (8.6%)	165 (1.1%)

^a A high class indicates more income than a low class^b The score ranges from 0 to 9 and a high score indicates more depressive symptoms than a low score^c The score ranges from 0 to 5 and a high score indicates better health status than a low score^d The score ranges of the immediate recall and delayed recall variables were 0–10 for HRS and 0–8 for MHAS

Table 2

Fit statistics and item difficulty hierarchy of the function questions with no DIF

Items	Activity	Measure (logits)	Model SE	Infit MnSq	Outfit MnSq
ClimbStairs	Mobility	3.31	0.02	1.03	1.24
Stooping	Mobility	3.14	0.02	1.06	1.16
GetUpChair	Mobility	2.24	0.02	1.13	1.35
WalkBlocks	Mobility	1.68	0.02	0.83	0.73
PushObject	Mobility	1.43	0.02	0.94	0.91
LiftWeight	Mobility	1.24	0.02	0.94	0.91
Walk1Block	Mobility	0.09	0.02	0.90	0.66
ReachArm	Mobility	0.08	0.02	1.26	1.68 ^a
Shopping	IADL	-0.82	0.02	0.88	0.95
Dressing	ADL	-0.84	0.02	0.90	0.74
PickupDime	Mobility	-1.42	0.03	1.29	2.12 ^a
GetOutBed	ADL	-1.45	0.03	0.90	0.66
Walking	Mobility	-1.56	0.03	0.79	0.43
Toileting	ADL	-1.81	0.03	0.91	0.67
TakeMedication	IADL	-2.59	0.03	1.11	2.62 ^a
Eating	ADL	-2.73	0.04	0.91	0.92

Climbing stairs is the most difficult item and eating is the easiest item

ADL activities of daily living; IADL instrumental activities of daily living; MnSq mean square residual

^a Misfit

Table 3

Regression coefficients, standard errors (SE), the 95% confidence interval, and *p* values from linear multiple regression model using the person measure (logits) estimated by the Rasch model

Variables	Regression coefficient (β)	SE	95% confidence interval	<i>p</i> value
Survey: HRS	− 0.258	0.031	− 0.319 − 0.197	< 0.0001 *
Age	− 0.025	0.001	− 0.028 − 0.022	< 0.0001 *
Female	− 0.425	0.023	− 0.470 − 0.379	< 0.0001 *
Single	− 0.210	0.024	− 0.256 − 0.163	< 0.0001 *
Health insurance	− 0.068	0.025	− 0.117 − 0.018	0.0007 *
Employed	0.461	0.026	0.410 0.512	< 0.0001 *
Monthly income	− 0.035	0.023	− 0.079 0.009	0.1210
Body mass index (BMI)	− 0.050	0.002	− 0.054 − 0.046	< 0.0001 *
Depression (CES-D)	− 0.173	0.005	− 0.182 − 0.164	< 0.0001 *
Self-rated health	0.739	0.012	0.714 0.763	< 0.0001 *
Immediate recall	0.315	0.067	0.185 0.445	0.0008 *
Delayed recall	0.245	0.092	0.065 0.425	< 0.0001 *
Stroke	− 0.754	0.056	− 0.863 − 0.645	< 0.0001 *
Arthritis	− 0.845	0.026	− 0.896 − 0.793	< 0.0001 *
Diabetes	− 0.228	0.027	− 0.280 − 0.175	< 0.0001 *
Cancer	− 0.028	0.039	− 0.105 0.049	0.4795
Lung problems	− 0.523	0.041	− 0.602 − 0.443	< 0.0001 *
Heart problems	− 0.285	0.033	− 0.351 − 0.220	< 0.0001 *
Hearing problems	− 0.037	0.050	− 0.134 0.061	0.4614

* Statistically significance at an alpha level of 0.05

The reference categories were HRS for survey, low age, low income, low depression score, low self-rated health score, and low word recall scores