Article

Physical Function in U.S. Older Adults Compared With Other Populations: A Multinational Study

Dana A. Glei, PhD[1], Noreen Goldman, DSc[2], Carol D. Ryff, PhD[3], and Maxine Weinstein, PhD[1]

Abstract

Objective: We compare physical performance from three U.S. national surveys and nationally representative surveys in England, Taiwan, and Costa Rica. **Method:** For each performance test, we use local mean smoothing to plot the age profiles by sex and survey wave and then fit a linear regression model to the pooled data, separately by sex, to test for significant differences across surveys controlling for age and height. **Results:** Age profiles of performance vary across U.S. surveys, but levels of lung function (peak expiratory flow) and handgrip strength in the United States are as high as they are in the other three countries. Americans also perform as well on the chair stand test as the English and Costa Ricans, if not better, but exhibit slower gait speed than the English at most ages. **Discussion:** With the exception of walking speed, we find little evidence that older Americans have worse physical performance than their peers.

Keywords

physical performance battery, chair stand test, gait speed, peak expiratory flow, nationally representative survey

[1]Georgetown University, Santa Rosa, CA, USA
[2]Princeton University, NJ, USA
[3]University of Wisconsin–Madison, USA

Corresponding Author:
Dana A. Glei, Center for Population and Health, Georgetown University, 5985 San Aleso Court, Santa Rosa, CA 95409-3912, USA.
Email: dag77@georgetown.edu

Introduction

As populations continue to extend life expectancy, a central concern is whether the added time comprises years of *healthy* life. One aspect of healthy life that has major consequences for individual well-being—and for health care costs and productivity—is physical function. An important but as yet unanswered question for Americans is whether their level of physical function is on par with people in other wealthy nations. Some comparative studies based on self-reported limitations suggest that older Americans are more likely to report physical limitations than their same age counterparts elsewhere (Avendano, Glymour, Banks, & Mackenbach, 2009; Crimmins, Garcia, & Kim, 2010; Wahrendorf, Reinhardt, & Siegrist, 2013). For example, 2006 nationally representative data from the United States and 14 European countries indicated that Americans had difficulty performing two out of 10 tasks at age 50, whereas similar difficulties were not evident until at least age 65 in Europe (Wahrendorf et al., 2013). In contrast, a recent analysis based on self-reports revealed so much variation in estimates of physical limitations across comparable nationally representative surveys of the U.S. population that the authors were unable to draw any conclusions about the relative position of Americans (Glei, Goldman, Ryff, & Weinstein, 2017).

In this article, we rely on physical performance assessments (e.g., timed walk, grip strength) administered by an external observer to compare levels of physical function across countries. Because they are more objective, these assessments may provide more consistent measures than self-reports, although we recognize that measured performance and self-reported limitations relate to different aspects of physical capacity. Still, both types of measures have been shown to be predictive of survival (Cooper, Kuh, Hardy, Mortality Review Group, & FALCon and HALCyon Study Teams, 2010; Goldman, Glei, Rosero-Bixby, Chiou, & Weinstein, 2014; Goldman, Glei, & Weinstein, 2016; Reuben et al., 2004; Rosero-Bixby & Dow, 2012; Swindell et al., 2010).

Our primary objectives are twofold. First, we examine the consistency of measured physical performance across three nationally representative surveys of the U.S. population. By comparing data from multiple surveys, all of which sample the U.S. national population, we can ascertain the degree of variability in the estimates. We would expect the estimates based on a comparable performance test to be similar across surveys representing the same population at a given time period. Second, we compare the U.S. estimates with those from nationally representative surveys in England, Taiwan, and Costa Rica. Although these countries are not representative of all wealthy nations, their populations share similar life expectancy and leading causes of

death with the United States. Moreover, a prior comparative study demonstrated that the best predictors of mortality at older ages are similar across these four countries (Goldman et al., 2016).

Background

Performance-based and self-reported measures of physical function are correlated and both are predictive of future health outcomes (Goldman et al., 2014; Guralnik et al., 1994). These two types of measures—often labeled "objective" and "subjective," respectively—capture different aspects of physical function (Reuben et al., 2004).

Performance tests are directly related to specific physiological capacities: walking speed and chair stand speed reflect lower extremity strength and mobility; grip strength represents overall muscle strength (Rantanen et al., 2003); and peak expiratory flow (PEF) indicates lung capacity and airway obstruction. Yet, these assessments are also likely to reflect underlying health and frailty more broadly (Cook et al., 1991; Cooper et al., 2010).

Researchers have emphasized the potential advantages of performance tests. These assessments, administered by a trained observer, are thought to have greater reproducibility, be more comparable across social and cultural contexts, and be more sensitive to minor impairment than self-reports of physical function (Guralnik, Branch, Cummings, & Curb, 1989; Guralnik et al., 1994; Myers, Holliday, Harvey, & Hutchinson, 1993). In terms of the disablement process (Nagi, 1976; Verbrugge & Jette, 1994), performance may capture physical impairment at an earlier stage before it progresses to functional limitation (Guralnik et al., 1989, 1994; Myers et al., 1993; Reuben et al., 2004).

Nonetheless, these assessments entail additional costs and logistical complications. The tests are not only costly to implement in large-scale surveys, but also place additional burden on the respondent and the interviewer. They are time-consuming, demand substantial effort for some older or weak respondents, require special equipment and space for administering, and may compromise response rates for the overall survey. Furthermore, physical performance can be measured only for those willing and able to participate, making such assessments susceptible to selection bias (Myers et al., 1993) and potentially jeopardizing comparability of estimates across surveys. Although performance assessments may be more sensitive to low levels of impairment than self-reported measures of physical function, they are less sensitive at high levels of disability (i.e., for those who are unable to complete the test). An additional concern is that physical performance is only one aspect of function (Reiman & Manske, 2011). Some might favor performance

assessments on the grounds that they are more objective, but a person's perception of his or her ability to function in his or her own environment can have important consequences for real-life physical functioning, which is situation-dependent.

There are few prior comparative studies based on physical performance assessments. One study compared individuals in the United States and India on grip strength and concluded that Americans are stronger than Indians (Albert, Alam, & Nizamuddin, 2005). Another study compared walking speed in China, India, Russia, South Africa, Ghana, and Mexico; the United States was not included (Capistrant, Glymour, & Berkman, 2014). To our knowledge, no one has compared physical performance between the United States and other high-income countries, while examining the consistency of the measurements.

Method

Data

For the purposes of examining the consistency of performance across U.S. surveys, we use cross-sectional data from three nationally representative surveys. Wave 2 of the Midlife in the United States (MIDUS) study included performance assessments during a clinic visit (fielded in 2004-2009). The National Health and Nutrition Examination Survey (NHANES) measured PEF and grip strength during the examination component of the 2011-2012 wave. We also use two waves from the Health and Retirement Survey (HRS) fielded around those same time periods: We compared MIDUS with the 2006 HRS wave (fielded in 2006-2007) and compared NHANES with the 2010 HRS wave (fielded in 2010-2011). None of these surveys included the institutionalized population in the initial sampling frame, but HRS included those who became institutionalized during longitudinal follow-up. Thus, to maximize comparability across surveys, we excluded institutionalized respondents from HRS ($n = 438$ in 2006; $n = 469$ in 2010).

For our cross-national analysis, we compare the U.S. estimates from MIDUS and the 2006 HRS with three population-based surveys in other countries that were fielded around that same time: Wave 2 (2004-2005) of the English Longitudinal Study of Aging (ELSA); the 2006-2007 wave of the Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan; and Wave 1 (2004-2006) of the Costa Rican Study on Longevity and Healthy Aging (CRELES). We selected these datasets because they include similar assessments and represent countries with similar life expectancy spanning four regions of the world: North America, Central America, Europe, and

Asia. Ideally, we would have data collected during the same time frame from multiple surveys for all of the countries in the analysis—not only the United States. Unfortunately, such data are not available. SEBAS and CRELES included institutionalized persons in their sampling frame, whereas ELSA did not. Thus, we excluded the small number of institutionalized respondents from SEBAS ($n = 1$) and CRELES ($n = 30$).

We also excluded respondents for whom age was top-coded ($n = 363$ aged 80+ in NHANES, $n = 109$ aged 90+ in ELSA) and the small number of respondents aged 35 to 36 in MIDUS who participated in the performance tests ($n = 3$). To avoid small cell sizes at very high ages, we restricted the upper end of the age range for each survey to just below the youngest age with fewer than five participants in the performance assessments. Thus, the age ranges of the analytic samples are as follows: MIDUS (ages 37-84), HRS 2006 (ages 52-94), HRS 2010 (ages 50-96), NHANES (ages 20-79), ELSA (ages 52-89), SEBAS (ages 53-87), CRELES (ages 60-101). The number of respondents who participated in at least one of the performance tests was as follows: $n = 1,240$ (MIDUS), $n = 7,173$ (2006 HRS), $n = 8,423$ (2010 HRS), $n = 4,808$ (NHANES), $n = 7,574$ (ELSA), $n = 1,230$ (SEBAS), $n = 2,561$ (CRELES). Table S1 summarizes sample designs, response rates, and restrictions on the analysis sample for each survey.

Measures

Physical performance assessments were administered during a clinic visit in the MIDUS study (all respondents from the national sample were targeted for participation), during a visit to the mobile exam center for NHANES, and in the respondent's home for all other surveys. In ELSA, the performance assessments were conducted during a second home visit, this time by a nurse.

PEF, a measure of lung function, was measured using a spirometer in NHANES and ELSA and a peak flow meter in the other surveys. We compute the maximum of three trials except for NHANES, which collected only one measurement.

Handgrip strength was measured using a dynamometer. CRELES included only two trials on the dominant hand, whereas the other surveys included two or three trials on both hands. To maximize comparability across surveys, we compute the maximum of (the first) two trials on the dominant hand for all surveys.

Chair stand speed is not available in HRS or NHANES, but the other four surveys administered similar assessments. From a sitting position, the respondent was asked to stand up and sit down again 5 times in a row as quickly as possible without using his or her arms. For those able to complete five stands, the completion time was recorded. Values of chair stand speed in SEBAS and

CRELES were adjusted for differences in chair height (because the test was conducted in the home and interviewers had to use whatever suitable chair was available). To adjust for differences in chair height, we regressed the completion time (c_i) for individual i on chair height (h_i) controlling for the respondent's age and height, with models fit separately by sex. The adjusted completion time was calculated as $\tilde{c}_i = c_i + \beta_s(\bar{h} - h_i)$, where β_s was the coefficient for h_i from the sex-specific model and $\bar{h}$ was the mean chair height in the sample (Cornman et al., 2011). Chair stand speed (i.e., number of stands per second) was computed as five (i.e., number of stands) divided by the adjusted completion time $\tilde{c}_i$ (i.e., seconds to complete the test adjusted for age and height). No information about chair height is available for ELSA; thus, those results remain unadjusted. In MIDUS, chair stands were administered in the clinic, where presumably the chair used for the test was standardized across all respondents.

The timed walk is the least comparable across surveys: It is not available for NHANES, and CRELES administered a "get up and go" test (i.e., the respondent started from a sitting position), which is not comparable. Among the other surveys, the length of the walking course is comparable only for HRS and ELSA (8 feet). Thus, we limit our comparisons of gait speed to those two surveys. Respondents were asked to walk the measured course at their normal speed and were allowed to use a walking aid (e.g., cane, walker). The surveys conducted two trials; we retain the maximum of the two trials. The walking test was administered only to respondents aged 65+ in HRS and 60+ in ELSA.

More specifics regarding the physical performance assessments administered in each survey are provided in Table S2. Variation across surveys in participation for each of the performance assessments is shown in Table S3. To facilitate comparisons of effect size, we have converted all of the performance assessments to *Z* scores by standardizing based on the distribution of the pooled samples (across both sexes and all surveys).

All analyses control for age and are carried out separately by sex. Because height can affect physical performance and may vary across populations, we further adjust for measured height in the regression models. We do not adjust for body mass index (BMI) because our intent is to estimate the magnitude of real differences in performance across populations. Adjusting for all of the relevant factors that affect physical function may eliminate between-country differences, but doing so imposes a hypothetical situation.

Analytical Strategy

The size of the analysis samples varies by performance test (see Table S1). Descriptive statistics for analysis variables are presented in Table S4. We use

the *lpoly* command in Stata 12.1 (StataCorp, 2011) to perform local mean smoothing—also known as the Nadaraya–Watson estimator (Nadaraya, 1964; Watson, 1964)—to plot the age profiles for each performance assessment, separately by sex and survey wave. For this method of scatterplot smoothing, a locally weighted average is computed for each point in the smoothing grid (in this case, each age) using a kernel (in this case, Epanechnikov) as the weighting function. One advantage of such smoothing procedures is that they do not impose a functional form on the age pattern, but rather reflect the observed data. With the exception of MIDUS, analyses are weighted to account for sampling design. Analysis weights are not available for the MIDUS biomarker sample.

To test whether differences across surveys are significant, we pool the data across surveys and fit a linear regression model for each performance assessment, separately by sex, controlling for survey wave, age, and height. We use a quadratic specification for age to better capture the slight curvature across age that is evident in the graphs for some assessments. We also include interactions between age and survey wave because the age curve appears to vary by country. These models are limited to respondents with valid data for height and the specified performance test. The *svy* commands in Stata 12.1 are used to account for stratification, clustering, and probability weights. To test for differences between the United States and the other countries, we use a Wald test to compare the mean coefficient across the U.S. surveys fielded in the mid-2000s (MIDUS and the 2006 wave of HRS) with the mean coefficient across the non-U.S. surveys (ELSA, SEBAS, and CRELES), which were fielded around that same period.

Given that participation in the physical performance assessments varies widely across surveys, we perform sensitivity analyses to explore how the selection process may have influenced the results. In these alternative analyses, we use multiple imputation to impute missing data for nonparticipants based on information about their health (e.g., self-reported physical and functional limitations), sociodemographic characteristics, and other factors correlated with participation or physical performance. Then, we reestimate the age profiles for physical performance among the full samples of respondents interviewed (MIDUS, *n* = 4277; HRS 2006, *n* = 8805; HRS 2010, *n* = 10341; NHANES, *n* = 5197; ELSA, *n* = 8671; SEBAS, *n* = 1265; CRELES, *n* = 2,763).

Results

Figures 1 and 2 show some variation in the estimated age curves for PEF and grip strength across the three U.S. surveys. Comparing estimates for individuals of the same sex and age, we see that PEF and grip strength in MIDUS
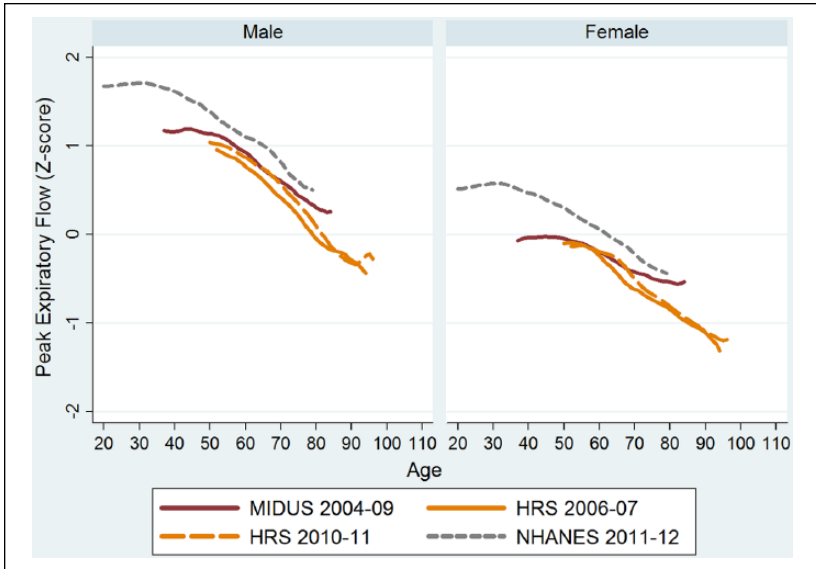
Figure 1. Smoothed age curves for PEF (lung function) by sex, U.S. surveys.
Note. PEF is plotted by age using local mean smoothing. The two surveys fielded in the mid-2000s (MIDUS and HRS 2006-2007) are shown with solid lines, whereas those fielded in the early 2010s (NHANES and HRS 2010-2011) are shown with dashed lines. NHANES included only one trial for PEF, whereas the estimates are based on the maximum of three trials for all the other surveys. MIDUS = Midlife in the United States; HRS = Health and Retirement Survey; NHANES = National Health and Nutrition Examination Survey; PEF = peak expiratory flow.

(2004-2009) appear somewhat higher than estimates based on the HRS wave (2006-2007) fielded around the same time period. Similarly, estimated PEF in NHANES (2011-2012) is better than the corresponding estimates in the HRS wave (2010-2011) during the same time period. In the case of grip strength, estimates based on NHANES and HRS 2010 look similar for women, but men in NHANES perform a bit more poorly than their counterparts in HRS 2010.

Linear regression models on the pooled data confirm that there are significant differences across the U.S. surveys (Table S5). Because the age profile varies by survey, the differences between surveys depend on age. Table S6 shows results from Wald tests comparing the main effects for the U.S. surveys at ages 60, 70, and 80, the age range in common across most of the surveys in this analysis. In most cases, respondents of both sexes in HRS perform significantly worse than their counterparts in MIDUS and NHANES

Figure 2. Smoothed age curves for grip strength by sex, U.S. surveys.
Note. Grip strength is plotted by age using local mean smoothing. The two surveys fielded in the mid-2000s (MIDUS and HRS 2006-2007) are shown with solid lines, whereas those fielded in the early 2010s (NHANES and HRS 2010-2011) are shown with dashed lines. MIDUS = Midlife in the United States; HRS = Health and Retirement Survey; NHANES = National Health and Nutrition Examination Survey.

in terms of PEF (Table S6). For example, at age 70, PEF is around one fifth of a *SD* higher in NHANES 2011-2012 compared with HRS 2010 and in MIDUS (2004-2009) compared with HRS 2006. For grip strength, the differences are smaller. The largest difference is between men in NHANES and HRS 2010: at age 70, grip strength is 0.17 *SD* ($p < .001$) lower in the former compared with the latter.

When we compare the age profiles of physical performance across countries, we find Americans (in MIDUS & HRS) perform at least as well as the English and better than the Taiwanese and Costa Ricans in terms of lung function (PEF; Figure 3), and demonstrate better grip strength than their counterparts in the other three countries (Figure 4). With respect to chair stand speed (Figure 5), Americans (in MIDUS) perform better than Costa Ricans and at least as well as the English. Compared with the Taiwanese, Americans exhibit somewhat slower chair stand speed at younger ages, but faster chair stand speed at the oldest ages. For walking speed, we make

Figure 3. Smoothed age curves for PEF (lung function) by sex and survey.
Note. PEF is plotted by age using local mean smoothing. MIDUS = Midlife in the United States; HRS = Health and Retirement Survey; PEF = peak expiratory flow.

comparisons based only on HRS 2006 and ELSA: below age 80, Americans exhibit slower walking speed than the English (Figure 6).

Some of the American advantage in physical performance could stem from the fact that Americans are taller than their counterparts in the other countries, particularly Taiwan and Costa Rica. Yet, even after accounting for differences in stature, Americans generally exhibit better lung function and grip strength than their counterparts in England, Taiwan, and Costa Rica (Table S6). Among men aged 60, the U.S. advantage in grip strength reaches nearly one third of a *SD*. For chair stand speed (available only for one U.S. survey), Americans (in MIDUS) are faster than their peers in the other three countries (on average), but the difference is bigger for men than women, especially at younger ages. Among men, the U.S. advantage is more than half a *SD* at age 70. In contrast, Americans (in HRS) have slower walking speed than their English counterparts, although the gap narrows at the oldest ages (men: −0.29 *SD* at age 70 to −0.17 at age 80; women: –0.27 *SD* to −0.13 *SD*, respectively).

One problem with comparing results from performance assessments across surveys is the wide variation in participation. Among all respondents who completed the initial interview, participation in at least one performance
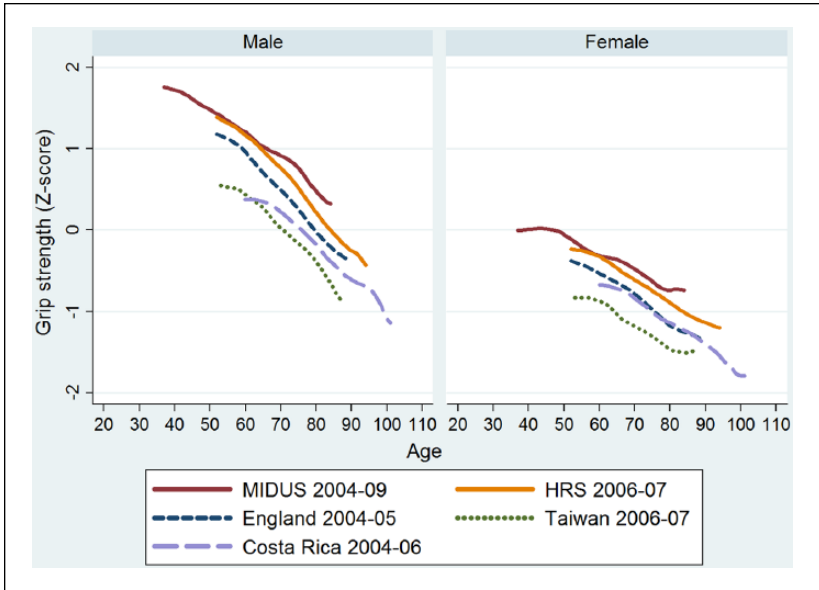
Figure 4. Smoothed age curves for grip strength by sex and survey.
Note. Grip strength is plotted by age using local mean smoothing. MIDUS = Midlife in the United States; HRS = Health and Retirement Survey.

test was much lower in MIDUS (29%) than in the other surveys (81%-97%; Table S3) because the assessments were administered during a comprehensive 2-day physical examination conducted in one of three MIDUS study clinics across the United States. In the other surveys, the tests were administered at the respondent's home (HRS, ELSA, SEBAS, CRELES) or a mobile exam center (NHANES).

Auxiliary analyses of the predictors of participation (Table S7) suggest that participants in the performance assessments are likely to represent a selective sample of more advantaged individuals. In an attempt to discern how selective participation might have influenced the results, we used all of the available information about nonparticipants' health, sociodemographic characteristics, and other factors correlated with participation or physical performance to impute missing data by multiple imputation. When we reestimate the age curves for performance among the full samples, levels of performance generally shift downward, particularly in MIDUS (implying that nonparticipants are likely to have worse performance than participants). Nonetheless, the comparisons across countries remain similar (Figures S1-S6). In the models that adjust for height, the biggest changes are in the
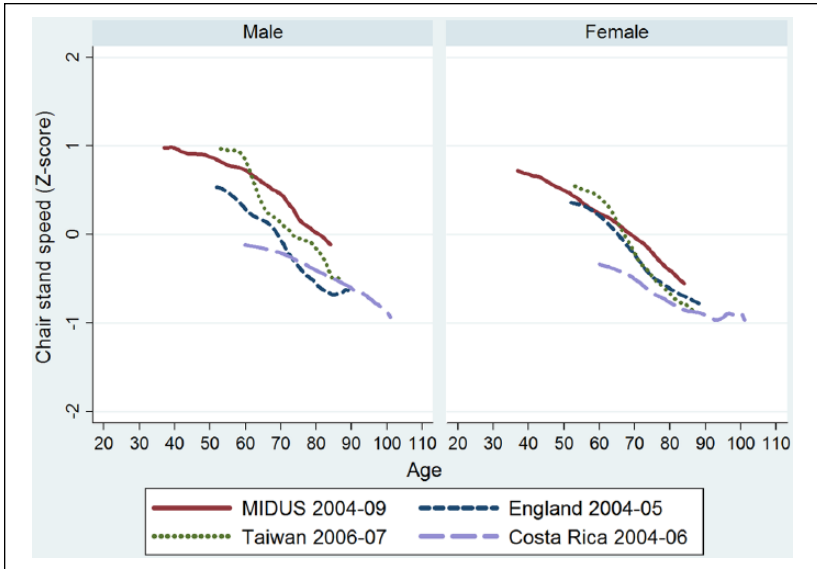
Figure 5. Smoothed age curves for chair stand speed by sex and survey.
Note. Chair stand speed is plotted by age using local mean smoothing. MIDUS = Midlife in the United States.

coefficients for MIDUS: when we include the full sample, the MIDUS advantage on the performance tests is attenuated, particularly for chair stand speed (see Tables S8 and S9). Consequently, the difference between the average coefficient of the U.S. surveys in the mid-2000s and the average coefficient of the non-U.S. surveys, which were fielded around the same time period, is generally reduced (Tables S10 and S11). Nonetheless, there is still a sizable and significant U.S advantage for PEF (except for men aged 60), grip strength, and chair stand speed (among men; differences are not significant for women). The U.S. disadvantage in walking speed relative to the English remains significant at age 70, but converges at the oldest ages.

Discussion

How do Americans fare on physical performance assessments compared with their sex- and age-matched counterparts in these other three countries? Given that we find variability across U.S. surveys in the age profiles of performance, we cannot definitively answer this question. Nonetheless, results from all three U.S. surveys suggest that levels of lung function and grip strength
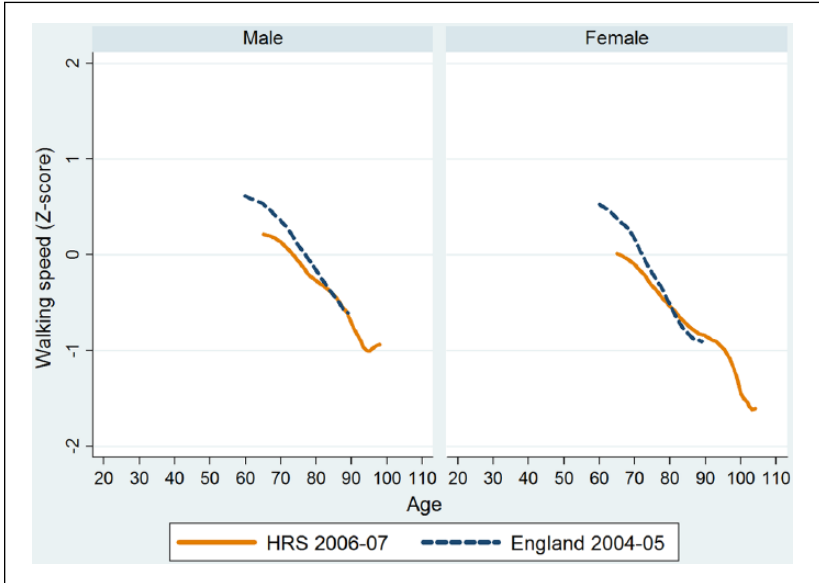
Figure 6. Smoothed age curves for walking speed by sex and survey.
Note. Walking speed is plotted by age using local mean smoothing. HRS = Health and Retirement Survey.

among Americans are as good as, if not better than, performance among their counterparts in three countries with similar life expectancy. Americans also perform as well on tests of chair stand speed as the English and Costa Ricans. These results stand in contrast to earlier comparative studies, based on self-reports, suggesting that Americans are more physically limited than their counterparts elsewhere (Avendano et al., 2009; Crimmins et al., 2010; Wahrendorf et al., 2013). A recent analysis based on self-reports reveals more ambiguous results: Some U.S. surveys suggest an American disadvantage, whereas others indicate similar or better physical function (Glei et al., 2017).

However, the United States does not fare as well at walking speed: below age 80, Americans exhibit slower walking speed than their English counterparts, but gait speed converges at the oldest ages. Unfortunately, this result is based on only one U.S. survey because of the lack of comparable performance tests.

Comparative analysis of physical function based on performance assessment would be greatly enhanced by the availability of multiple surveys representing each population and encompassing a wider variety of high-income

countries. Our study is necessarily limited by the data that are available. Although we have data from three different nationally representative samples for the United States, only one survey is available for each of the other countries in this analysis. The variability in estimates that we observe across surveys within the United States illustrates the importance of studying consistency and reproducibility of research results. Without more data from different surveys that follow a comparable protocol and sample the same population, it may be impossible to determine how Americans compare with their counterparts in other high-income countries in terms of performance-based physical function.

The main focus of this article is on measurement and on evaluating the comparability of estimates across different datasets. Certainly, there could be real differences across countries in levels of physical function, which may be the result of a variety of mechanisms such as individual lifestyle choices and contextual factors related to social, cultural, and policy influences. Yet, before one can explain *why* differences exist, we must first determine whether there are any real differences to explain. One important conclusion from our findings is that lack of comparability of estimates across difference surveys may compromise our ability to identify whether or not there are notable differences.

Although performance assessments may be more comparable across countries than self-reports, they too have limitations. First, there is variation across surveys in the nature of the assessment. For example, the length of the walking course varied from eight to 50 feet across surveys making it impossible to compare walking speed. Second, there may be differences among surveys in the selection process determining who completes the assessments (e.g., variation in exclusion criteria, differences in protocol, location of the tests). Both of these issues bear on the comparability of the estimates across surveys. Third, performance tests can only tell us about a person's capacity to perform the specified task in a contrived situation and, by their very nature, reflect a particular type of functioning (e.g., PEF and grip strength do not reflect lower extremity function). Moreover, they do not measure disability, which needs to be evaluated in terms of individuals' ability to perform roles and tasks expected in their own social environment (Verbrugge & Jette, 1994). Finally, performance assessments may not be as objective as we expect; some researchers have suggested that they may be influenced by sociocultural factors that differ across context (Jeune et al., 2006). For example, social norms with respect to walking speed could affect how fast participants walk when asked to walk at their "normal" pace. In cultures where strength is viewed as a sign of masculinity, men may be more strongly motivated than women (or than men in cultures that do not place such a high value on strength) to perform well on grip strength.

The relative value of self-reported and performance-based measures of physical function may depend on the goal of the study. Self-reported measures of physical function are likely to be useful within a survey sample, but less so for comparing absolute levels across populations. In particular, it is not possible to determine the extent to which observed differences in self-reported function represent true disparities in ability versus other factors that influence reporting. What is less clear, however, is whether performance assessments are the most suitable criteria for comparing physical function across populations. Previous work has shown that within a given survey, self-reported measures and performance assessments are both among the best predictors of survival (Goldman et al., 2016). A thornier question, which we have begun to address in this article but requires additional investigation, is how to interpret cross-national results.

Ethics Statement

Data Sharing Statement

Acknowledgments

Declaration of Conflicting Interests

Funding

Supplemental Material

Supplemental material is available for this article online.

References

Albert, S. M., Alam, M., & Nizamuddin, M. (2005). Comparative study of functional limitation and disability in old age: Delhi and New York City. *Journal of Cross-Cultural Gerontology*, *20*, 231-241. doi:10.1007/s10823-006-9014-2

Avendano, M., Glymour, M. M., Banks, J., & Mackenbach, J. P. (2009). Health disadvantage in US adults aged 50 to 74 years: A comparison of the health of rich and poor Americans with that of Europeans. *American Journal of Public Health*, *99*, 540-548. doi:10.2105/AJPH.2008.139469

Capistrant, B. D., Glymour, M. M., & Berkman, L. F. (2014). Assessing mobility difficulties for cross-national comparisons: Results from the World Health Organization Study on Global AGEing and Adult Health. *Journal of the American Geriatrics Society*, *62*, 329-335. doi:10.1111/jgs.12633

Cook, N. R., Evans, D. A., Scherr, P. A., Speizer, F. E., Taylor, J. O., & Hennekens, C. H. (1991). Peak expiratory flow rate and 5-year mortality in an elderly population. *American Journal of Epidemiology*, *133*, 784-794.

Cooper, R., Kuh, D., Hardy, R., Mortality Review Group, & FALCon and HALCyon Study Teams. (2010). Objectively measured physical capability levels and mortality: Systematic review and meta-analysis. *British Medical Journal*, *341*, Article c4467. doi:10.1136/bmj.c4467

Cornman, J. C., Glei, D., Rodríguez, G., Goldman, N., Hurng, B., & Weinstein, M. (2011). Demographic and socioeconomic status differences in perceptions of difficulty with mobility in late life. *Journal of Gerontology: Social Sciences, 66*(2), 237-248. doi:10.1093/geronb/gbq087

Crimmins, E. M., Garcia, K., & Kim, J. K. (2010). Are international differences in health similar to international differences in life expectancy? In E. M. Crimmins, S. H. Preston & B. Cohen (Eds.), International Differences in Mortality at Older Ages: Dimensions and Sources. Panel on Understanding Divergent Trends in Longevity in High-Income Countries. Committee on Population, Division of Behavioral and Social Sciences and Education. (pp. 68-101). Washington, D.C.: The National Academies Press.

Glei, D. A., Goldman, N., Ryff, C. D., & Weinstein, M. (2017). Can we determine whether physical limitations are more prevalent in the US than in countries with comparable life expectancy? *SSM Population Health, 3*, 808-813.

Goldman, N., Glei, D. A., Rosero-Bixby, L., Chiou, S., & Weinstein, M. (2014). Self-reported versus performance-based measures of physical function: Prognostic value for survival. *Demographic Research*, *30*, 227-252.

Goldman, N., Glei, D. A., & Weinstein, M. (2016). What matters most for predicting survival? A multinational population-based cohort study. *PLoS ONE*, *11*(7), e0159273. doi:10.1371/journal.pone.0159273

Guralnik, J. M., Branch, L. G., Cummings, S. R., & Curb, J. D. (1989). Physical performance measures in aging research. *Journal of Gerontology*, *44*(5), M141-M146.

Guralnik, J. M., Simonsick, E. M., Ferrucci, L., Glynn, R. J., Berkman, L. F., Blazer, D. G., . . . Wallace, R. B. (1994). A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology*, *49*(2), M85-M94.

Jeune, B., Skytthe, A., Cournil, A., Greco, V., Gampe, J., Berardelli, M., . . . Robine, J. M. (2006). Handgrip strength among nonagenarians and centenarians in three European regions. *The Journals of Gerontology, Series A: Biological Sciences & Medical Sciences*, *61*, 707-712.

Myers, A. M., Holliday, P. J., Harvey, K. A., & Hutchinson, K. S. (1993). Functional performance measures: Are they superior to self-assessments? *Journal of Gerontology*, *48*(5), M196-M206.

Nadaraya, E. A. (1964). On estimate regression. *Theory of Probability and Its Application*, *9*, 141-142.

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

Rantanen, T., Volpato, S., Ferrucci, L., Heikkinen, E., Fried, L. P., & Guralnik, J. M. (2003). Handgrip strength and cause-specific and total mortality in older disabled women: Exploring the mechanism. *Journal of the American Geriatrics Society*, *51*, 636-641.

Reiman, M. P., & Manske, R. C. (2011). The assessment of function: How is it measured? A clinical perspective. *The Journal of Manual & Manipulative Therapy*, *19*, 91-99. doi:10.1179/106698111X12973307659546

Reuben, D. B., Seeman, T. E., Keeler, E., Hayes, R. P., Bowman, L., Sewall, A., . . . Guralnik, J. M. (2004). Refining the categorization of physical functional status: The added value of combining self-reported and performance-based measures.

The Journals of Gerontology, Series A: Biological Sciences & Medical Sciences, *59*, 1056-1061.

Rosero-Bixby, L., & Dow, W. H. (2012). Predicting mortality with biomarkers: A population-based prospective cohort study for elderly Costa Ricans. *Population Health Metrics*, *10*(1), Article 11. doi:10.1186/1478-7954-10-11

StataCorp. (2011). *Stata: Release 12. Statistical software*. College Station, TX: Author.

Swindell, W. R., Ensrud, K. E., Cawthon, P. M., Cauley, J. A., Cummings, S. R., Miller, R. A., & Study Of Osteoporotic Fractures Research Group. (2010). Indicators of "healthy aging" in older women (65-69 years of age): A data-mining approach based on prediction of long-term survival. *BMC Geriatrics*, *10*, Article 55. doi:10.1186/1471-2318-10-55

Verbrugge, L. M., & Jette, A. M. (1994). The disablement process. *Social Science & Medicine*, *38*, 1-14.

Wahrendorf, M., Reinhardt, J. D., & Siegrist, J. (2013). Relationships of disability with age among adults aged 50 to 85: Evidence from the United States, England and continental Europe. *PLoS ONE*, *8*(8), e71893. doi:10.1371/journal.pone.0071893

Watson, G. S. (1964). Smooth regression analysis. *Sankhya, Series A*, *26*, 359-372.