SSM - Population Health 3 (2017) 808–813

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Article

Can we determine whether physical limitations are more prevalent in the US than in countries with comparable life expectancy?

Dana A. Glei[a,*], Noreen Goldman[b], Carol D. Ryff[c], Maxine Weinstein[d]

[a] Center for Population and Health, Georgetown University, Mailing address: 5985 San Aleso Court, Santa Rosa, CA 95409-3912, USA
[b] Office of Population Research and Woodrow Wilson School of Public and International Affairs, Princeton University, 243 Wallace Hall, Princeton, NJ 08544-2091, USA
[c] Department of Psychology, University of Wisconsin-Madison, 1300 University Avenue, 2245 MSC, Madison, WI 53706, USA
[d] Center for Population and Health, Georgetown University, 312 Healy Hall, 37th & O Streets, NW, Washington, DC 20057-1197, USA

ARTICLE INFO

Keywords:
Survey data
Self-report
Comparative
Physical function
Physical limitations
United States

ABSTRACT

We evaluate the variability in estimates of self-reported physical limitations by age across four nationally representative surveys in the US. We consider its implications for determining whether, as previous literature suggests, the US estimates reveal limitations at an earlier age than in three countries with similar life expectancy: England, Taiwan, and Costa Rica. Based on cross-sectional data from seven population-based surveys, we use local mean smoothing to plot self-reported limitations by age for each of four physical tasks for each survey, stratified by sex. We find substantial variation in the estimates in the US across four nationally-representative surveys. For example, one US survey suggests that American women experience a walking limitation 15 years earlier than their Costa Rican counterparts, while another US survey implies that Americans have a 4-year advantage. Differences in mode of survey may account for higher prevalence of limitations in the one survey that used a self-administered mail-in questionnaire than in the other surveys that used in-person or telephone interviews. Yet, even among US surveys that used the same mode, there is still so much variability in estimates that we cannot conclude whether Americans have better or worse function than their counterparts in the other countries. Seemingly minor differences in question wording and response categories may account for the remaining inconsistency. If minor differences in question wording can result in such extensive variation in the estimates within a given population, then lack of comparability is likely to be an even greater problem when examining results across countries that do not share the same language or culture. Despite the potential utility of self-reported physical function within a survey sample, our findings imply that absolute estimates of population-level prevalence of self-reported physical limitations are unlikely to be strictly comparable across countries—or even across surveys within the same population.

1. Introduction

Self-reported measures of physical function are included in virtually all large-scale health interview surveys and are widely used in aging-related research. In addition to being easily obtainable, such measures are an important component of prognostic indexes for predicting survival. The subjective nature of self-reports may capture valuable information about underlying health and wellbeing not easily measured by clinical tests. Indeed, research has demonstrated that self-reported measures of physical function are among the strongest predictors of survival at older ages, outperforming standard clinical biomarkers (Goldman et al., 2016; Swindell et al., 2010).

Based on these self-reports, previous comparative studies have concluded that older Americans are more likely to report physical limitations than their same age counterparts in many other countries (Avendano, Glymour, Banks, & Mackenbach, 2009; Crimmins, Garcia, & Kim, 2010; Wahrendorf, Reinhardt, & Siegrist, 2013). A National Academy of Sciences panel found that the percentage of those aged 50 and older reporting a physical limitation was higher in the U.S. than the other seven countries considered—Japan and six European countries (Crimmins et al., 2010). Another study of persons aged 50–85 documented that Americans report a higher number of physical limitations, on average, than their counterparts in 12 out of 13 European countries (Poland was the exception) (Wahrendorf et al., 2013).

Yet, it is unclear whether such self-reported measures are truly comparable across populations that vary in terms of language, culture,

Abbreviations: CRELES, Costa Rican Study on Longevity and Healthy Aging; ELSA, English Longitudinal Study of Aging; HRS, Health and Retirement Survey; MIDUS, Midlife in the United States study; NHANES, National Health and Nutritional Examination Survey; NHIS, National Health Interview Survey; TLSA, Taiwan Longitudinal Study of Aging

* Corresponding author.
E-mail addresses: dag77@georgetown.edu (D.A. Glei), ngoldman@princeton.edu (N. Goldman), cryff@wisc.edu (C.D. Ryff), weinstma@georgetown.edu (M. Weinstein).

http://dx.doi.org/10.1016/j.ssmph.2017.07.008
Received 28 March 2017; Received in revised form 14 July 2017; Accepted 21 July 2017
2352-8273/ <outputStr5>

and social norms (Meijer, Kapteyn, & Andreyeva, 2011). The subjective nature of these measures may make them sensitive to variation in question wording, response categories, and ordering of the questions; to mode of interview and other survey methods; and to differences between individuals in the interpretation of "difficulty."

In this paper, we take advantage of similar questions about physical limitations administered in four US nationally representative surveys, fielded in a similar period, to evaluate the variability in estimates representing the same population. We then consider the implications of this variability for determining whether Americans have more physical limitations than their counterparts in three countries with similar life expectancy—78.9 years in the US versus 80.6 in England/Wales; 78.8 years in Taiwan; 78.7 in Costa Rica as of 2010 (The World Bank, 2015; University of California, Berkeley (USA) & Max Planck Institute for Demographic Research (Germany), 2016).

1.1. Background

Previous research suggests that several factors may affect self-reports of physical function and disability. One important consideration is mode of survey (e.g., face-to-face, telephone, or mail-in questionnaire). Walsh and Khatutsky (2007) demonstrate that estimates of disability vary considerably by survey mode. Second, variation in response may be attributable to differences in sequencing, question wording, and response categories (Dillman & Christian, 2005; Picavet & van den Bos, 1996; Rodgers & Miller, 1997). A third issue is use of proxy respondents, whose assessments can differ from those of the respondents themselves (Rodgers & Miller, 1997). A fourth major concern, particularly for comparative research, is that responses about physical limitations may reflect variation in the threshold for reporting difficulty, owing to such factors as personality, expectations, cultural norms, and physical environments. For example, Melzer, Lan, Tom, Deeg and Guralnik (2004) identify significant differences in thresholds between American and Dutch older adults, as well as across age and income groups within the US; they conclude that part of the apparent Dutch advantage in walking ability results from their higher threshold for reporting difficulties.

2. Materials and methods

2.1. Data

We use cross-sectional data from seven population-based surveys, the first four of which represent the US: wave 2 (2004-06) of the Midlife in the United States (MIDUS) study; the 2006-07 wave of the Health and Retirement Survey (HRS); the 2005-06 National Health and Nutritional Examination Survey (NHANES); the 2006 National Health Interview Survey (NHIS); wave 2 (2004-05) of the English Longitudinal Study of Aging (ELSA); the 2003-04 wave of the Taiwan Longitudinal Study of Aging (TLSA); and wave 1 (2004-06) of the Costa Rican Study on Longevity and Healthy Aging (CRELES). We selected these datasets because they were fielded during a similar period (2003–2007), include similar questions about physical limitations, and represent countries with similar life expectancy spanning four regions of the world: North America, Central America, Europe, and Asia. The availability of four nationally-representative datasets for the US allows us to examine the consistency of the estimates across surveys representing the same population.

Table S1 summarizes sample designs, response rates, and restrictions on the analysis sample for each dataset. For comparability across surveys, we exclude institutionalized respondents. Given that age is top-coded at age 85 and older in NHANES and NHIS and top-coded at age 90 and older in ELSA, we exclude respondents aged 85 and older. In auxiliary analyses (not shown), we test the sensitivity of the results to the exclusion of interviews that were completed by proxy; the conclusions remain unchanged. Among community-dwelling respondents younger than 85, missing data for our key dependent variable (walking limitation) is highest in NHIS (2%) and lowest in TLSA ($<$ 0.05%); those respondents are excluded from analysis following common practice. Our analysis samples comprise: n = 1784 for MIDUS (ages 30–84); n = 15,609 for HRS (ages 52–84); n = 4788 for NHANES (ages 20–84); n = 23,193 for NHIS (ages 18–84); n = 8350 from ELSA (ages 52–84); n = 5040 for TLSA (ages 50–84); and n = 2128 for CRELES (ages 60–84).

2.2. Measures

Each survey asks respondents whether they have difficulty performing four tasks: walking a short distance, lifting/carrying, climbing stairs, and bending/stooping/kneeling/crouching/squatting (see Table S2 for details). This information was collected via a self-administered mail-in questionnaire in MIDUS, a phone interview for a random half of the HRS sample, and a face-to-face interview for the other half of the HRS sample and all respondents in the other surveys. Although there are some differences in the question wording and in the response categories, the walking task is the most comparable across all surveys (but probably less exacting for Taiwan). Thus, we focus primarily on the results for walking. Respondents are coded as having a limitation on the specified task if they report any level of difficulty.

All analyses control for age and sex. For comparisons across US datasets, we further adjust for race/ethnicity and education to account for potential between-survey differences in the demographic characteristics of the samples.

2.3. Analytical strategy

All analyses are weighted using survey-provided probability weights (rescaled as needed so that the sum of weights equals the unweighted sample size for each dataset) to account for the sampling design. We use local mean smoothing to plot the reports of difficulty by age for each physical task, separately by sex and dataset. To quantify differences across datasets in these smoothed curves, we use an age-equivalent formulation (Zajacova, Montez, & Herd, 2014). To test for significant differences across US surveys, we pool the data and fit a logit model for each type of limitation controlling for age, sex, and survey. In a subsequent model, we further adjust for race/ethnicity and education. The "svy" commands in Stata 12.1 are used to fit the models while accounting for survey design (i.e., stratification, clustering, and probability weights).

3. Results

3.1. Comparisons across US datasets

As shown in Fig. 1, about 40 percent of men aged 75 in NHIS—who serve as the reference group—report having a walking limitation. The equivalent age at which a similar percentage of men report a walking limitation is 6 years higher (81) in NHANES, where the question wording is most comparable, while it is age 71 in MIDUS and 84 in HRS (Table 1). Among women, the corresponding age is 71 in NHIS versus 77 in NHANES, but much lower in MIDUS (57); the equivalent age in HRS (76) is similar to NHANES. Thus, there is extensive variation in the sex- and age-specific prevalence of self-reported walking limitation across different datasets representing the US non-institutionalized, national population around 2005.

For the other physical tasks, MIDUS respondents consistently report limitations at a younger age than respondents in the other US surveys (Table 1 and Figs. S3.1–3.3). In the case of lifting/carrying and stair climbing, the US surveys with the most comparable question (NHIS and NHANES) yield similar estimates for stair climbing (Fig. S3.2), but the equivalent ages for lifting/carrying differ by seven years for men and five years for women (Fig. S3.1). For bending/kneeling/stooping, NHIS

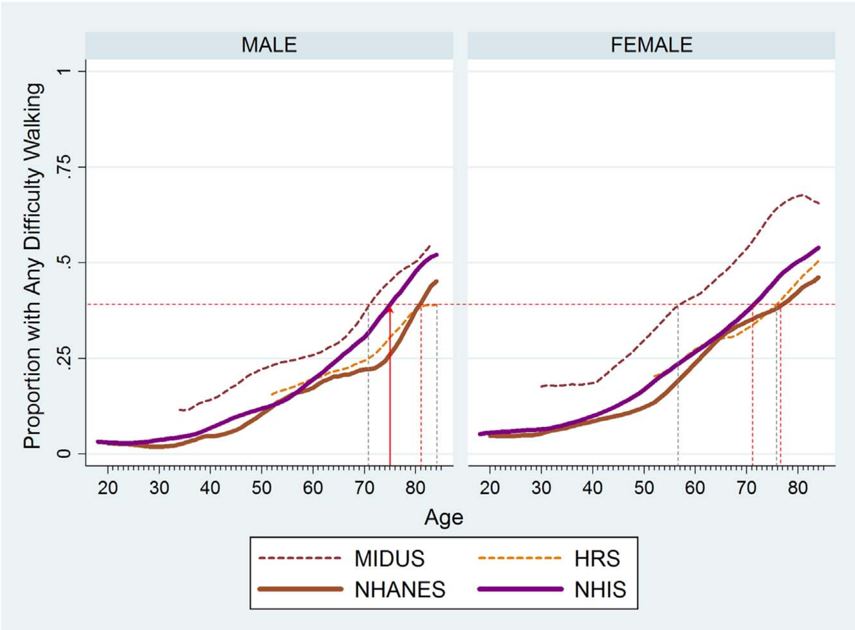

Fig. 1. Smoothed age curves for self-reported walking limitation by sex and dataset, US surveys, weighted. *Note:* Physical limitations by age are plotted using local mean smoothing—also known as the Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964). For each point in the smoothing grid (in this case, each age), a locally weighted average is computed using a kernel (in this case, Epanechnikov) as the weighting function. The question was most comparable in NHANES and NHIS (solid lines). The definition of the task was identical in MIDUS and HRS (dashed lines).

and MIDUS are virtually identical in terms of the defined task, but the equivalent age differs by 23 years in men (75 in NHIS vs. 52 in MIDUS) and 19 years in women (69 and 50, respectively). Across all four US surveys, the estimated age equivalent differs by as much as 26 years (lifting/carrying among men in MIDUS versus NHIS).

Results from regression models on the pooled data indicate that all of the pairwise differences between US surveys are significant. Compared with NHIS, respondents are significantly less likely to report a walking limitation in HRS (OR = 0.92, $p < 0.05$) and NHANES (OR = 0.79, $p < 0.001$), but much more likely to do so in MIDUS (OR = 1.87, $p < 0.001$) (Model 1a, Table S3). For each of the other three physical tasks (Models 2a, 3a, & 4a, Table S3), again all of the differences among US surveys are significant. Even when comparing the surveys with the most comparable questions, reports of a lifting/carrying and a stair climbing limitation are significantly lower in NHANES compared with NHIS (OR = 0.79, $p < 0.001$ and OR = 0.83, $p < 0.01$, respectively) and the probability of reporting a limitation bending/kneeling/stooping is much higher in MIDUS than in HRS (OR = 2.19, $p < 0.001$).

Compared with the other US surveys, MIDUS has fewer non-whites and Latinos and greater representation of better educated respondents (see Table S4). Yet, even among non-Latino whites of similar education, MIDUS respondents report a walking limitation at a younger age than respondents in the other US surveys (Fig. S5). When we adjust for race/ethnicity and education, there is little change in the odds ratios representing between-survey differences in reporting a walking limitation (relative to NHIS, OR = 0.78, $p < 0.001$ for NHANES; OR = 0.93, $p < 0.10$ for HRS; OR = 2.01, $p < 0.001$ for MIDUS; Model 1b, Table S3) or any other type of limitation (Models 2b, 3b, & 4b).

3.2. Comparisons with other countries

Fig. 2 shows a comparison of the age curves for self-reported walking limitation in the four US surveys, England, Taiwan, and Costa

Table 1
Equivalent ages reporting the same level of physical limitation as men aged 75 in NHIS, by sex and survey, weighted analyses.

	USA				England	Taiwan	Costa Rica
	NHIS	NHANES	MIDUS	HRS	ELSA	TLSA	CRELES
Men							
Walking	75[a]	81[a]	71[b]	84[b]	77	> 84[c,d]	79[b]
Lifting/carrying	75[a]	82[a]	49	79[b]	74[b]	68[c]	N/A
Stair climbing	75[a]	77[a]	50	56[b]	62[b]	76	62
Stooping/bending/squatting	75[b]	78	52[b]	75[a]	80[a]	> 84[c,d]	N/A
Women							
Walking	71[a]	77[a]	57[b]	76[b]	73	77[c]	72[b]
Lifting/carrying	57[a]	62[a]	34	52[b]	< 52[b,d]	56[c]	N/A
Stair climbing	64[a]	65[a]	< 30[d]	< 52[b,d]	52[b]	66	< 60[d]
Stooping/bending/squatting	69[b]	62	50[b]	60[a]	70[a]	71[c]	N/A

Abbreviations: CRELES = Costa Rican Study on Longevity and Healthy Aging; ELSA = English Longitudinal Study of Aging; HRS = Health and Retirement Survey; MIDUS = Midlife in the United States study; N/A = Not Available; NHANES = National Health and Nutritional Examination Survey; NHIS = National Health Interview Survey; TLSA = Taiwan Longitudinal Study of Aging.

Note: We use men aged 75 in NHIS as the reference group; for all other subgroups, we show the locally-weighted mean age (based on the smoothed curves in Figs. 2 and S6.1–S6.3) at which men/women in each of the other datasets report the same level of limitation as men aged 75 in NHIS.

[a] These surveys appear to be the most strictly comparable (in terms of question wording, defined task, and response categories).

[b] The defined task is virtually identical in these surveys.

[c] The defined task in this survey is the most different from the others.

[d] Equivalent age is outside of observed age range for this sample.

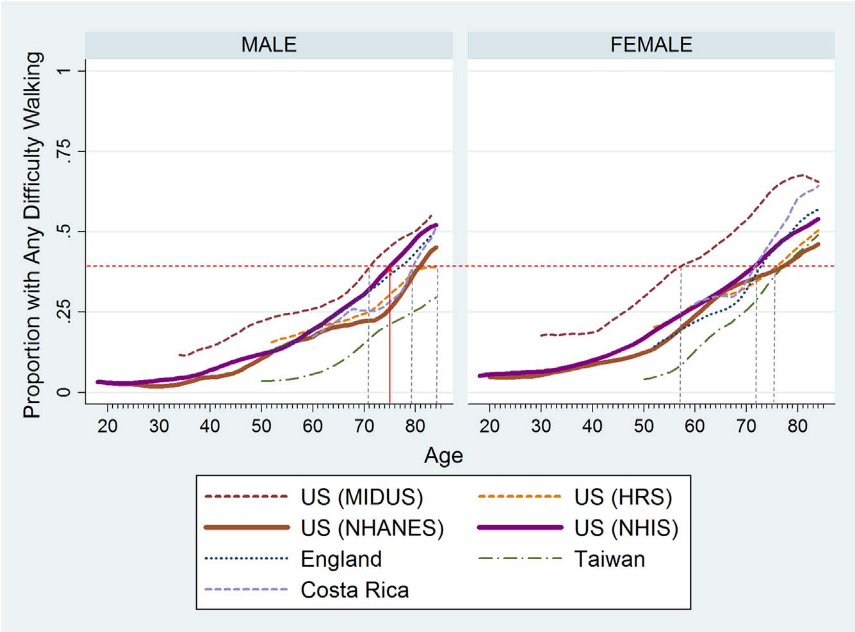

Fig. 2. Smoothed age curves for self-reported walking limitation by sex and dataset, weighted. *Note:* We use local mean smoothing to plot physical limitations by age for two of the US datasets (MIDUS and NHANES) alongside the corresponding curves based on ELSA (England), TLSA (Taiwan), and CRELES (Costa Rica). The question was most comparable in NHANES and NHIS (solid lines). The definition of the task was identical in MIDUS, HRS, and Costa Rica (dashed lines).The wording differed in England (dotted line) and Taiwan (dash-dotted line).

Rica. There is so much variation in the estimated age curves across US datasets that we would draw a completely different conclusion about whether Americans have better or worse function than their counterparts in these three countries. For example, the defined task is identical in MIDUS, HRS, and CRELES, but MIDUS suggests that American men report a walking limitation eight years earlier than their counterparts in Costa Rica, whereas HRS implies that American men look *more* mobile than Costa Ricans (Table 1). Among women, MIDUS respondents also report a walking limitation 15 years earlier than their Costa Rican counterparts, but HRS suggests that Americans have a four year *advantage*.

With respect to the other physical tasks, estimates from MIDUS consistently suggest that Americans experience physical limitations earlier than their counterparts in England, Taiwan, and Costa Rica, yet results from other US surveys are mixed: in some cases, they suggest a US disadvantage, and in other cases, Americans look better (Table 1 and Figs. S6.1–6.3). If we restrict our comparisons to the surveys where the defined tasks are virtually identical (HRS and ELSA), Americans appear to exhibit limitations earlier than the English for stair climbing and stooping/kneeling/crouching, but later than the English for lifting/carrying.

3.3. Sensitivity analyses

In the analyses presented above, respondents who reported they "do not do" a specified activity (in the surveys that included that response option: HRS, NHANES, NHIS, and CRELES) were treated as missing data and excluded from analysis. We explored the sensitivity of the results recoding those respondents as having difficulty. Because they comprise only a small fraction of the sample in any given survey ($\leq$ 1% for kneeling/stooping; $<$ 2% for walking; $<$ 3% for lifting/carrying; $<$ 6.5% for climbing stairs), including them in the analysis as having a limitation for that task has little effect on the results. The smoothed age profiles suggest slightly higher levels of limitation, particularly at older ages. The odds ratios from the regression models also change very little (results not shown).

4. Discussion

Prior studies have documented notable variation in the estimated prevalence of self-reported disability and physical function across surveys of the same population (Freedman et al., 2004; Freedman et al., 2013; Picavet & van den Bos, 1996; Rodgers and Miller, 1997; Wiener, Hanley, Clark, & Van Nostrand, 1990), but the extent to which this variability constrains any attempt to make comparisons across countries has not been fully appreciated. A National Academy of Sciences study and two other comparative studies (Avendano et al., 2009; Crimmins et al., 2010; Wahrendorf et al., 2013) found that Americans report more physical limitations than individuals in other high-income countries. Given such findings, it is tempting to conclude that Americans are, in fact, more physically limited. Yet, our results demonstrated so much variability in the estimated age curves for physical limitations across samples intended to represent the national US population that it is impossible to say whether Americans have worse physical function than their counterparts in other countries with similar life expectancy. If we had multiple datasets for the other countries with which to assess consistency of those estimates, we suspect that the conclusions would have been even more ambiguous.

Mode of survey may account for at least some of the higher reporting of physical limitations in MIDUS compared with the other US surveys. MIDUS used a mail-in self-administered questionnaire, whereas the other three US surveys asked about physical function in face-to-face interviews (with HRS also using phone interviews for half the sample). Earlier studies found higher prevalence of disability and physical limitations for self-administered questionnaires surveys than for in-person interviews (Picavet & van den Bos, 1996; Walsh & Khatutsky, 2007). Thus, results from mail-in surveys should not be compared with estimates based on in-person interviews. In HRS, we tested whether the probability of reporting a limitation for each of the four tasks differed between those interviewed by phone versus in-person controlling for age and sex, but we found no significant difference by mode of survey for any of the tasks (results not shown). Excluding MIDUS, we find more consistency in the estimates across the other three US surveys. Nonetheless, enough variability remains that we still cannot conclude whether Americans have better or worse function than individuals in the other countries.

There were also differences across surveys in question wording and response categories (see Section S2 of Supplementary material for details). Picavet & van den Bos (1996) showed that even seemingly minor differences in wording could generate major differences in estimated prevalence of functional limitations. For example, they concluded that about one-quarter of reported mobility limitations are temporary (which some surveys screen out by using introductions that emphasize long-standing disability). For the walking task, HRS specifically excludes short-term limitations, whereas none of the other surveys (in the US and elsewhere) do so. Other differences in the wording of the specified task, whether the question emphasized the ability to perform the task without any assistance or special equipment, the number and nature of the response categories (e.g., whether there is a category for "does not do it" or distinctions between different levels of difficulty), and even the ordering of questions within the larger survey could have important effects on the resulting estimates (Picavet & van den Bos, 1996). None of these questions is strictly comparable across any pair of surveys, even where efforts have been made to harmonize them (e.g., HRS and ELSA). For example, both the question wording and the response categories for the walking task are very different in HRS compared with ELSA (Table S2).

Thresholds for acknowledging activity limitations (i.e., definition of "difficulty") may also vary across individuals and by societal characteristics within and across surveys. For example, people who live in a very hot climate or hilly terrain may find it more difficult to walk a quarter mile. Although researchers have used externally administered performance assessments (e.g., gait speed, grip strength) and anchoring vignettes to adjust for such reporting heterogeneity (Meijer et al., 2011; Melzer et al., 2004; Salomon, Tandon, & Murray, 2004), these techniques are generally designed to improve estimates of *relative* rather than absolute levels of physical limitations.

In future research, we will examine whether performance assessments are less variable across surveys than self-reported physical limitations. Although the more objective nature of performance assessments may render more consistent estimates, such assessments have their own drawbacks. Not only are they more difficult and expensive to collect, but they provide no information about individuals who are unable or unwilling to perform the test. Comparability and absolute validity of performance assessments across surveys may also be compromised by differences in the nature of the assessment and by differences in selective participation (e.g., variation in exclusion criteria, differences in protocol that influence willingness to participate, and variation in the amount of effort or discomfort required before a person concludes s/he cannot do the task). Moreover, performance assessments are unlikely to provide a complete picture of physical function because self-reports capture a subjective evaluation that reflects underlying health beyond what can be determined by objective measures.

Our study is limited by cross-sectional data (which provide estimates of age-specific prevalence, but do not yield information regarding the age of onset) and by the lack of multiple surveys for countries other than the US. Yet, the inconsistencies we observe in cross-sectional estimates are likely to plague longitudinal studies as well. In addition, there is no reason to believe that estimates would be more consistent in other countries. Indeed, Picavet and van den Bos (1996) showed extensive variability in estimated physical limitations across multiple surveys in the Netherlands.

5. Conclusions

Despite the potential utility of self-reported physical function within a survey sample, our findings imply that absolute estimates of population-level prevalence of self-reported physical limitations are unlikely to be strictly comparable across countries—or even across surveys within the same population. Indeed, the observed variability in the estimates across high quality US surveys representing the same population suggests that we may not be able to estimate absolute levels for a given population let alone compare levels across countries. If minor differences in question wording can result in substantial variation in the estimates within the same population, then lack of comparability is likely to be an even greater problem when contrasting results across countries that do not even share the same language or culture. Beyond the problems of cross-national and cross-survey comparability, comparisons across time may also be affected if cultural and individual factors that influence reporting (e.g., thresholds for reporting "difficulty") vary over time or there are changes in survey methods (e.g., ordering or wording of the questions, survey mode). Further work is needed to isolate the possible sources of this variability and to identify modifications to survey design that can help maximize comparability. Also of importance is additional research to clarify whether and how perceptions of physical function have their own health consequences.

Conflict of interest statement

There are no known conflicts of interest associated with this publication.

Financial disclosure statement

There has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

This work was supported by the National Institute on Aging [grant numbers R01AG16790, R01AG16661, P01 AG020166]; the Eunice Kennedy Shriver National Institute of Child Health and Human Development [grant number P2CHD047879]; the General Clinical Research Centers Program at the National Institutes of Health [grant numbers M01-RR023942, M01-RR00865]; the National Center for Advancing Translational Sciences, National Institutes of Health [grant number UL1TR000427], and the Graduate School of Arts and Sciences, Georgetown University.

We are grateful to the staff at the Center for Population and Health Survey Research (Health Promotion Administration at the Ministry of Health and Welfare in Taiwan) who designed the TLSA and supervised all aspects of the fieldwork and data processing. We also thank all of the individuals who participated in the CRELES, ELSA, HRS, MIDUS, NHANES, NHIS, and TLSA surveys for their vital contributions to the resulting datasets.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ssmph.2017.07.008.

References

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*(3), 540–548. http://dx.doi.org/10.2105/AJPH.2008.139469.

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.

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods, 17*(1), 30–52. http://dx.doi.org/10.1177/1525822X04269550.

Freedman, V. A., Crimmins, E., Schoeni, R. F., Spillman, B. C., Aykan, H., Kramarow, E., Land, K., Lubitz, J., Manton, K., Martin, L. G., Shinberg, D., & Waidmann, T. (2004). Resolving inconsistencies in trends in old-age disability: Report from a technical working group. *Demography, 41*(3), 417–441.

Freedman, V. A., Spillman, B. C., Andreski, P. M., Cornman, J. C., Crimmins, E. M., Kramarow, E., Lubitz, J., Martin, L. G., Merkin, S. S., Schoeni, R. F., Seeman, T. E., & Waidmann, T. A. (2013). Trends in late-life activity limitations in the United States: An update from five national surveys. *Demography, 50*(2), 661–671. http://dx.doi.org/10.1007/s13524-012-0167-z (doi).

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. http://dx.doi.org/10.1371/journal.pone.0159273.

Meijer, E., Kapteyn, A., & Andreyeva, T. (2011). Internationally comparable health indices. *Health Economics, 20*(5), 600–619. http://dx.doi.org/10.1002/hec.1620.

Melzer, D., Lan, T., Tom, B. D. M., Deeg, D. J. H., & Guralnik, J. M. (2004). Variation in thresholds for reporting mobility disability between national population subgroups and studies. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences, 59*(12), 1295–1303.

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

Picavet, H. S., & van den Bos, G. A. (1996). Comparing survey data on functional disability: The impact of some methodological differences. *Journal of Epidemiology and Community Health, 50*(1), 86–93.

Rodgers, W., & Miller, B. (1997). A comparative analysis of ADL questions in surveys of older people. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 52B*(Special Issue), 21–36.

Salomon, J. A., Tandon, A., & Murray, C. J. (2004). Comparability of self rated health: Cross sectional multi-country survey using anchoring vignettes. *British Medical Journal (Clinical Research Ed), 328*(7434), 258. http://dx.doi.org/10.1136/bmj.37963.691632.44 (doi).

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. *BioMed Central Geriatrics, 10*. http://dx.doi.org/10.1186/1471–2318-10-55 (55-2318-10-55).

The World Bank (2015). *Life expectancy at birth, total (years)*. Retrieved May 27, 2015, from 〈http://data.worldbank.org/indicator/SP.DYN.LE00.IN〉.

University of California, Berkeley (USA), & Max Planck Institute for Demographic Research (Germany) (2016). *Human mortality database.* Retrieved 5 December 2016, from 〈www.mortality.org〉.

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. http://dx.doi.org/10.1371/journal.pone.0071893.

Walsh, E. G., & Khatutsky, G. (2007). Mode of administration effects on disability measures in a sample of frail beneficiaries. *The Gerontologist, 47*(6), 838–844.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics; Series A, 26*, 359–372.

Wiener, J. M., Hanley, R. J., Clark, R., & Van Nostrand, J. F. (1990). Measuring the activities of daily living: Comparisons across national surveys. *Journal of Gerontology, 45*(6), S229–S237.

Zajacova, A., Montez, J. K., & Herd, P. (2014). Socioeconomic disparities in health among older adults and the implications for the retirement age debate: A brief report. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 69*(6), 973–978. http://dx.doi.org/10.1093/geronb/gbu041.