



Age Effects of Frames of Reference in Self-Reports of Health, Well-Being, Fatigue and Pain

Ania Filus¹ · Doerte U. Junghaenel¹ · Stefan Schneider¹ · Joan E. Broderick¹ · Arthur A. Stone^{1,2}

Received: 4 December 2017 / Accepted: 25 September 2018 / Published online: 3 December 2018
© Springer Nature B.V. and The International Society for Quality-of-Life Studies (ISQOLS) 2018

Abstract

Self-reports of health and well-being are central for population monitoring, so it is paramount that they are measured accurately. The goal of this study was to examine the impact of age on the use of the comparison standards or frames of reference (FoRs) in self-reports of health, life-satisfaction, fatigue, and pain, and to determine if the age-health outcome associations were affected by age differences in FoRs. Respondents ($n = 2000$) selected from the U.S. general population self-rated their life-satisfaction and health outcomes. Following this, they were asked to indicate if they used any comparisons (FoRs) when making their rating and the direction of these comparisons (upward, lateral or downward). Analyses examined (a) whether age groups differed in the type and direction of FoRs, and (b) whether age patterns in health, life-satisfaction, fatigue, and pain were altered when FoRs were statistically controlled. Compared to middle-aged and older people, younger respondents were more likely to compare themselves with other people when self-rating their health and life-satisfaction. They were also more likely to use a hypothetical situation when evaluating their pain and fatigue. Younger participants used lateral comparisons less often and downward comparisons more often than middle-aged and older participants. When these age differences in FoRs were statistically controlled, the observed age patterns in self-reported health outcomes were somewhat reduced. The results show that people of different ages use different FoRs when self-reporting their life-satisfaction and health outcomes.

Key words Frames of reference · Self-reports · Health · Life-satisfaction · Fatigue · Pain · Age

✉ Ania Filus
afilus@usc.edu

¹ USC Dornsife Center for Self-Report Science, University of Southern California, Suite 405, 635 Downey Way, Los Angeles, CA 90089-3332, USA

² Department of Psychology, University of Southern California, Los Angeles, CA, USA

Introduction

Self-reports of health, subjective well-being, and symptoms are widely used in gerontology, medical, and epidemiological research because they can serve as valuable indicators of general health and health-related outcomes across the lifespan (Schnittker 2005). Two of the measures studied here, subjective health and life-satisfaction, are also considered important predictors of individuals' health trajectories and have robust links with objective physical health indices, healthcare utilization, and longevity (DeSalvo et al. 2005, 2006 for review see Diener and Chan 2011; Franks et al. 2003; Ostbye et al. 2006; Sargent-Cox et al. 2008). The other two measures studied here, pain and fatigue, are common symptoms in the general population and are associated with healthcare use, work absenteeism, and overall productivity (Cote et al. 2001; J. Ricci et al. 2006, 2007; van't Leven et al. 2010; Vasseljen et al. 2013).

A potentially important factor in studies employing these four outcome measures that has sometimes been overlooked is the comparison standards respondents use to answer questions, also known as frames of reference. Frames of reference (FoRs) are well-established in survey methodology and refer to time frame or reference group comparisons that respondents use when answering questions (Fienberg et al. 1985; Schwarz 1999). When asked about their life-satisfaction, for instance, some individuals may compare themselves to when they were younger (a time frame) or to other people of similar age (a reference group) (Kaplan and Baron-Epel 2003; Fayers, Langston, Robertson, and group 2007). Such comparisons can impact the way answers are formed.

FoRs might be a concern in self-reports if the comparison standard is not provided in survey questions. This is because different individuals or different groups of individuals may use different types of comparisons and this "natural" variation in the use of FoRs may compromise the utility of self-reports for making comparisons across people of different ages, demographic, or medical circumstances (Sargent-Cox et al. 2008; Stone et al. 2008; Ubel et al. 2005). A hypothetical example demonstrates the concern: if younger people used their impression of an "ideal self" more often than older people and if comparisons with an ideal self were associated with lower life-satisfaction, then life-satisfaction would be artificially low in younger age groups as compared to older age groups. Thus, it is important to know whether participants of different subgroups consider different FoRs when self-rating their health or life-satisfaction. This paper aims to describe frames of reference that are naturally used in self-reports of health and life-satisfaction and the associations between FoRs and age.

Evidence supporting this examination comes from several lines of research. For example, three qualitative studies showed age differences in the use of reference standards when answering the single-item self-rated health question. Peersman and colleagues (Peersman et al. 2012) found that older people were more inclined to use health comparisons with oneself in the past as a reference when responding to self-rated health questions, whereas younger people were more likely to use health behaviors (e.g., smoking, exercising) as a referent. Similar age difference were found by Krause and Jay (1994). They showed that younger respondents were more likely to use health behaviors as a reference standard when responding to the global self-rated health item, while middle aged and older participants were more likely to think about health problems they had or could have. Finally, in a study by Simon and colleagues

(Simon et al. 2005), older participants were more likely to refer to physical and functional aspects of their health when responding to global self-rated health items but younger participants were more likely to use aspects of well-being.

Furthermore, research shows that explicit manipulation of a FoR in the question wording can impact the age-pattern of health and health-related outcomes. Self-rated health items containing a FoR directing respondents to use comparisons with “other people” have been shown to reduce age differences compared to self-rated health items without an explicit FoR or those with a time-comparative FoR (Baron-Epel and Kaplan 2001; Sargent-Cox et al. 2008, 2010a). Questions explicitly asking respondents to “rate your health for someone your age” also yield less pronounced age differences than health ratings without a specific FoR (Roberts 1999; Ubel et al. 2005; Vuorisalmi et al. 2006).

Questions using different FoRs have also been found to yield differential associations with health-related outcomes, such as survival, and differential trajectories of self-rated health over time (Ferraro and Wilkinson 2015; Sargent-Cox et al. 2008, 2010a, b). Sargent-Cox and colleagues (Sargent-Cox et al. 2010a) found that a measure of self-rated health that contained neither an age-comparative FoR (comparisons to people who are younger/older/or of the same age) nor a self-comparative FoR (comparing current health to previous health) in the item wording was the best predictor of mortality in older adults. Self-rated health based on an age-comparative FoR included in the item wording improves with age in older adults (Dening et al. 1998; Seitsamo and Klockars 1997), whereas health ratings based on items with a self-comparative FoR decline with age (McCullough and Laurenceau 2004).

There is also reason to believe that the “directionality” of a FoR comparison (upward (better than oneself), lateral, or downward (worse than oneself) relative to the respondent’s perception of themselves) differs by age. Drawing on comparison theories (Albert 1977; Festinger 1950, 1954), it has been suggested that older adults are more likely to engage in downward social comparisons (i.e., comparisons with a social reference group that one perceives as inferior or downgraded) than younger adults in order to elevate their perspective on themselves and how they are doing (Heckhausen and Brim 1997; Suls et al. 1991). Temporal comparisons (comparing one’s current state with the past), in contrast, may yield less favorable self-evaluations in older adults due to age-related declines in physical health (Leinonen et al. 2001). For instance, prior research that manipulated FoRs in the item wording has reported that an age-comparative FoR in self-ratings of health yields better scores for older adults (Baron-Epel et al. 2004).

Despite the importance of FoRs in self-report research, we know very little about the FoRs - if any - that people *naturally* use when a FoR is not made explicit in a survey question, that is, when a FoR is not included in questions. There are only a few qualitative studies that explored which FoRs individuals naturally use and they were relatively small-scale investigations with non-representative samples. Kaplan and Baron-Epel (Kaplan and Baron-Epel 2003) examined self-rated health in Israeli residents and found that most respondents compared themselves to “Other People.” Fayers and colleagues (Fayers, Langston, Robertson, and group 2007) examined FoRs about health-related quality of life in patients with Paget’s disease and found that the majority of patients compared themselves to (1) how they were a year ago, (2) before they became ill, and (3) other people who are healthy. These two studies focused on health-related ratings.

In our own recent work (Junghaenel et al. 2018), we extended this prior research by querying one hundred adults from a community sample of the U.S. general population about their natural use of FoRs when rating the outcomes examined in this study: self-rated health, life-satisfaction, and two common, self-reported symptoms, pain and fatigue. In this qualitative work we found that participants reported three broad FoR categories: (1) *interpersonal comparisons* (references to other people), (2) *historical comparisons* (references to an earlier time in life or an important event in the past), and (3) *imaginary comparisons* (references to an imaginary situation). We use these categories in the current study.

The goal of the present study was to examine whether or not people of different age groups in the general U.S. population vary in their use of FoRs when responding to health-related outcomes. The study aims were: (i) to examine the prevalence of FoRs used by respondents of different age groups for self-rated health, life-satisfaction, pain, and fatigue; (ii) to evaluate whether the natural use of FoRs (type and direction of comparisons) differs by outcome and age group; (iii) to examine whether the differential use of FoRs by age (including type and direction of comparisons) could potentially affect the observed age-patterns of health, life-satisfaction, pain, and fatigue.

Methods

Participants and Procedure

Participants ($n = 2000$) were recruited from a U.S. national Internet panel of about one million households, hosted by Survey Sampling International (SSI). The opt-in panel consists of people who volunteered to periodically participate in Internet surveys for which they receive modest compensation. Invitations to participate in the study were sent to panelists in three age groups (age: 21–45 years, 46–64 years, and 65–85 years, $n = 648$ –683 in each group) until the targeted sample size of 2000 participants was reached. Recruitment strategies for the study also included an equal proportion of female and male participants and racial/ethnic demographics based on the 2010 Census. Inclusion criteria did not include specific health conditions.

Participants completed the questionnaire online and were presented with one item at a time. The FoR checklist consisted of a computer-generated branching system that was programmed by the survey host SSI. Each participant was randomly assigned to two of the four outcome variables (i.e., health, life-satisfaction, fatigue, and pain). The checklist first presented the item for one of the two selected study outcomes and asked participants to provide their self-report rating. This was followed by an open-ended question about what participants were thinking about or which comparisons they were making when answering the survey item. Next, participants were presented with a list of specific FoRs (derived from our prior qualitative work) and were asked to indicate which of these, if any, they used when making their self-report rating (“When you answered the health question, did you make any of the following comparisons?”). The list included references to other people, i.e. *Interpersonal comparisons* (“I compared myself with another person or other people”), past events, i.e. *Historical comparisons* (“I made a comparison with how I was some time ago”), a hypothetical situation, i.e. *Imaginary comparisons* (“I thought about how I would feel if something about me or

my life were different”) as well as an option that allowed them to say that they were not thinking of any of the listed FoRs (“I did not think about any of the above when answering the question”). If a participant indicated that s/he did not think of any of these three FoRs, the checklist moved on to the next outcome variable. Participants could check all options that applied and were later asked to select the one (if multiple FoRs were endorsed) that was “most important” to them when they were thinking about their answer to the survey item. Participants were then asked to think about their selected single or most important FoR and rated whether this FoR contained an upward, lateral, or downward comparison. For example, when participants selected that they made *Interpersonal comparisons*, the subsequent question on the internet checklist asked them to rate whether this person/these people were 1) better off (had better health, greater life-satisfaction, less fatigue, or less pain), 2) similar to the participant (had similar health, life-satisfaction, fatigue, or pain) or 3) were worse off (worse health, lower life-satisfaction, more fatigue, or more pain) than the participant. These steps were repeated for both outcome variables that were assigned to the participant. The online branching system also administered other survey questions not reported in the present study. For example, these questions included asking participants what specific types of past events they were thinking about in their *historical comparisons*, whether participants were thinking about one or more people and what their characteristics were in their *interpersonal comparisons*, and what aspects of themselves or their lives participants were thinking about in their *imaginary comparisons*. Information about the FoR checklist is available from the authors.

Measures

The self-report outcomes were derived from the following sources. For all of the selected questions, none had a specific FoR in the standard item wording (i.e., specifying a comparison standard for the question). For subjective health, the global health item from the SF-36v2 (Ware Jr. and Sherbourne 1992; J. E. Ware et al. 2000) was used (“In general, would you say your health is...”; response options: excellent, very good, good, fair, poor). For life-satisfaction the World Values Survey (www.worldvaluessurvey.org) life-satisfaction question was applied (“All things considered, how satisfied are you with your life as a whole these days?”; response options ranging from 1 = completely dissatisfied to 10 = completely satisfied). The fatigue question was taken from the Brief Fatigue Inventory (Mendoza et al. 1999) (“Please rate your fatigue (weariness, tiredness) by circling the one number that best describes your usual level of fatigue”; response options ranging from 0 = no fatigue to 10 = fatigue as bad as you can imagine). The pain question was selected from the Brief Pain Inventory (Cleeland 1994) (“Please rate your pain by circling the number that best describes your pain on the average”; response options ranging from 0 = no pain to 10 = pain as bad as you can imagine).

Data Analysis

For the analyses each participant’s most salient FoR was used, which included: *Interpersonal*, *Historical* and *Imaginary comparisons* or *None of the FoRs*. If a participant checked multiple FoRs, the one selected as “most important” was used. Overall there were only 340 instances of multiple FoRs (8.5%). The analysis showed that younger and

middle-aged people were more likely to use multiple FoRs as compared to older people (younger vs. older: OR = 2.48, $p < .001$, 95%CI [1.86, 3.29]; and middle aged vs. older: OR= 1.42, $p < 0.05$, 95%CI [1.04, 1.93]). The analyses of direction of comparisons were conducted excluding the subsample of participants who reported using None of the FoRs. There were no missing data. To investigate whether the use of FoRs (type and direction) differ by outcome (health, life-satisfaction, fatigue and pain), we used a series of logistic regression models (McCullagh and Nelder 1989; Venables and Gardner 2002) followed by pairwise comparisons with Bonferroni corrections to control for Type 1 error (Agresti 2007; MacDonald and Gardner 2000). Two approaches were used to evaluate the age effects on the use of FoRs. For the analyses conducted on the data pooled across outcomes, we used a series of logistic regression models followed by pairwise comparisons with Bonferroni correction. To properly account for non-independence (each participant evaluated themselves on two outcomes hence, outcomes were clustered within individuals) Huber-White cluster-robust standard errors were applied in the analyses (Williams 2000). For the analyses conducted by each outcome domain we used a series of logistic regression models, where the predictor variable “Age” was re-coded into two dummy indicators contrasting young and middle-aged participants with older participants. The dependent variable FoR was recoded into three dummy indicators contrasting *Interpersonal*, *Historical* and *Imaginary comparisons* to the absence of these FoRs (*None of the FoRs*). The second dependent variable, direction of comparison, was also recoded into two dummy variables contrasting *downward* and *lateral* comparisons to the *upward* comparisons. We checked for possible confounding effects of demographic variables (gender, SES and marital status). Inclusion of covariates did not change the substantive interpretation of the results; therefore, the final regression analyses were conducted using only participants’ age as the predictor. These analyses were conducted using R statistical software.

To investigate how the age differences in the use of FoRs affected the observed age-differences in ratings of health, life-satisfaction, fatigue and pain, we conducted two sets of univariate ANOVAs in STATA that examined: (i) age differences in each outcome variable without any covariates (i.e., testing the main effect of age groups), and (ii) age differences in each outcome holding differences in FoRs constant (i.e., testing the main effect of age groups controlling for FoR categories). The age effects obtained from models i) and ii) were statistically compared using ‘seemingly unrelated equation’ models (Baum 2006; Srivastava and Giles 1987), in which the results from separate models with different sets of explanatory variables are combined into a single model with a joint variance-covariance matrix. This makes it possible to directly compare parameter estimates (i.e., age effects) across the different models.

Results

Participants

Participants had a mean age of 53.01 years ($SD=16.0$). The gender distribution was 52.6% female. Most participants were married or living as married (57.7%). Our sample was predominantly White (80.2%), non-Hispanic (86.7%) and educated (81% with college education or more). Table 1 presents the demographic characteristics.

Table 1 Demographic characteristics of the study sample

	Young (21-45 years)	Middle-Age (46-64 years)	Older (65-85 years)
Age, mean (SD, n)	33.0 (6.3, n =648)	55.2 (5.2, n =669)	69.9 (4.2, n =683)
Gender, %			
Female	44.6%	58.0%	54.8%
Male	55.4%	42.0%	45.2%
Marital Status, %			
Married or Living as Married	61.1%	54.3%	57.8%
Widowed	1.1%	4.3%	14.9%
Separated	1.4%	2.2%	1.2%
Divorced	3.4%	20.2%	17.1%
Never Married	33.0%	19%	8.9%
Employment Status, %			
Employed	78.1%	58.6%	19.5%
Out of Work	5.5%	7.9%	1.9%
Homemaker	6.9%	7.0%	2.5%
Student	5.4%	0.4%	0.1%
Retired	1.4%	16.7%	75.3%
Unable to Work	2.6%	9.3%	0.7%
Education, %			
Education up to High School Graduate	15.3%	20.1%	21.4%
Some College	22.2%	32.7%	30.2%
College Graduate	41.8%	33.3%	31.3%
Graduate Degree	20.6%	13.7%	17.1%
Race, %			
Caucasian	75%	79.5%	85.7%
African-American	11.9%	15.2%	10.8%
Asian	6.9%	4.3%	2.8%
Native Hawaiian or Other Pacific Islander	1.1%	0.1%	0.1%
American Indian or Alaska Native	3.1%	1.8%	0.9%
Other	5.4%	2.1%	1.0%
Ethnicity, %			
Hispanic or Latino	25.3%	7.3%	7.8%
Non-Hispanic	74.7%	92.7%	92.2%

Prevalence of FoRs Overall and Across Outcomes

Overall, slightly over one-third of respondents (36.7%) did not use any of the FoRs in the questionnaire. For those who reported using one of the FoRs, *Historical comparisons* were most frequent (35%), followed by *Imaginary* (15.1%) and *Interpersonal comparisons* (13.2%).

To examine the differences in the use of FoRs by outcome, we applied a series of logistic regression models. Table 2 presents marginal means and the summary of pairwise comparisons, where each column of the table represents a separate logistic regression analysis. For *Interpersonal comparisons*, the main effect of Outcome was significant ($\chi^2(3) = 84.25, p < 0.001$). *Interpersonal comparisons* were used more frequently for health as compared to the other three outcomes, which did not differ from each other. There was also a significant effect of Outcome for *Historical comparisons* ($\chi^2(3) = 10.3, p < 0.05$). This FoR was used somewhat more frequently for fatigue as compared to health and life-satisfaction, but the frequency of *Historical comparisons* for fatigue and pain was not significantly different. For *Imaginary comparisons* the main effect of Outcome was also significant ($\chi^2(3) = 93.71, p < 0.001$). The pairwise tests showed that *Imaginary comparisons* were used more frequently for life-satisfaction compared with the other three outcomes, which did not differ among themselves. Finally, there was also a significant effect of Outcome for *None of the FoRs* ($\chi^2(3) = 33.07, p < 0.001$), which was reported most frequently for pain as compared to health and life-satisfaction. The frequency of *None of the FoRs* did not differ between pain and fatigue.

Do Young, Middle-Aged and Older People Report FoRs Differently?

In the first step we examined differences between the age groups in the use of FoRs using the data pooled across the four outcome domains. Table 3 presents marginal means and the summary of pairwise comparisons. For *Interpersonal comparisons*, the main effect of Age was significant ($\chi^2(2) = 114.67, p < 0.001$). Younger participants used *Interpersonal comparisons* more frequently than did middle-aged and older participants. For *Imaginary comparisons* the main effect of Age was also significant ($\chi^2(2) = 18.35, p < 0.001$). Younger participants made *Imaginary comparisons* more frequently than did older participants, whereas there were no significant age differences between middle-aged and older participants. Finally, there was also a significant effect of Age for *None of the FoRs* ($\chi^2(2) = 114.22, p < 0.001$). Younger participants used *None of the FoRs* less frequently than middle-aged and older participants who did not differ from each other. No significant age differences were found for *Historical comparisons* ($\chi^2(2) = 0.18, p = 0.91$).

Table 2 The use of FoRs by outcome

Outcome	FoRs			
	Interpersonal Comparisons	Historical Comparisons	Imaginary Comparisons	None of the FoRs
Health ($n = 1027$)	.22b	.33a	.11a	.34a
Life-satisfaction ($n = 1003$)	.13a	.32a	.24b	.31b,a
Fatigue ($n = 990$)	.10a	.38b	.13a	.39c,a
Pain ($n = 980$)	.10a	.35a,b	.12a	.43c

The numbers represent the distribution of a given FoR across domains (marginal means). Within-column comparisons: rows with the same subscript do not differ from one another at $p < 0.05$. Bonferroni corrections were applied

Table 3 The use of FoRs by age across outcome

Age Group	Interpersonal Comparisons	Historical Comparisons	Imaginary Comparisons	None of the FoRs
Young (<i>n</i> = 1302)	.22b	.35a	.18a	.25a
Middle (<i>n</i> = 1338)	.08a	.35a	.15a,b	.42b
Old (<i>n</i> = 1366)	.10a	.35a	.12b	.43b

The numbers represent the distribution of a given FoR across age groups (marginal means). Column comparisons: rows with the same subscript do not differ from one another at $p < 0.05$. Bonferroni corrections were applied

Table 4 Effects of age on the use of FoRs by outcome

	B	SE	Wald test χ^2	df	p	95% CI for B	OR
Health							
Interpersonal comparisons							
Young	0.41	0.18	5.4	1	< 0.05	0.065 – 0.763	1.51
Middle	-0.30	0.20	2.4	1	0.12	-0.686 – 0.082	0.74
Historical comparisons							
Young	0.10	0.16	0.34	1	0.56	-0.224 – 0.416	1.10
Middle	0.18	0.16	1.2	1	0.26	-0.136 – 0.495	1.19
Imaginary comparisons							
Young	0.78	0.25	9.7	1	< 0.001	0.287 – 1.266	2.17
Middle	0.23	0.27	0.72	1	0.39	-0.300 – 0.761	1.36
Life-satisfaction							
Interpersonal comparisons							
Young	0.89	0.24	13.7	1	< 0.001	0.419 – 1.364	2.44
Middle	-0.15	0.28	0.3	1	0.58	-0.706 – 0.398	0.86
Historical comparisons							
Young	0.13	0.17	0.59	1	0.44	-0.198 – 0.454	1.14
Middle	0.03	0.16	0.04	1	0.84	-0.289 – 0.357	1.04
Imaginary comparisons							
Young	0.22	0.18	1.5	1	0.22	-0.135 – 0.584	1.25
Middle	0.26	0.18	2.1	1	0.14	-0.090 – 0.615	1.30
Fatigue							
Interpersonal comparisons							
Young	1.70	0.31	30.2	1	< 0.001	1.088 – 2.294	5.54
Middle	0.03	0.39	0.01	1	0.94	-0.725 – 0.788	1.03
Historical comparisons							
Young	-0.04	0.16	0.07	1	0.79	-0.354 – 0.270	0.96
Middle	-0.10	0.16	0.42	1	0.52	-0.415 – 0.209	0.90
Imaginary comparisons							
Young	0.48	0.23	4.3	1	< 0.05	0.027 – 0.938	1.62
Middle	0.18	0.24	0.55	1	0.46	-0.295 – 0.657	1.20
Pain							
Interpersonal comparisons							
Young	1.37	0.27	26.1	1	< 0.001	0.847 – 0.390	3.95
Middle	-0.29	0.35	0.72	1	0.4	-0.985 – 0.390	0.74
Historical comparisons							
Young	-0.07	0.16	0.19	1	0.67	-0.392 – 0.251	0.93
Middle	0.02	0.16	0.02	1	0.89	-0.296 – 0.341	1.02
Imaginary comparisons							
Young	0.60	0.25	5.7	1	< 0.05	0.106 – 1.097	1.82
Middle	0.40	0.26	2.4	1	0.12	-0.105 – 0.914	1.49

B = unstandardized regression coefficient, OR = Odds Ratio, SE = Standard Error, CI = Confidence Interval. The outcome variable (FoRs) was dummy coded with absent FoR as a reference category. The predictor variable (Age group) was also dummy coded with older age as a reference category

In the next step we investigated age differences in the use of FoRs for each outcome separately. Results from the logistic regression models are presented in Table 4 and Fig 1. The analyses showed that for health younger participants were 1.5 times more likely than older participants to make *Interpersonal comparisons* (OR=1.51, $p < 0.001$) and 2.2 times more likely than older participants to use *Imaginary comparisons* (OR=2.17, $p < 0.001$). No significant age effects for references to *Historical comparisons* for the health were found. For life-satisfaction younger participants were almost 2.4 times more likely than older participants to use *Interpersonal comparisons* (OR=2.44, $p < 0.001$). No

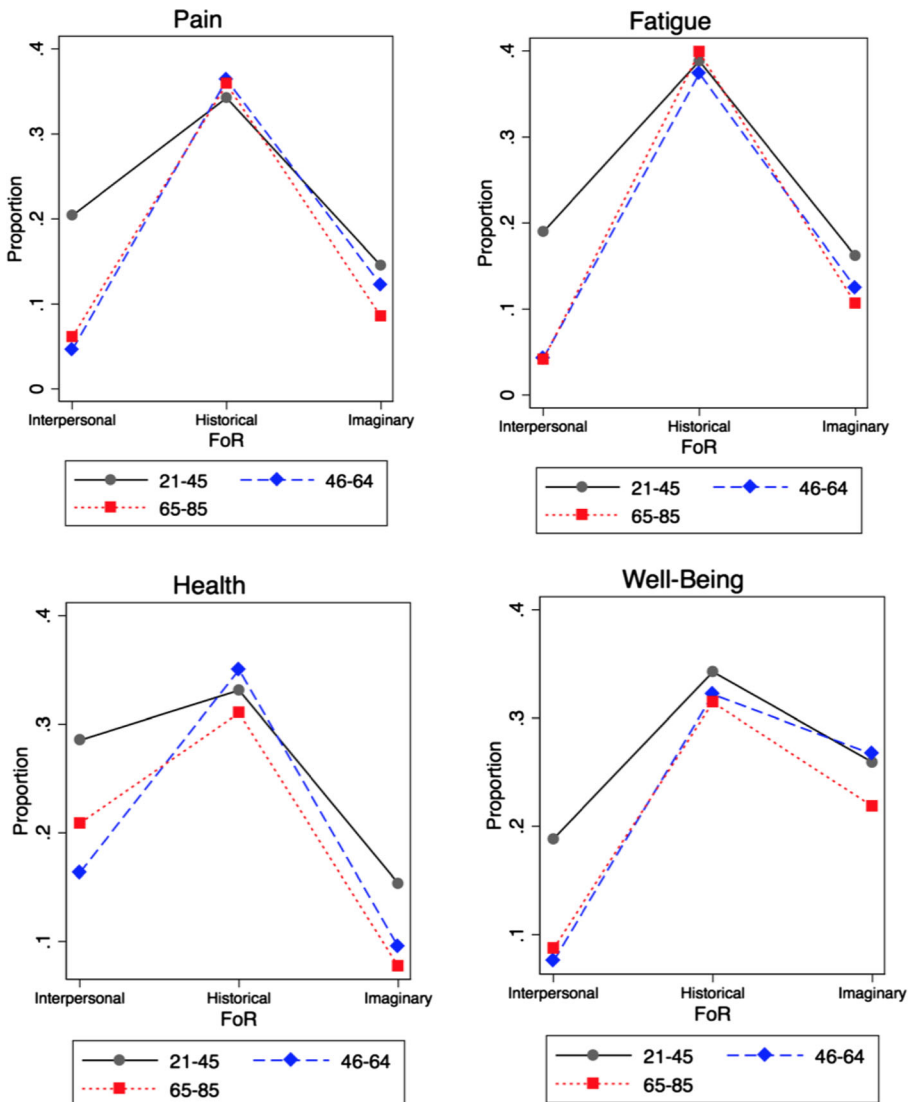


Fig 1 The use of FoRs (Interpersonal, Imaginary and Historical Comparisons) by age group and outcome

significant age effects for *Historical* or *Imaginary references* were found. For fatigue, younger participants were over 5 times more likely than older participants to make *Interpersonal comparisons* (OR=5.54, $p < 0.001$) and 1.6 times more likely than older participants to make *Imaginary comparisons* (OR=1.62, $p < 0.05$). No significant age effects for *Historical comparisons* were found. For pain, younger participants were almost 4 times more likely than older participants to make *Interpersonal comparisons* (OR = 3.95, $p < 0.001$). Furthermore, they were almost two times more likely than older participants to make *Imaginary references* (OR=1.82, $p < 0.001$). There were no significant age effects for *Historical comparisons*. Finally, for each outcome and FoR no significant difference was found between middle-aged and older participants.

Prevalence of Downward, Lateral, and Upward Comparisons Overall and Across Outcomes

For each of the broad FoR categories (notably, with the exception of the “None of the FoRs” category) respondents were subsequently asked about the direction of their comparison (i.e., *upward*, *lateral* or *downward* comparison). Investigation of the frequency of these comparisons indicated that *upward* comparisons were most frequently used (52%), followed by *lateral* (32%) and *downward* comparisons (16%).

We were also interested if the use of *upward*, *lateral* and *downward* comparisons differs by outcome. Table 5 presents a summary of the logistic regression analyses, including marginal means and the summary of pairwise comparisons. For each of the three FoRs the effect of domain was significant (*Upward*: $\chi^2(3) = 38.82$; *Lateral*: $\chi^2(3) = 21.32$, $p < 0.001$; *Downward*: $\chi^2(3) = 18.64$, $p < 0.001$). The *upward* comparisons were used more frequently for health and pain than for fatigue and life-satisfaction. The *lateral* comparisons were used more frequently for health and pain and less frequently for life-satisfaction and fatigue. Finally, *downward* comparisons were used more frequently for health and life-satisfaction, which did not differ from each other, than for fatigue and pain, which also did not differ from each other.

Table 5 The use of upwards, lateral and downwards comparisons by outcome

	Upwards	Lateral	Downwards
Outcome			
Health ($n = 645$)	.46b	.34b	.20a
Life-satisfaction ($n = 658$)	.53a	.29a	.18a
Fatigue ($n = 575$)	.62a	.26a	.12b
Pain ($n = 538$)	.46b	.39b	.15b

The numbers represent the distribution of a given FoR across domains (marginal means). Column comparisons: rows with the same subscript do not differ from one another at $p < 0.05$. Bonferroni corrections were applied

Do young, middle-aged, and older people differ in their reporting of upward, downward, and lateral comparisons?

We tested for the effects of age on the use of *upward*, *lateral*, and *downward* comparisons using data pooled across four outcomes and logistic regression models. Table 6 presents the summary of the analyses. We found significant effects of age for *lateral* and *downward* comparisons (*Lateral*: $\chi^2(2) = 26.51, p < 0.001$; *Downward*: $\chi^2(2) = 15.49, p < 0.001$). The *lateral* comparisons were used less frequently and the *downward* comparisons were used more frequently by younger participants as compared to middle-aged and older participants, which did not differ significantly from each other.

Does the use of FoRs Bias Observed Age-Patterns of Health, Life-Satisfaction, Pain, and Fatigue?

Our previous analyses showed that age groups differ in the use of FoRs, both the type and direction of comparisons. We next investigated whether we obtain different age patterns in health, life-satisfaction, fatigue and pain before versus after statistically controlling for differences in FoRs. First, we investigated how the age differences in the use of broad FoR categories (i.e., *Interpersonal*, *Historical*, *Imaginary comparisons* or *None of the FoRs*) affect age-differences in ratings of health, life-satisfaction, fatigue, and pain. Table 7 presents the ANOVA results (overall F-test) for the age group differences in means before and after adjusting for FoRs (first two columns), the estimated age group means, and the χ^2 -value for the test of differences in age effects between the two models (last column). Results show that taking into consideration FoRs did not influence age differences in ratings of life-satisfaction and health ($p > .10$). However, FoRs significantly affected age differences in ratings of fatigue and pain ($p < .01$). As shown in Fig 2, for the fatigue and pain items, after accounting for the effects of FoRs the age differences are still present, however, they are somewhat reduced.

Finally, we examined if differences in the direction of comparisons affect age differences in ratings of health, life-satisfaction, fatigue, and pain. The analyses steps were the same as the ones described in the previous paragraph. Results shown in Table 7 indicate that the used direction of comparison did not significantly affect age

Table 6 The use of upward, lateral and downward comparisons by age group

Age Group	Type of comparison		
	Upward	Lateral	Downward
Young ($n = 938$)	.35a	.45a	.20a
Middle ($n = 742$)	.31a	.56b	.13b
Old ($n = 736$)	.29a	.56b	.15b

The numbers represent the distribution of a given FoR across age groups (marginal means). Column comparisons: rows with the same subscript do not differ from one another at $p < 0.05$. Bonferroni corrections were applied

Table 7 Estimated means by age group before and after adjustment for foers and direction of comparisons, and tests for differences in age effects between the two models

	F-test		Estimated means			Difference in age-effect between models
	<i>dfs</i>	<i>F</i>	Young	Middle	Older	χ^2 (2)
Health (1-5 scale)						
Unadjusted	2,	17.98***	3.67	3.23	3.32	
	10-24					
Adjusted for FoRs	2,	18.68***	3.67	3.24	3.30	3.30
	10-21					
Adjusted for direction of comparison	2,	21.26***	3.68	3.25	3.29	3.75
	10-20					
Life-satisfaction (1-10 scale)						
Unadjusted	2,	9.04***	7.50	6.90	7.40	
	10-00					
Adjusted for FoRs	2,997	9.14***	7.53	6.91	7.36	3.85
Adjusted for direction of comparison	2,996	8.20***	7.47	6.94	7.38	1.26
Fatigue (0-10 scale)						
Unadjusted	2,987	8.32***	4.60	4.13	3.81	
Adjusted for FoRs	2,984	4.09*	4.44	4.21	3.89	13.63**
Adjusted for direction of comparison	2,983	6.19**	4.52	4.19	3.84	5.39
Pain (0-10 scale)						
Unadjusted	2,977	10.85***	3.96	3.72	2.97	
Adjusted for FoRs	2,974	7.15***	3.77	3.80	3.08	15.52***
Adjusted for direction of comparison	2,973	7.92***	3.81	3.79	3.06	13.42**

* $p < .05$, ** $p < .01$, *** $p < .001$

differences in ratings of life-satisfaction, health, and fatigue ($p > .10$), but it significantly affected the age differences in pain ratings ($p < .01$). Looking at Fig 2 we see that after accounting for the effects of FoRs, the age differences in the ratings of pain are still present, however, they are reduced.

Discussion

Self-reported survey data is used to monitor the health and well-being of the population by both governments and scientists. For some of these outcomes, self-reports are the only practical way to acquire the information, e.g., pain, fatigue, and life-satisfaction. As with any measurements that are central for population monitoring and decision-making, self-reports need to be measured as precisely as possible. Despite the

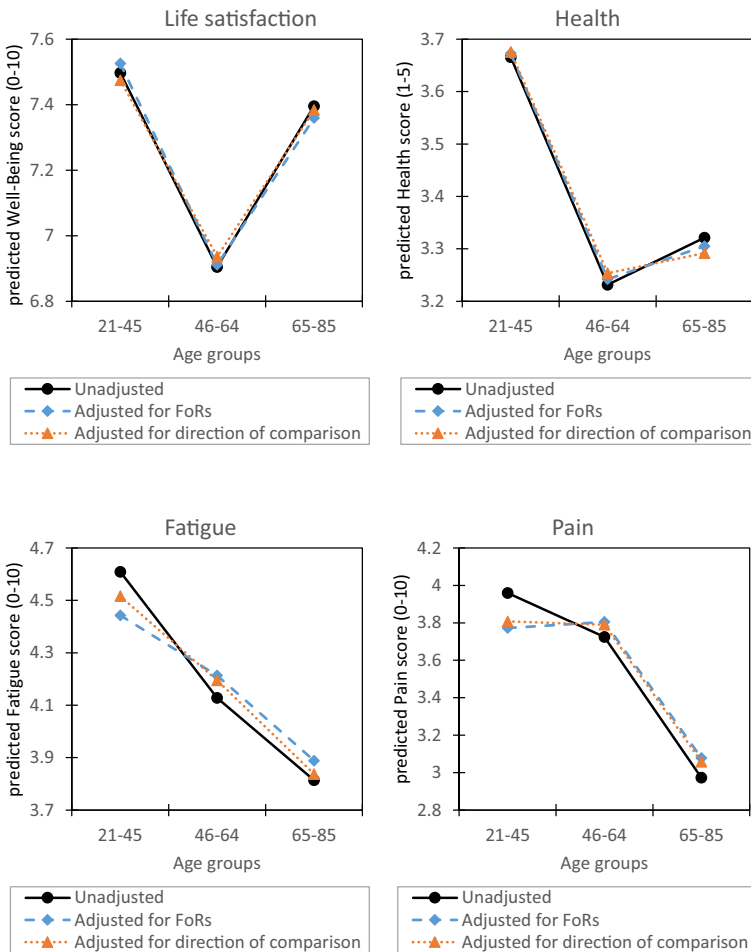


Fig 2 Estimated age patterns before and after adjustment of FoRs (Interpersonal, Historical, and Imaginary Comparisons) and direction of comparisons (Upward, Downward, Lateral) by outcome

widespread use of self-reports, knowledge of the processes underlying people’s evaluation of their general health or life-satisfaction is limited (Kaplan and Baron-Epel 2003; Simon et al. 2005; Peersman et al. 2012). In this study we focused on the comparison standards or FoRs people use when responding to survey questions about their health, life-satisfaction, pain and fatigue, a topic for which there are only a handful of studies (Kaplan and Baron-Epel 2003; Fayers, Langston, Robertson, and group 2007; Jungthaenel et al. 2018). We examined the use of FoRs in questions that do not explicitly provide instructions about reference frames, because the field lacks empirical evidence regarding the age differences in the natural use of FoRs and whether or not such differences - if observed - affect observed age patterns for the outcomes. The findings of this study make several contributions to the exiting literature. These are outlined below.

The first contribution of this study is that we found that in the U.S. general population most participants used three broad-defined FoRs: comparisons to other

people (*Interpersonal comparisons*), to past events (*Historical comparisons*), or to a hypothetical situation (*Imaginary comparisons*). This is consistent with our previous research (Junghaenel et al. 2018) that showed that participants mostly used these three broad categories when evaluating their health and life-satisfaction. In this study the most frequently used FoR was a comparison concerning past events. In fact, *Historical comparisons* were used almost twice as frequently as the other two FoRs investigated in this study, namely *Interpersonal* and *Imaginary comparisons*. Further, we also showed that slightly over one-third of respondents reported not using any of the FoRs specified in the survey. This is comparable to our previous qualitative research (Junghaenel et al. 2018) where 20–40% of respondents reported not using FoRs when self-evaluating their health or life-satisfaction. This is interesting and calls into question whether or not there are implicit FoRs that are not accessible to individuals or if such comparisons are truly not used at all. We recommend that future studies investigate this issue further using mixed-methods approaches.

The second contribution of this study is that it showed that the use of frames of reference differs by the outcome measure. Namely, for self-rated health the dominating FoR was *Interpersonal comparisons*, for life-satisfaction it was *Imaginary comparisons*, whereas for pain and fatigue respondents mostly used *Historical comparisons*. This is consistent with previous research suggesting that respondents most often make comparisons to other people or past events when self-rating their health (e.g., Kaplan and Baron-Epel 2003; Fayers et al. 2007; Peersman et al. 2012). These findings also align with our previous qualitative work (Junghaenel et al. 2018) that showed that references to a hypothetical situation were mainly reported for ratings of life-satisfaction.

The third contribution is that we found that there were significant and substantial age differences in the reporting of FoRs. Younger participants used *Interpersonal comparisons* more often than the other two age groups, regardless of the domain being assessed. One possible explanation here is that younger people tend to have more exposure to various social groups in their daily lives than the middle-aged and older people, which could trigger social comparisons. Carstensen (1992) has shown that size of social networks reduces gradually with age. Indeed, a recent meta-analysis (Wrzus et al. 2013) suggests that there is evidence of age differences in the size and composition of social networks, with the peak in young adulthood and a steady decline in middle-age and older adults. This finding does not align with theoretical positions proposing that older people, not younger ones, use interpersonal comparisons more frequently as a strategy that protects them from negative effects of aging (e.g., Baron-Epel and Kaplan 2001; Albert 1977; Festinger 1950, 1954). This literature suggested that by comparing themselves with other people who are doing worse than them, older people manage to maintain positive well-being as evident in self-reports, a phenomenon known as the *paradox of aging* (Mather 2012, Zhou et al. 2017). Nevertheless, these theoretical considerations have not been comprehensively examined. The association between age and the use of *Interpersonal comparison* in self-reports and possible explanations of these differences require further research.

Furthermore, we found that younger people used *Imaginary comparisons* somewhat more frequently than middle-aged and older participants when self-reporting their fatigue and pain. To date there is no empirical evidence regarding the age differences in the natural use of this FoR. However, these age differences may be understandable

when one considers that in this study more frequent use of *Imaginary comparisons* by younger people related to self-reports of one's pain and fatigue. Younger people, on average, may have less experience with chronic health conditions or fatigue as compared with other age groups. Epidemiological studies indicate that people in their 40s and 50s most often experience chronic fatigue (e.g., Santhouse et al. 2010) and that there is generally a higher rate of chronic pain in older age groups (e.g., Docking et al. 2011). Hence, the use of hypothetical situations may be a useful reference standard in self-reporting fatigue and pain among younger respondents.

Another noteworthy contribution of this study is that we also investigated the direction of the comparisons (i.e., *upward*, *lateral* or *downward*) used by respondents in self-reports. Our findings were that overall *upward* comparisons were used most frequently and *downward* comparisons least frequently. We also found that the use of the direction of comparisons depended on the outcome being assessed. *Upward* comparisons were used most often for life-satisfaction and fatigue, *lateral* comparisons were used most often for health and pain, whereas *downward* comparisons were most frequent for health and life-satisfaction. No previous studies have examined the effects of outcome domain being evaluated on the use of the direction of comparisons in self-report surveys.

The results also revealed significant age effects in the direction of comparisons used for self-reported fatigue: younger participants used *lateral* comparisons less often and *downward* comparisons more often than middle-aged and older participants, whereas middle-aged and older participants did not differ. This finding again is inconsistent with prior research suggesting that older people use downward comparisons more frequently than younger respondents to self-enhance their perspectives on themselves and their overall health (Heckhausen and Brim 1997; Suls et al. 1991). These results suggest that self-enhancement via downward comparisons does not explain the *paradox of aging* and hence emphasize the need to investigate other cognitive or psychometric processes responsible for this effect.

The final contribution of this study is that we found that age differences in the use of FoRs slightly impacted the age patterns for self-reported fatigue, and pain. The overall effect of naturally used FoRs is that they magnify the existing age differences in self-reports of pain, and fatigue, but do not change their direction or pattern. This is somewhat surprising given the existing literature that suggests that explicit manipulations of a FoR in the item wording can impact the age pattern of an outcome such as self-rated health (e.g., Baron-Epel and Kaplan 2001; Sargent-Cox et al. 2008, 2010a), or even the pattern of changes in self-rated health over time (e.g., Sargent-Cox et al. 2008, 2010a, b). We recommend that future research should examine the impact of natural FoRs versus FoR specified in questions to resolve this inconsistency.

There are limitations of the study that should be considered in interpreting the findings. The present study used the representative U.S. population, however, it consisted predominantly of White and highly educated respondents. This is often the case for studies using internet survey samples (e.g., Hays et al. 2015). Future studies should include more diverse samples to address whether the findings obtained in this study are consistent among respondents of different demographic composition. Moreover, this study focused on three FoRs, namely *Interpersonal*, *Historical*, and *Imaginary comparisons*. This choice was based on our previous research that showed that participants mostly used these three broad categories when evaluating their health and

life-satisfaction (Junghaenel et al. 2018). It is quite possible there may be other comparison standards that people naturally use when self-rating their health or life-satisfaction that we did not include in this study. Finally, it should be noted that in the study design we only used one question for each outcome domain; however many self-reports of health or life-satisfaction are based on multiple-item measures. It remains an open research question if people use the same comparison standard for each item in a multiple-item scale.

The limitations notwithstanding, this study adds to the existing literature on the effects of aging on self-report processes. The results revealed significant differences between younger participants and other age groups in the natural use of FoRs in self-reports of health and life-satisfaction. Specifically, our study found that age differences relate not only to the types of FoR (i.e., *Interpersonal, Historical, and Imaginary*) but also the direction of these comparisons (i.e., *upward, lateral* or *downward*). This suggests that researchers and practitioners working with self-report measures should be aware of these age differences in the natural use of FoRs. Furthermore, we showed that the age differences in the natural use of FoRs enhance the observed age differences for self-reported pain and fatigue, but do not change their direction and that they have no pronounced effect on observed age patterns in self-reported health and life-satisfaction. More research is needed on the interactions between age, type of FoRs and direction of comparisons, to provide clear guidelines to professionals working with self-report measures.

Acknowledgments This work was supported by a grant from the National Institute on Aging (R01 AG042407, PI: Arthur A. Stone, Ph.D.)

Compliance with Ethical Standards

Disclosure Arthur A. Stone is a Senior Scientist with the Gallup Organization, and a consultant for Adelphi Values, inc., and Precision Health Economics.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2nd ed.). New York: John Wiley & Sons.
- Albert, S. (1977). Temporal Comparison Theory. *Psychological Review*, *84*(6), 485–503. <https://doi.org/10.1037//0033-295x.84.6.485>.
- Baron-Epel, O., & Kaplan, G. (2001). General subjective health status or age-related subjective health status: does it make a difference? *Social Science & Medicine*, *53*(10), 1373–1381.
- Baron-Epel, O., Shemy, G., & Carmel, S. (2004). Prediction of survival: a comparison between two subjective health measures in an elderly population. *Social Science & Medicine*, *58*(10), 2035–2043. [https://doi.org/10.1016/S0277-9536\(03\)00412-X](https://doi.org/10.1016/S0277-9536(03)00412-X).
- Baum, C. F. (2006). *An Introduction to Modern Econometrics Using Stata* (pp. 236–242). College Station: Stata Press.
- Carstensen, L. L. (1992). Social and emotional patterns in adulthood: support for socioemotional selectivity theory. *Psychology and Aging*, *7*(3), 331–338.
- Cleeland, C. (1994). Pain assessment: global use of the Brief Pain Inventory. *Annals of Academic Medicine Singapore*, *23*, 129–138.

- Cote, P., Cassidy, J. D., & Carroll, L. (2001). The treatment of neck and low back pain - Who seeks care? Who goes where? *Medical Care*, 39(9), 956–967. <https://doi.org/10.1097/00005650-200109000-00006>.
- Dening, T. R., Chi, L. Y., Brayne, C., Huppert, F. A., Paykel, E. S., & O'Connor, D. W. (1998). Changes in self-rated health, disability and contact with services in a very elderly cohort: a 6-year follow-up study. *Age and Ageing*, 27(1), 23–33. <https://doi.org/10.1093/ageing/27.1.23>.
- DeSalvo, K. B., Fan, V. S., McDonell, M. B., & Fihn, S. D. (2005). Predicting mortality and healthcare utilization with a single question. *Health Services Research*, 40(4), 1234–1246. <https://doi.org/10.1111/j.1475-6773.2005.00404.x>.
- DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality prediction with a single general self-rated health question. A meta-analysis. *Journal of General Internal Medicine*, 21(3), 267–275. <https://doi.org/10.1111/j.1525-1497.2005.00291.x>.
- Diener, E., & Chan, M. Y. (2011). Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity. *Applied Psychology: Health and Wellbeing*, 3, 1–43.
- Docking, R. E., Fleming, J., Brayne, C., Zhao, J., Macfarlane, G. J., Jones, G. T., & Cambridge City over-75s Cohort Study, c. (2011). Epidemiology of back pain in older adults: prevalence and risk factors for back pain onset. *Rheumatology (Oxford)*, 50(9), 1645–1653. <https://doi.org/10.1093/rheumatology/ker175>.
- Fayers, P. M., Langston, A. L., Robertson, C., & group, P. T. (2007). Implicit self-comparisons against others could bias quality of life assessments. *Journal of Clinical Epidemiology*, 60(10), 1034–1039. <https://doi.org/10.1016/j.jclinepi.2007.03.005>.
- Ferraro, K. F., & Wilkinson, L. R. (2015). Alternative Measures of Self-Rated Health for Predicting Mortality Among Older People: Is Past or Future Orientation More Important? *Gerontologist*, 55(5), 836–844. <https://doi.org/10.1093/geront/gnt098>.
- Festinger, L. (1950). Informal Social Communication. *Psychological Review*, 57(5), 271–282. <https://doi.org/10.1037/h0056932>.
- Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7(2), 117–140. <https://doi.org/10.1177/001872675400700202>.
- Fienberg, S. E., Loftus, E. F., & Tanur, J. M. (1985). Cognitive aspects of health survey methodology: an overview. *The Milbank Memorial Fund Quarterly. Health and Society*, 63(3), 547–564.
- Franks, P., Gold, M. R., & Fiscella, K. (2003). Sociodemographics, self-rated health, and mortality in the US. *Social Science & Medicine*, 56(12), 2505–2514.
- Hays, R. D., Liu, H. H., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behavior Research Methods*, 47(3), 685–690. <https://doi.org/10.3758/s13428-015-0617-9>.
- Heckhausen, J., & Brim, O. G. (1997). Perceived problems for self and others: Self-protection by social downgrading throughout adulthood. *Psychology and Aging*, 12(4), 610–619. <https://doi.org/10.1037/0882-7974.12.4.610>.
- Junghaenel, D. U., Broderick, J. E., Schneider, S., May, M., Bolton, A., McCarrier, K. P., . . . Stone, A. A. (2018). Frames of Reference in Self-Reports of Health, Well-being, Fatigue, and Pain: A Qualitative Examination. *Applied Research in Quality of Life*, 31, 583–601.
- Kaplan, G., & Baron-Epel, O. (2003). What lies behind the subjective evaluation of health status? *Social Science & Medicine*, 56(8), 1669–1676.
- Krause, N. M., & Jay, G. M. (1994). What do global self-rated health items measure? *Medical Care*, 32, 930–942.
- Leinonen, R., Heikkinen, E., & Jylha, M. (2001). Predictors of decline in self-assessments of health among older people - a 5-year longitudinal study. *Social Science & Medicine*, 52(9), 1329–1341. [https://doi.org/10.1016/S0277-9536\(00\)00249-5](https://doi.org/10.1016/S0277-9536(00)00249-5).
- MacDonald, P. L., & Gardner, R. C. (2000). Type I error rate comparisons of post hoc procedures for I x J chi-square tables. *Educational and Psychological Measurement*, 60(5), 735–754. <https://doi.org/10.1177/00131640021970871>.
- Mather, M. (2012). The emotion paradox in the aging brain. *Annals of the New York Academy of Sciences*, 1251, 33–49.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- McCullough, M. E., & Laurenceau, J. P. (2004). Gender and the natural history of self-rated health: a 59-year longitudinal study. *Health Psychology*, 23(6), 651–655. <https://doi.org/10.1037/0278-6133.23.6.651>.
- Mendoza, T. R., Wang, X. S., Cleeland, C. S., Morrissey, M., Johnson, B. A., Wendt, J. K., & Huber, S. L. (1999). The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. *Cancer*, 85(5), 1186–1196.
- Ostbye, T., Krause, K. M., Norton, M. C., Tschanz, J., Sanders, L., Hayden, K., . . . Cache County, I. (2006). Ten dimensions of health and their relationships with overall self-reported health and survival in a predominantly religiously active elderly population: the cache county memory study. *Journal of the American Geriatrics Society*, 54(2), 199–209. <https://doi.org/10.1111/j.1532-5415.2005.00583.x>

- Peersman, W., Cambier, D., Maeseneer, J. D., & Willems, S. (2012). Gender, educational and age differences in meanings that underline global self-rated health. *International Journal of Public Health*, *57*, 513–523.
- Ricci, J., Chee, E., & Lorandeanu, A. (2006). Fatigue in the US workforce: Prevalence and cost of lost productive work time. *Value in Health*, *9*(3), A166–A166. [https://doi.org/10.1016/S1098-3015\(10\)64818-8](https://doi.org/10.1016/S1098-3015(10)64818-8).
- Ricci, J. A., Chee, E., Lorandeanu, A. L., & Berger, J. (2007). Fatigue in the US workforce: Prevalence and implications for lost productive work time. *Journal of Occupational and Environmental Medicine*, *49*(1), 1–10. <https://doi.org/10.1097/01.jom.0000249782.60321.2a>.
- Roberts, G. (1999). Age effects and health appraisal: a meta-analysis. *The Journals of Gerontology: Series B, Psychological Sciences and Social Sciences*, *54*(1), S24–S30.
- Santhouse, A. M., Hotopf, M., & David, A. S. (2010). Chronic fatigue syndrome. *BMJ*, *340*, c738. <https://doi.org/10.1136/bmj.c738>.
- Sargent-Cox, K. A., Anstey, K. J., & Luszcz, M. A. (2008). Determinants of self-rated health items with different points of reference: implications for health measurement of older adults. *Journal of Aging and Health*, *20*(6), 739–761. <https://doi.org/10.1177/0898264308321035>.
- Sargent-Cox, K. A., Anstey, K. J., & Luszcz, M. A. (2010a). The choice of self-rated health measures matter when predicting mortality: evidence from 10 years follow-up of the Australian longitudinal study of ageing. *BMC Geriatrics*, *10*, 18. <https://doi.org/10.1186/1471-2318-10-18>.
- Sargent-Cox, K. A., Anstey, K. J., & Luszcz, M. A. (2010b). Patterns of longitudinal change in older adults' self-rated health: the effect of the point of reference. *Health Psychology*, *29*(2), 143–152. <https://doi.org/10.1037/a0017652>.
- Schnittker, J. (2005). When mental health becomes health: age and the shifting meaning of self-evaluations of general health. *The Milbank Quarterly*, *83*(3), 397–423. <https://doi.org/10.1111/j.1468-0009.2005.00407.x>.
- Schwarz, N. (1999). Self-reports: how the questions shape the answers. *American Psychologist*, *54*, 93–105.
- Seitsamo, J., & Klockars, M. (1997). Aging and changes in health. *Scandinavian Journal of Work, Environment & Health*, *23*(Suppl 1), 27–35.
- Simon, J. G., De Boer, J. D., Joung, I. M. A., Bosma, H., & Mackenbach, J. P. (2005). How is your health in general? A qualitative study on self-assessed health. *Perceived Health*, *15*, 200–208.
- Srivastava, V. K., & Giles, D. E. A. (1987). *Seemingly unrelated regression equations models: estimation and inference*. New York: Marcel Dekker.
- Stone, A. A., Broderick, J. E., Schwartz, J. E., & Schwarz, N. (2008). Context effects in survey ratings of health, symptoms, and satisfaction. *Medical Care*, *46*(7), 662–667. <https://doi.org/10.1097/MLR.0b013e3181789387>.
- Suls, J., Marco, C. A., & Tobin, S. (1991). The Role of Temporal Comparison, Social-Comparison, and Direct Appraisal in the Elderly Self-Evaluations of Health. *Journal of Applied Social Psychology*, *21*(14), 1125–1144. <https://doi.org/10.1111/j.1559-1816.1991.tb00462.x>.
- Ubel, P. A., Jankovic, A., Smith, D., Langa, K. M., & Fagerlin, A. (2005). What is perfect health to an 85-year-old?: evidence for scale recalibration in subjective health ratings. *Medical Care*, *43*(10), 1054–1057.
- van't Leven, M., Zielhuis, G. A., van der Meer, J. W., Verbeek, A. L., & Bleijenberg, G. (2010). Fatigue and chronic fatigue syndrome-like complaints in the general population. *European Journal of Public Health*, *20*(3), 251–257. <https://doi.org/10.1093/eurpub/ckp113>.
- Vasseljen, O., Woodhouse, A., Bjorngaard, J. H., & Leivseth, L. (2013). Natural course of acute neck and low back pain in the general population: The HUNT study. *Pain*, *154*(8), 1237–1244. <https://doi.org/10.1016/j.pain.2013.03.032>.
- Venables, W. N., & Gardner, R. C. (2002). *Modern Applied Statistics with S*. New York: Springer.
- Vuorisalmi, M., Lintonen, T., & Jylha, M. (2006). Comparative vs global self-rated health: associations with age and functional ability. *Aging Clinical and Experimental Research*, *18*(3), 211–217.
- Ware Jr., J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, *30*(6), 473–483.
- Ware, J. E., Kosinski, M., & Dewey, J. E. (2000). *How to Score Version 2 of the SF-36 Health Survey*. Lincoln: QualityMetric Incorporated.
- Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*, *56*, 645–646.
- Wrzus, C., Hänel, M., Wagner, J., & Neyer, F. J. (2013). Social network changes and life events across the life span: a meta-analysis. *Psychological Bulletin*, *139*(1), 53–80.
- Zhou, L., Lu, J., Chen, G., Dong, L., & Yao, Y. (2017). Is there a paradox of aging: When the negative aging stereotype meets the positivity effect in older adults. *Experimental Aging Research*, *43*, 80–89.