



Original article

The English are healthier than the Americans: really?

Alarcos Cieza,^{1,2,3*} Cornelia Oberhauser,² Jerome Bickenbach,^{3,4}
Richard N Jones,⁵ Tefvik Bedirhan Üstün,⁶ Nenad Kostanjsek,⁶
John N Morris⁷ and Somnath Chatterji⁸

¹Faculty of Social and Human Sciences, University of Southampton, Southampton, UK, ²Department of Medical Informatics, Biometry and Epidemiology, Pettenkofer School of Public Health (PSPH^{LMU}), Research Unit for Biopsychosocial Health, Ludwig-Maximilians University, Munich, Germany, ³Swiss Paraplegic Research, Nottwil, Switzerland, ⁴Department of Health Sciences and Health Policy, University of Lucerne, Lucerne, Switzerland, ⁵Department of Psychiatry and Human Behavior, Warren Alpert Medical School, Brown University, Providence, RI, USA, ⁶Classification, Terminology and Standards, Department of Health Statistics and Informatics, World Health Organization, Geneva, Switzerland, ⁷Institute for Aging Research, Social and Health Policy Research, Hebrew SeniorLife, Boston, MA, USA and ⁸Surveys, Measurement and Analysis, Department of Measurement and Health Information Systems, World Health Organization, Geneva, Switzerland

*Corresponding author. Faculty of Social and Human Sciences, Academic Unit Psychology, University of Southampton, Southampton SO17 1BJ, UK. E-mail: a.cieza@soton.ac.uk

Accepted 5 August 2014

Abstract

Background: When comparing the health of two populations, it is not enough to compare the prevalence of chronic diseases. The objective of this study is therefore to propose a metric of health based on domains of functioning to determine whether the English are healthier than the Americans.

Methods: We analysed representative samples aged 50 to 80 years from the 2008 wave of the Health and Retirement Study (N = 10 349) for the US data, and wave 4 of the English Longitudinal Study of Ageing (N = 9405) for English counterpart data. We first calculated the age-standardized disease prevalence of diabetes, hypertension, all heart diseases, stroke, lung disease, cancer and obesity. Second, we developed a metric of health using Rasch analyses and the questions and measured tests common to both surveys addressing domains of human functioning. Finally, we used a linear additive model to test whether the differences in health were due to being English or American.

Results: The English have better health than the Americans when population health is assessed only by prevalence of selected chronic health conditions. The English health advantage disappears almost completely, however, when health is assessed with a metric that integrates information about functioning domains.

Conclusions: It is possible to construct a metric of health, based on data directly collected from individuals, in which health is operationalized as domains of functioning.

Its application has the potential to tackle one of the most intractable problems in international research on health, namely the comparability of health across countries.

Key words: Health, functioning, health state, cross-cultural comparison, Rasch model, health metric

Key Messages

- Comparing the health of populations based on the prevalence of health conditions does not give the full story about health comparisons. This approach does not take the severity of health conditions into account and assumes that two people with a health condition such as diabetes always have the same level of health, or that someone with hypertension is equally unhealthy as someone with diabetes.
- It is possible to construct a metric of health based on data directly collected from individuals, in which health is operationalized as domains of functioning.
- The English have better health than the Americans when population health is assessed only by prevalence of selected chronic health conditions. The English health advantage disappears, however, when health is assessed with a metric that integrates information about functioning domains.

Introduction

Our common sense notion of population health tells us that if two populations are identical in every respect except that the prevalence of chronic diseases is higher in one, then the population with the lower prevalence is healthier. Based on this intuition, Banks and colleagues have carried out several studies using population-based data, coming to the conclusion that the English are healthier than the Americans, for all socioeconomic groups and across the lifespan.^{1,2} This conclusion has been highlighted by a panel of experts convened by the US National Research Council and the Institute of Medicine that has recently reported its findings.³ This approach to comparing the health of populations, however, does not take the severity of health conditions into account and assumes that two people with a health condition such as diabetes always have the same level of health, or that someone with hypertension is equally unhealthy as someone with diabetes.

We think that this approach does not give the full story about health comparisons. We endeavoured to generate a composite measure of overall health that takes into consideration chronic disease severity in terms of the impact of health conditions on the person. As opposed to the Global Burden of Disease studies,⁴ which compare population health using a synthetically constructed proxy measure combining mortality and psychometrically weighted morbidity, drawing from this measure important conclusions about population health change over time, we propose to use data directly derived from individuals. The question we

address in this paper is whether the English would still be healthier than the Americans if we took this approach.

Following the World Health Organization (WHO), we operationalize health with the notion of ‘health state’ understood as: (i) an intrinsic attribute of an individual that can be aggregated to the population level; and (ii) comprising domains of human functioning that describe the actual impact of health conditions on people’s lives.⁵ We treat health state as a unidimensional construct, recognizing that at some level of precision any construct is unidimensional and at another level of precision no construct would be.⁶ We compare the health of the English and American populations by constructing a cardinal metric of health state with the data from the Health and Retirement Study (HRS) in the USA and the English Longitudinal Study of Ageing (ELSA) for England.

Methods

Data

We use data from the 2008 wave of the HRS for the US data, and wave 4 of ELSA (May 2008–July 2009), for English counterpart data. The HRS⁷ and ELSA⁸ are biannual, longitudinal and nationally representative surveys that focus on adults aged 50 and over. Both datasets are openly available after registration. HRS data are available from the corresponding website [<http://hrsonline.isr.umich.edu/index.php?p=data>] and ELSA data are available from the UK data service [<http://discover.ukdataservice.ac.uk/>].

To operationalize health state in terms of domains of human functioning, we identified questions and measured tests addressing those domains that are common to both surveys. 34 self-report questions were identified. They consisted of impairments in body and mental functions ('Are you often troubled with pain?' and 'How much of the time during the past week did you feel depressed?'), and difficulties in activities of daily living (ADLs) and instrumental activities of daily living (IADLs) ('Because of a health problem, do you have any difficulty with bathing or showering?' and '...do you have any difficulty with managing your money – such as paying your bills and keeping track of expenses?'). The response options for these selected questions were coded or recoded so that higher values indicated worse health.

Six variables were selected from the measured tests used in both surveys. Grip strength was assessed with a hand dynamometer. Lung function was assessed with peak expiratory flow rate (PEFR). Balance was evaluated with three progressively more difficult stances: side-by-side, semi-tandem and tandem.⁹ These results were recoded into a polytomous variable with four response options (0 = 'ability to perform tandem stand', 1 = 'ability to perform semi-tandem but not tandem stand', 2 = 'ability to perform side-by-side stand but not semi-tandem stand' and 3 = 'not able to perform side-by-side stand'). Cognitive functions were assessed by immediate and delayed recall of 10 common nouns and an orientation test consisting of reporting day, date, month and year. For grip strength, lung function, immediate recall and delayed recall the sample was then divided into three groups: low (<one standard deviation (SD) below the mean), medium (\pm one SD around the mean) and high (>one SD above the mean). The distributions of all measured tests in both populations were very similar so that the same thresholds, based on the mean of both surveys, were applied; for grip strength and lung function, the thresholds were defined separately for males and females.

Analysis strategy

Descriptive statistics were used to characterize the sample, taking the sampling weights into account.

Our strategy was, first, to replicate the analysis of Banks *et al.*¹ with the more recent waves of data to ensure consistency of results when using the population between ages 50 and 80. We calculated the age-standardized disease prevalence of diabetes, hypertension, all heart diseases, stroke, lung disease, cancer and obesity. We selected these conditions because they were the ones selected by Banks *et al.* and because other conditions had been captured with different approaches in both surveys, making the

comparison very difficult. Also following Banks *et al.*, we selected from both surveys the demographic variables of age, sex and ethnicity and the socioeconomic (SES) variables of education and household income, and divided the SES variables into three groups: low, medium and high education and income. To ensure that the results are not affected by health-related features of minority populations in the two countries (Blacks and Hispanics in the USA, Blacks and Asian immigrants in England), analyses were restricted to White populations in the two countries. As a result, the sample sizes were 10 349 for HRS and 9405 for ELSA. We recognize that this limits the generalizability of the results and that when we talk about 'the English' and 'the Americans' we exclusively refer to the White populations in both countries.

Second, item response theory (IRT) was used to construct a metric of health with all self-report questions and measured tests addressing the selected domains in each survey. For each survey separately, we evaluated the assumptions of IRT, namely unidimensionality, local independence and monotonicity. We then combined the data of the surveys after collapsing response options with very low frequencies and then, by using the Polytomous Rasch Model, we created a single health scale.^{10,11} Then we tested for differential item functioning (DIF) for survey, gender and age groups (≤ 64 and > 64) using iterative hybrid ordinal logistic regression with change in McFadden's pseudo R-squared measure (> 0.02) as DIF criterion.^{12,13} Questions and measured tests showing DIF were calibrated separately for each of the two groups showing DIF. After DIF correction, we calculated a final Rasch model. Based on the calibrations of the included questions and measured tests, a summary score of the health state of each of the individuals in the sample was calculated. We transformed the resulting scores into more meaningful values¹⁴ ranging from worst health in the sample (value 0) to best health (value 100).

Third, to test whether the health of the English was better than that of the Americans, we calculated a linear additive model¹⁵ controlling for socio-demographic and SES variables taking the sampling weights into account. Age was modelled as a non-parametric effect using P-splines. ELSA was used as the reference population, male for gender and low income and low education for the SES variables.

All analyses were performed with R version 2.15.1.¹⁶

Results

Table 1 presents sample characteristics (age, gender, education and income) for the two surveys. The table also includes the percentages of the population, rating their

overall health as 'excellent', 'very good', 'good', 'fair' and 'poor'. Overall, the two populations are very similar in characteristics.

Table 2 shows the age-standardized prevalence of health conditions by education and income and in total. The results of Table 2 confirm that the Banks strategy produces the same results when using the 2008 wave for HRS and wave 4 of ELSA, namely that the prevalence of the reported conditions was higher in the American population than in the English. We also confirmed the negative gradient across education and income. Only the prevalence of

lung disease is higher in England for these waves, due to the fact that the ELSA survey included asthma in lung disease whereas the HRS did not.

In the analysis creating the metric of health, the evaluation of the Rasch model assumptions showed that the three assumptions were reasonably justified. Unidimensionality was probed with bifactor analysis. Bifactor analysis assumes the presence of a single general factor and multiple independent group factors.^{17,18} Bifactor analyses supported the assumption of a strong general factor, but the questions and measured tests from the domains cognition, emotion, sleep, vision and hearing loaded higher in their respective group factors than the general factor. Nevertheless since these domains also loaded high in the general factor and because conceptually these domains contribute to the hypothesized dimension of health, we decided to proceed with unidimensional Rasch. To check whether this decision affected the results, we repeated the Rasch analyses with and without cognition, emotion, sleep, vision and hearing and confirmed that the results (inferences) did not change. The Pearson correlation of the person's abilities produced in both Rasch analyses was 0.92. For local independency, the low percentage of residual correlations above 0.25 (0.9% in both HRS and ELSA) resulting from a single factor confirmatory factor analysis supports the assumption that most of the questions are conditionally independent given an individual score on the latent trait. After collapsing the response options of two items because of low frequencies, all items satisfied the monotonicity assumption.

With regards to DIF analyses, nine of the 40 questions and measured tests showed DIF and were separately calibrated in the two groups. Three variables—questions

Table 1. Sample characteristics of HRS and ELSA populations, including response frequencies of the general health question

Characteristics	USA	England
	Non-Hispanic Whites aged 50 to 80 (N = 10 349, N* = 9720)	Whites aged 50 to 80 (N = 9405, N* = 8577)
Age (mean; median)	64.5; 63	63.3; 62
Gender: female (%)	52.6	52.0
Education: low (%)	46.0	44.4
Education: medium (%)	24.4	27.2
Education: high (%)	29.7	28.4
Income: low (%)	25.4	30.6
Income: medium (%)	35.0	33.4
Income: high (%)	39.6	36.0
General health: excellent (%)	11.8	13.4
General health: very good (%)	35.0	29.5
General health: good (%)	30.9	31.1
General health: fair (%)	15.3	18.5
General health: poor (%)	7.0	7.4

N is the number of persons in the respective group in the dataset, N* is the subgroup with positive sampling weight. All data are population weighted.

Table 2. Self-reported health conditions and health state variables, by education and income, in the USA and England

Health condition	Education										Income							
	USA					England					USA				England			
	N	Low	Medium	High	Total	N	Low	Medium	High	Total	Low	Medium	High	Total	Low	Medium	High	Total
Diabetes	9715	16.3	14.9	13.6	15.2	8568	9.5	8.4	7.0	8.6	20.2	16.5	11.7	15.2	10.6	8.6	7.6	8.6
Hypertension	9713	51.9	50.7	41.4	48.2	8570	37.8	34.0	31.5	34.8	55.3	52.0	42.6	48.2	38.1	36.5	31.4	34.8
All heart disease	9713	21.8	21.2	15.6	19.5	8568	13.1	12.6	13.7	13.2	26.4	20.2	16.0	19.5	14.3	13.3	12.8	13.2
Stroke	9715	4.0	4.2	3.2	3.9	8568	3.6	2.7	2.2	3.0	5.5	4.8	2.2	3.9	3.4	3.7	2.1	3.0
Lung disease ^a	9717	13.0	10.4	5.4	10.3	8571	16.7	14.1	12.1	14.7	16.5	10.4	6.6	10.3	17.5	16.2	11.7	14.7
Cancer	9713	13.3	10.9	11.6	11.6	8569	5.0	4.9	5.4	5.0	18.9	11.1	10.2	11.6	5.0	5.3	4.6	5.0
Obesity	3981	46.5	45.3	38.3	43.5	7034	37.9	33.5	25.4	32.9	41.9	43.7	42.0	43.5	35.2	34.0	29.5	32.9

N is the number of persons with positive sampling weight and valid value for the respective problem. Family income is adjusted for family size, divided into equal income tertiles with one-third of the weighted population in each group. In the USA, education is divided into high school or less (0–12 years), more than high school but not a college graduate (13–15 years), and college or more (>=16 years). In England the education division is from a level lower than 'O-level' or equivalent (typically 0–11 years of schooling), qualified to a level lower than 'A-level' or equivalent (typically 12–13 years), and a higher qualification (typically >13 years). All data are weighted and age-standardized.

^aLung disease includes asthma in ELSA but excludes asthma in HRS.

Note: Myocardial infarction is not presented here as in the HRS data represent information on heart attacks in the past 2 years, not over the life span.

Table 3. Health state variables included in the single health scale and their threshold parameters (Thr 1–3) for the final Rasch model

Component	Additional information	Question	Split into	Thr 1	Thr 2	Thr 3
Grip strength	High, medium, low		Old	-2.488	1.790	
			Young	-1.275	2.723	
Lung function	High, medium, low		Old	-2.393	1.810	
			Young	-1.351	2.559	
Cognition	Delayed recall Immediate recall			-1.661	1.285	
			Old	-1.054	2.509	
			Young	-0.382	3.277	
	Any problems in orientation			1.846		
Memory		How would you rate your memory at the present time?		-0.895	0.716	
Balance	Tandem, semi-tandem, or side-by-side stand	Balance ^a		1.568	2.502	
Seeing		Is your eyesight excellent, very good, good, or fair using glasses or corrective lenses as usual?		-1.579	-0.277	1.733
Hearing		Is your hearing excellent, very good, good, or fair using a hearing aid as usual?	Female	-0.796	0.148	1.717
			Male	-1.352	-0.609	0.652
Energy	How much of the time during the past week ...	I feel full of energy these days (only in HRS ^b)		0.308		
		you had a lot of energy? (only in ELSA ^b)		-1.349	0.884	
		you felt that everything you did was an effort?		1.818		
		you could not get going?		1.747		
Sleep	How much of the time during the past week ...	your sleep was restless?		0.776		
Depression or sadness	How much of the time during the past week ...	you felt depressed?		2.326		
		you felt sad?		1.795		
		you were happy?		2.407		
Dizziness		Persistent dizziness or lightheadedness?		2.467		
Pain		Are you often troubled with pain?		0.620		
Incontinence	During the past 12 months ...	Have you lost any amount of urine beyond your control?	Female	1.227		
			Male	2.485		
Mobility	Do you have any difficulty with ...	walking one or several blocks?		1.258		
		sitting 2 hours?		2.066		
		getting up from a chair after sitting for long periods?		0.963		
		climbing one or several flights of stairs?		0.853	1.400	
		stooping, kneeling or crouching?		0.459		
		reaching or extending arms above shoulder level?		2.394		
		pulling or pushing large objects?		1.762		
		lifting weights?	HRS	1.764		
			ELSA Female	1.266		
			ELSA Male	2.158		
ADLs	Do you have any difficulty with ...	picking up a dime from a table?		3.303		
		dressing, including putting on shoes and socks?	HRS	2.833		
			ELSA	2.313		
		walking across a room?		3.745		
		bathing or showering?		3.053		
		eating, such as cutting up food?		4.353		
		getting in and out of bed?	HRS	3.523		
	ELSA	3.372				
IADLs	Do you have any difficulty with ...	using the toilet?		3.738		
		using maps?		3.071		
		preparing a hot meal?		3.808		
		shopping for groceries?		3.201		
		making phone calls?	Female	4.788		
			Male	4.003		
		managing money, such as bills and expenses?		4.202		

^aFor balance the response options 2 = 'ability to perform side-by-side stand but not semi-tandem stand' and 3 = 'not able to perform side-by-side stand' were collapsed.

^bThe wording as well as the response options of these two questions on energy were very different in HRS and ELSA. Therefore, they were included as separate items in the analysis.

about lifting weights, dressing and getting in and out of bed—showed DIF by country. Three showed DIF by gender: hearing, incontinence and making phone calls. Four measured tests showed DIF by age group: grip strength, lung function, immediate recall, and lifting weights in ELSA. Table 3 presents the questions and measured tests included in the metric of health and their threshold parameters for the final Rasch model. The threshold parameters provide an overview of the item difficulty. Making phone calls in females showed the highest response threshold on the logit scale (4.8) and therefore constitutes the most difficult item. Grip strength in the older age group showed the lowest response threshold (−3.8), thereby representing the easiest item.

The minimum and maximum person levels on the latent health state were −2.5 and 4.5. Figure 1 presents the respective distribution of the (transformed) person's abilities and item difficulties of the questions and measured tests along the health continuum.

Table 4 shows the regression coefficients from the linear additive model. The English have a slightly better health than the Americans. Females have poorer health than males and persons with low income and low level of education have poorer health than those with medium and high income and medium and high level of education.

Finally, Figure 2 shows the non-parametric effect of age resulting from the linear additive model. The graph

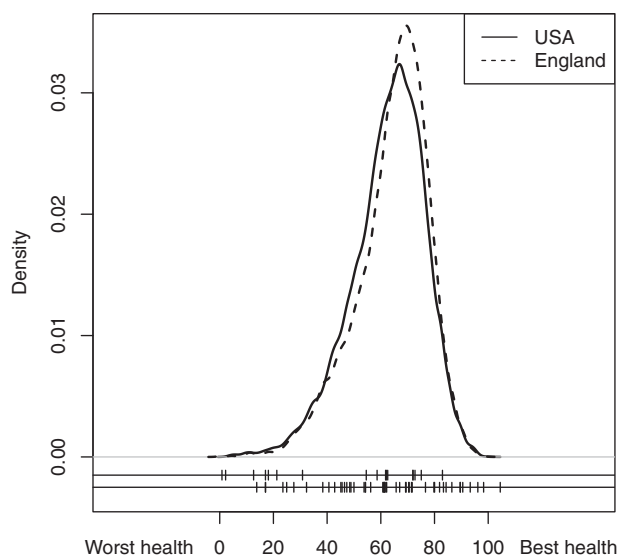


Figure 1. Density curves showing the distribution of health in the USA and England after transforming the health score to a scale from 0 to 100. The two lines below the density curves indicate the item thresholds of the measured tests (upper line) and the questions (lower line) resulting from the Rasch model, indicating the levels of health that we capture with the measured tests and questions selected.

represents the expected change in the intercept for the different levels of age. The values of the solid line can be interpreted in the same manner as the regression coefficients in Table 4, in the sense that if one were to add all of the regression coefficients that apply to one person, this sum would predict the health level for that individual. Concretely, a 50-year-old English woman with low income and education will have a health state of 60.2 (coefficient of 2.5), from age 50 to 68 this health state constantly worsens (with a coefficient of 0.1 at the age of 68) and after this 68 it worsens faster (to a coefficient of −5.8 at the age of 80).

To see whether the identified differences between the countries remain after controlling for gender, age, education and income, we calculated a second linear additive model including these covariates but excluding the survey. The residuals from this model are depicted in Figure 3 and show that their distribution is very similar and that the slight health advantage for England remains. This confirms the results of the regression model including the survey and that the differences remain when controlling for gender, age, education and income.

Discussion

The English have better health than Americans when health is assessed only by counting chronic health conditions. The English health advantage almost disappears, however, when health is assessed with a metric that integrates information about functioning domains.

Table 4. Regression coefficients, standard errors (SE), the 90% confidence interval and *p*-values resulting from the linear additive model using the health score resulting from the constructed health metric as independent variable. Age was modelled as non-parametric effect using P-splines

	Coefficient	SE	90% confidence interval		<i>p</i> -value
Intercept	57.57	0.23	57.19	57.95	<0.0001
Survey: HRS	−0.26	0.18	−0.56	0.05	0.1614
Gender: female	−1.23	0.18	−1.53	−0.93	<0.0001
Age ^a					
Income: medium	3.77	0.23	3.38	4.15	<0.0001
Income: high	7.35	0.25	6.95	7.76	<0.0001
Education: medium	3.44	0.23	3.06	3.82	<0.0001
Education: high	6.62	0.23	6.24	7.00	<0.0001

The reference categories were ELSA for survey, male for gender, low income and low education.

^aFor the effect of age see Figure 2.

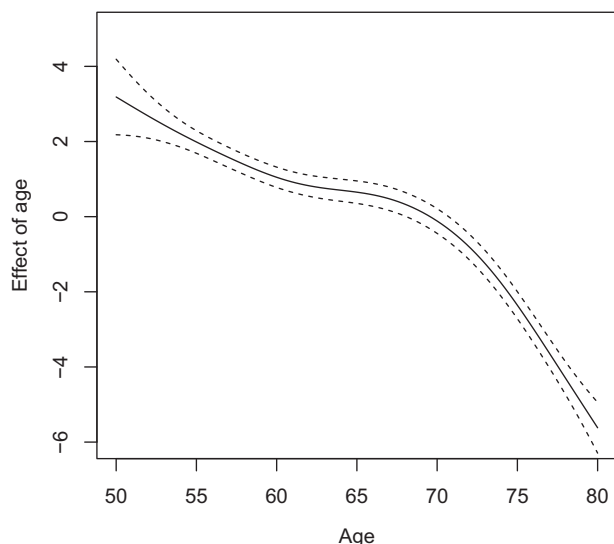


Figure 2. Effect of age resulting from the linear additive model (solid line) and pointwise 90% credible intervals (dashed lines).

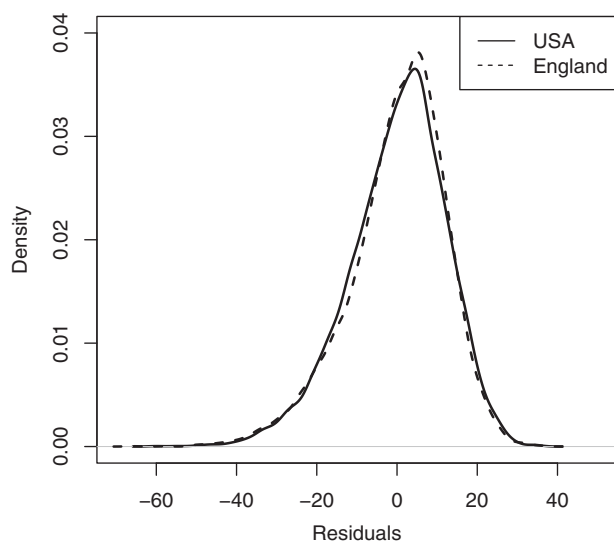


Figure 3. Density curves showing the distribution of the residuals obtained from a second linear additive model including gender, age, education and income but excluding the survey.

Strengths and weaknesses in relation to other studies

The novelty of our approach consists first in the use of IRT to calibrate in a single metric the information from two independent health surveys. The creation of the single metric made possible a cross-population comparison. Second, we used information directly collected from individuals by means of questions and measured tests. Third, we operationalized health as a continuous variable based on domains of functioning and not as a dichotomous variable (healthy/unhealthy).

IRT methods have been previously used for analyses of general population surveys. WHO constructed a metric of disability, using the data of the World Health Survey¹⁹ with similar methods to ours, for its World Report on Disability.²⁰ Thereafter, Hosseinpoor *et al.*^{21,22} used the score derived from the WHO metric of disability to investigate differences between men and women in the context of socio-demographic factors, and Chatterji *et al.*²³ used the same score to compare two populations, China and India. All these studies, however, used data of a single survey implemented in different countries. In our case, we combined the data of two independent surveys. The challenge we faced was data harmonization, so that the surveys could be analysed together. It was a time consuming exercise to identify those questions and measured tests that are common between surveys and thereafter to recode if necessary the response options of questions and to harmonize the data collection approaches of the measured tests. This could be the reason why Chan *et al.*²⁴ used only a small subset of questions when comparing the HRS and ELSA populations. They used, as we did, an IRT model, but their results and ours are not comparable because they compared both populations with a small number of questions from a limited age group (≥ 65). Future health comparisons like ours will be facilitated by recent initiatives to use the same data collection approaches in different countries. One example, that hopefully will be followed by other initiatives, is the effort to standardize surveys on ageing across the world.²⁵

To capture health we use information directly obtained from individuals by means of questions and measured tests about domains of functioning. This approach contrasts with indirect approaches for comparing the health of populations, such as health gaps and health expectancy that rely on existing population data, e.g. mortality and morbidity statistics. We opted for a direct approach because we wanted to propose a methodology based on which the information from health surveys could be utilized for comparing health not only at the population level but also at the subgroup or individual level. When health differences are found, we can specify the extent of those differences. For example, based on our metric, people between 50 and 80 with diabetes in the USA are in slightly worse health ($\bar{x} = 49.99$; SD 14.6) than the English counterparts ($\bar{x} = 51.32$; SD = 15.0) (analyses are not shown but can be obtained from the authors).

Our intention was to capture health from the perspective of the intrinsic capacity of the person, without taking into consideration whether the environment had a positive or negative influence on that capacity. Measured tests clearly reflect intrinsic attributes of the individual, but so do questions that require the respondent to focus on the

condition of her or his body, for example the question in ELSA: 'Because of a physical or health problem, do you have any difficulty getting up from a chair after sitting for long periods?'. We carefully selected this kind of question for our investigation and disregarded those in which the exclusive focus on the internal capacity was not as clear.

When we operationalize health as a continuous variable based on domains of functioning, we implicitly reveal our understanding of health as a unidimensional construct to which different human functioning domains contribute. In this investigation we took into consideration only a limited number of those domains, namely those that were in common in the two surveys. This was based on the practical consideration of data harmonization. Nevertheless, as the estimates for the position of persons and items in the health continuum (Figure 1) reveal, we successfully covered all health levels of the sample. To our knowledge this is the first time that measured tests have been combined with questions in a single metric using IRT methodology. This seems to be a good decision, since measured tests proved to be especially useful as they increased measurement precision in the lower margins of the person distribution.

The validity of our metric of health can also be derived from several results of this study. First of all, all of the well-known gradients of health—age, education and income levels—are captured by our metric of health. Second, the well-documented but variable differential in health between men and women is also captured by the metric.^{26,27} Third, although it is well known that one has to be careful when interpreting the results of single self-rated health questions,²⁸ the results reported in Table 1 are consistent with the results of our health metric that shows little difference in health between the two populations. When we group together the responses to the self-rated health question 'excellent' and 'very good', there is also little difference between English and Americans.

Although we rely on a unidimensional conceptualization of health, we do appreciate that, as commonly understood and confirmed in the literature on health status measurement,^{29,30} health can also be treated as a multidimensional construct. We also appreciate that health can be conceptualized exclusively as the absence of disease, as Banks *et al.* and many others do. It is important, however, to be aware of the conceptualization used in each investigation because it will guide methodological decisions and the interpretation of results. Our conceptualization of health guided our decision to use a unidimensional approach despite the fact that cognition, emotion, sleep, vision and hearing were also loading high in other group factors in the bifactor analyses. Our decision was then confirmed in the sensitivity analyses.

It could be seen as a limitation of the study that we did not question whether health state is a linear function of the dummy variables income, education, gender and country. Should this not have been the case, the predicted values could have gone beyond the health state range 0 to 100. We decided to assume a linear function in order to facilitate the interpretation of the results and for two further reasons that can be inferred from Figure 1. First, there are very few extreme cases and, second, that the distribution of the health metric is close to normal. Thus, we can assume that the predicted values will fall within the range between 0 and 100 and that the linear model was the most appropriate.

Meaning of the study: possible explanations and implications for clinicians and policy makers

Based on these considerations, we think that the interpretation of the results of Banks *et al.* is not that English are healthier than Americans but that, in that section of the population that is comparable between both countries and that exclude minority populations such as Black, Hispanic and Asian immigrants, the prevalence of health conditions is higher in America than in England. As we have seen in this investigation, however, in the same section of the population English and Americans do not differ in health when assessed with a metric based on domains of functioning and when controlling for age, gender, income and education. To validate our results, we calculated the amount of variance explained by the model (using adjusted R^2) including and excluding the survey. In both cases the variance explained was 17.7%, which confirms the tendency of no difference between both countries.

The most intuitive explanation of why there is little health difference, even though Americans have a higher prevalence of chronic health conditions, is that English are indeed doing worse than Americans. Langa *et al.*³¹ have found, using data from HRS and ELSA, that US adults scored better than English adults on the 24-point cognitive scale they created. To see if this holds in other domains, we compared the populations with respect to the percentage of persons having problems in specific functioning domains. The results, which are not presented in this investigation but can be obtained from the authors, show that the percentage of English having problems in memory, energy, sleep, depression or sadness, dizziness and pain is higher than corresponding percentages in America. The percentages are lower in favour of the English only in hearing, seeing and mobility. Unfortunately, the percentage in all domains cannot be compared because HRS used filters for a relatively high number of questions and the data from the whole population in each domain are not available. Nevertheless, the results for this small number of domains

already supports our explanation. This explanation raises important questions for health policy, such as: 'Is England doing a good job in prevention, but does not sufficiently take care of persons who already have health conditions?'. Further studies could shed light on this by investigating health care utilization and quality of care.

Conclusion

It is possible to construct a metric of health, based on data directly collected from individuals, in which health is operationalized as domains of functioning that describe the actual impact of health conditions on people's lives. Its application to comparing the health of Americans and the English shows that it has the potential to tackle one of the most intractable problems in international research on health, namely the comparability of health across countries.^{32,33} Additional studies are needed to further understand the consequences of our result, namely that the English are in the end not healthier than the Americans.

Funding

This study was supported by the United States National Institute on Aging's Division of Behavioral and Social Research through Interagency Agreements (OGHA 04034785; YA1323-08-CN-0020; Y1-AG-1005-01) and through research grants (R01-AG034479 and R21-AG034263).

Conflict of interest: None declared.

References

1. Banks J, Marmot M, Oldfield Z, Smith JP. Disease and disadvantage in the United States and in England. *JAMA* 2006;**295**: 2037–45.
2. Martinson ML, Teitler JO, Reichman NE. Health across the life span in the United States and England. *Am J Epidemiol* 2011;**173**:858–65.
3. Woolf SH, Aron L (eds). National Research Council and Institute of Medicine. *U.S. Health in International Perspective: Shorter Lives, Poorer Health*. Washington, DC: National Academies Press, 2013.
4. Murray CJL, Ezzati M, Flaxman AD *et al*. GBD 2010: design, definitions, and metrics. *Lancet* 2012;**380**:2063–66.
5. Salomon JA, Mathers CD, Chatterji S, Sadana R, Üstün TB, Murray CJL. Quantifying individual levels of health: definitions, concepts and measurement issues. In: Murray CJL, Evans DB (eds). *Health Systems Performance Assessment: Debates, Methods and Empiricism*. Geneva: World Health Organization, 2003.
6. Andrich D. *Rasch Models for Measurement*. Newbury Park, CA: Sage, 1988.
7. Health and Retirement Study. *About the Health and Retirement Study*. <http://hrsonline.isr.umich.edu/> (19 August 2014, date last accessed).
8. English Longitudinal Study of Ageing. *Insight Into a Maturing Population*. <http://www.elsa-project.ac.uk/> (19 August 2014, date last accessed).
9. Crimmins E, Guyer H, Langa K, Ofstedal MB, Wallace R, Weir D. *Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health and Retirement Study*. Ann Arbor, MI: Survey Research Center, University of Michigan, 2008.
10. Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982;**47**:149–74.
11. Andrich D. Application of a psychometric rating model to ordered categories which are scored with successive integers. *Appl Psychol Measurement* 1978;**2**:581–94.
12. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. *Med Care* 2006;**44**(Suppl 3): S115–23.
13. Choi SW, Gibbons LE, Crane PK. Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Soft* 2011;**39**:1.
14. Andrich D, Sheridan B, Luo G. *RUMM2020: Rasch Unidimensional Models for Measurement*. Perth, WA: RUMM Laboratory, 2002.
15. Wood SN. *Generalized Additive Models: An Introduction With R*. Boca Raton, FL: Chapman Hall/CRC, 2006.
16. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2013.
17. Reise SP. The rediscovery of bifactor measurement models. *Multivar Behav Res* 2012;**47**:667–96.
18. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res* 2007;**16**(Suppl 1):19–31.
19. World Health Organization. *World Health Survey Instruments and Related Documents*. <http://www.who.int/healthinfo/survey/instruments/en/> (19 August 2014, date last accessed).
20. World Health Organization. *Technical Appendix C: Design and Implementation of the World Health Survey*. Geneva: World Health Organization, 2011.
21. Hosseinpoor AR, Stewart Williams J, Amin A *et al*. Social determinants of self-reported health in women and men: understanding the role of gender in population health. *PLoS One* 2012;**7**:e34799.
22. Hosseinpoor AR, Williams JS, Jann B *et al*. Social determinants of sex differences in disability among older adults: a multi-country decomposition analysis using the World Health Survey. *Int J Equity Health* 2012;**11**:52.
23. Chatterji S, Kowal P, Mathers C *et al*. The health of aging populations in China and India. *Health Aff (Millwood)* 2008;**27**:1052–63.
24. 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 ELISA. *J Gerontol B Psychol Sci Soc Sci* 2012;**67**:121–32.
25. RAND Corporation. *RAND Survey Meta Data Repository: Welcome to the Survey Meta Data Repository*. <https://mmicdata.rand.org/megadata/> (4 August 2013, date last accessed).
26. Verbrugge LM. A health profile of older women with comparisons to older men. *Res Aging* 1984;**6**:291–322.

27. Gorman BK, Read JG. Gender disparities in adult health: an examination of three measures of morbidity. *J Health Soc Behav* 2006;**47**:95–110.
28. Salomon JA, Nordhagen S, Oza S, Murray CJL. Are Americans feeling less healthy? The puzzle of trends in self-rated health. *Am J Epidemiol* 2009;**170**:343–51.
29. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;**31**:247–63.
30. Prieto L, Alonso J, Lamarca R. Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. *Health QualLife Outcomes* 2003;**1**:27.
31. Langa KM, Llewellyn DJ, Lang IA *et al*. Cognitive health among older adults in the United States and in England. *BMC Geriatr* 2009;**9**:23.
32. Murray CJL, Frenk J. Health metrics and evaluation: strengthening the science. *Lancet* 2008;**371**:1191–99.
33. Meijer E, Kapteyn A, Andreyeva T. Internationally comparable health indices. *Health Econ* 2011;**20**:600–19.