

Frames of Reference in Self-Reports of Health, Well-Being, Fatigue, and Pain: a Qualitative Examination

Doerte U. Junghaenel¹ · Joan E. Broderick¹ ·
Stefan Schneider¹ · Marcella May¹ · Alicia Bolton¹ ·
Kelly P. McCarrier² · Larissa M. Stassek² ·
Sarah C. Keithly² · Arthur A. Stone¹

Received: 24 November 2016 / Accepted: 6 June 2017 / Published online: 6 July 2017
© Springer Science+Business Media B.V. and The International Society for Quality-of-Life Studies (ISQOLS) 2017

Abstract Self-reports in survey research can be affected by internal comparison standards, or Frames of Reference (FoRs), that people apply when making their ratings. The goal of this study was to determine which FoRs people naturally use when rating their health, subjective well-being, fatigue, and pain. We further examined whether FoRs varied by content domain and age group. One hundred adults from a community sample of the US general population participated in individual semi-structured qualitative interviews. Participants provided self-report ratings on two of the four content domains and were then systematically queried about FoRs. Interview responses were summarized and coded into broad FoR categories. Four broad FoR categories emerged: References to (1) *Other People*, (2) an *Earlier Time in Life*, (3) an *Important Event in the Past*, and (4) a *Hypothetical Situation*. FoRs were reported in the majority (80.5%) of responses and multiple FoRs were reported in 34% of responses. The reporting of FoRs was evident for all domains, but was more prevalent for well-being compared to pain. References to a *Hypothetical Situation* were only mentioned in the well-being and health domains. For health, references to *Other People* were more frequently reported at older compared to younger ages. Our results extend prior work by demonstrating that participants' reporting of FoRs is evident in ratings of various content domains. They further suggest that a limited number of FoRs are used and that their identification holds promise for understanding and controlling systematic group differences in FoRs.

✉ Doerte U. Junghaenel
junghaen@usc.edu

¹ USC Dornsife Center for Self-Report Science & Center for Economic and Social Research, University of Southern California, Verna & Peter Dauterive Hall, 635 Downey Way, Los Angeles, CA 90089-3332, USA

² Health Research Associates (HRA), Mountlake Terrace, WA, USA

Keywords Frames of reference · Self-report · Comparison standard · Health · Well-being · Qualitative research · United States · Symptoms

Introduction

Asking individuals to self-report on their health, subjective well-being, and symptoms is common in population and health survey research. Self-rated health, for instance, is seen as a valid index of a person's general health status (Kaplan and Baron-Epel 2003; Baron-Epel and Kaplan 2001); it is related to mortality risk and is an indicator of physical and mental health (e.g., Sargent-Cox et al. 2008; Ostbye et al. 2006), and it has been found to be a good predictor of long-term health outcomes, such as survival, healthcare utilization, and disability (Ostbye et al. 2006; Franks et al. 2003; DeSalvo et al. 2005, 2006; Jylha 2009). Similarly, self-reports of subjective well-being have been linked to various indicators of objective health status and longevity (for review see Diener and Chan 2011). Pain and fatigue are indicators of health-related quality of life and are among the most common self-reported symptoms in the general population. They have been found to be positively linked to healthcare utilization and work absenteeism, and negatively associated with overall productivity (Ricci et al. 2006, 2007; Cote et al. 2001; van't Leven et al. 2010; Vasseljen et al. 2013; Jordan et al. 2014; Efficace et al. 2013).

Although self-reports are ubiquitous in survey research, their validity relies on the assumption that respondents share similar interpretations of the content of the survey questions and the response scale used to measure it (Schwarz 1999). One concern is that individuals may use different types of comparison standards (Fienberg et al. 1985), or Frames of Reference (FoRs), while making their self-report ratings. For example, respondents may compare themselves to how they felt during an earlier time in their life, to other people, or to other situations (real or imagined). Research has shown that manipulating the FoR in the wording of survey items can impact study results. Several studies have found that an item that explicitly asks respondents to "rate your health for someone your age" yields less pronounced age differences than health ratings without a stated comparison standard (Roberts 1999; Vuorisalmi et al. 2006; Ubel et al. 2005). Self-reports of health have also been found to be differentially associated with health-related outcomes, such as survival, depending on the type of FoR specified in a survey item; for example, self-comparative health ratings have demonstrated weaker positive associations with life expectancy than age-comparative health ratings, and ratings of past health have demonstrated weaker negative associations with mortality than ratings of future health expectations (Sargent-Cox et al. 2008, 2010; Ferraro and Wilkinson 2015).

Despite the likely importance of FoRs in self-reports, there is little research about the specific FoRs that people use when a FoR is not made explicit in a survey question. Most self-report items do not specify a FoR, leaving the choice of FoR to the respondent (Sargent-Cox et al. 2008), and researchers, therefore, have no information about the particular FoRs that may have been employed. If the natural use of FoRs varies between individuals, and perhaps demographically-defined groups, this may reduce the comparability of responses across people of different ages or different

circumstances (Sargent-Cox et al. 2008; Ubel et al. 2005; Stone et al. 2008). Limited prior research has attempted to delineate FoRs that individuals naturally use in their self-reported health ratings. Kaplan and Baron-Epel (2003) interviewed Israeli residents about their health and found that most respondents compared themselves to other people. Fayers et al. (2007) examined FoRs about health-related quality of life in patients with Paget's disease. Patients were presented with a list of possible FoRs and were asked to select which one they mainly used for their rating of health-related quality of life. They found that the majority of patients compared themselves to (1) how they were a year ago, (2) how they were before they became ill, and (3) other people who are healthy.

Several important issues remain unresolved from this prior research. First, the two studies were based on limited samples that may preclude generalization to community samples in the United States (US). Fayer's et al. (2007) study was based on a patient sample and Kaplan and Baron-Epel's (2003) sample was comprised of Israeli residents. Second, these studies focused on health-related ratings, and it may be that FoRs found for health do not generalize to other self-report domains. We are not aware of any qualitative research that has examined FoRs in other content domains, such as subjective well-being and common symptoms, like fatigue and pain. This is a significant limitation in prior research on FoRs, given the widespread examination of these constructs with self-report methods in population and health survey research. In addition, we are unaware of any research that directly examines differential use of FoRs by age group in different content domains. This is important since differences in the natural use of FoRs could yield outcomes that look age-related, but may actually be FoR-related.

The present study reports on the initial qualitative step in a broader research effort to determine the prevalence and nature of FoRs for self-report questions in health, well-being, and common symptom domains in a community sample of the US general population. In the future study, participants will be presented with an online checklist of FoRs and will be asked to select the one(s) they naturally use when making their self-report rating of the specified domain. To ensure that the planned checklist covers the full breadth of FoRs, we conducted the present qualitative study to examine which FoRs people apply when rating their health, subjective well-being, fatigue, and pain.

The first aim of this study was to determine which FoRs were reported in an adult sample of US community participants. The second aim was to examine whether these FoRs were reported in all four content domains (health, well-being, fatigue, and pain). Finally, the relationship between FoRs and age is of particular interest to aging and gerontology research. We therefore explored which FoRs were most often reported by young, middle, and older age groups, first combined across all content domains and then by domain, in an effort to increase our understanding of whether age may be differentially associated with the reporting of FoRs.

Methods

Participants

A purposive, non-probability sample of 100 adults was recruited from five metropolitan areas of the US to participate in this study. Study participants were recruited through

web-based advertisements and from within a panel of study volunteers managed by Health Research Associates (HRA), a research consultancy specializing in the development and evaluation of health-related questionnaire instruments. Participants were considered eligible for inclusion in the study if they were (a) 18 years of age or older, (b) able to read, write, and speak in English, and (c) available to travel to the site of the scheduled study interview visit within their city. Individuals were excluded from the study if they reported conditions that would interfere with engaging in the study procedures, such as vision problems, severe mental illness, cognitive impairment, terminal illness, and recent major surgery. The screening form was developed such that participants self-reported on overall ability to participate (“Are you capable of completing paper questionnaires and a spoken interview?”), but over the entire screening process, the study coordinator conducting the screening also paid close attention to subject responses and comments, listening for any sign of cognitive or other forms of incapacitation that warranted review with the principal investigator for potential exclusion. At the interview visit, the trained interviewer again gauged subject capability to participate. Ultimately, across all individuals screened for this study, there were no instances in which a potential subject was deemed unfit to give informed consent or undergo the interview process. Recruitment quotas were employed to yield a sample of the US population across a range of age, race, and socioeconomic groups. Quotas were also used to ensure the inclusion of individuals with varying responses to the four content domains (i.e., ensuring approximately equal numbers of participants self-reporting good vs. poor general health, high vs. low levels of pain, etc.).

Procedure

The study was approved by the Institutional Review Board of the University of Southern California and Quorum Review IRB (Seattle, WA). Study advertisements were placed in geographically-limited online forums for each of the five metropolitan areas. Copies of the study advertisement text were also sent via email to members of HRA’s participant database (composed of individuals who previously provided consent to be contacted by HRA about study opportunities). Interested participants then completed a telephone screening to determine eligibility to participate in the study and quota group assignment.

Following the telephone screening, eligible participants were scheduled for an in-person study appointment where they provided written informed consent, completed study questionnaires, and were engaged in a qualitative interview about their use of FoRs when responding to the self-report assessments. Each participant completed self-report assessments for two of the four study domains (i.e., health, well-being, fatigue, and pain) to maximize efficiency and to minimize interview redundancy across multiple domains. The domains were randomly selected for each participant until quota groups were filled and ordering was counterbalanced to control for the possibility of spill-over effects of FoRs from the first to the second domain. This resulted in each of the four domains being discussed within approximately 50 qualitative interviews. All interviews were audio-recorded for analyses. At the end of each interview session, participants were compensated \$80 for their time and travel expenses.

Measures

The self-report questions for the four content domains were derived from the following sources. For subjective health, the global health item from the SF-36v2 (Ware et al. 2000; Ware and Sherbourne 1992) was used (“In general, would you say your health is...”; response options: excellent, very good, good, fair, poor). The well-being question was abstracted from the World Values Survey (www.worldvaluessurvey.org) (“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).

Qualitative Interviews

The HRA interviewers were each experienced in interviewing participants for the purposes of developing and validating assessments of patient-reported outcomes. An interview guide, based on the Four-Stages Model of the Survey Response Process, was developed to provide the conceptual approach for the cognitive interviews (Tourangeau 1991; Willis 2005). Within the interview guide, two primary cognitive interview methods were utilized: the “think-aloud” approach and interviewer probes (Willis 2005). The think-aloud approach presents an item and asks the respondent to verbally describe what she/he is thinking as she/he reads and interprets the question and formulates a response. In the second approach, the interviewer systematically probes the respondent with a set of queries to uncover the basis of her/his response to the item, i.e., both the response selected and a probing of the FoR(s) used. Interviewers used both approaches to elicit the information from respondents, with the think-aloud procedure preceding the probes.

Data Analysis

Data were analyzed using inductive thematic content analysis, in which themes emerge from the data through researchers’ careful interpretation and discussion (Braun and Clarke 2006). First, the qualitative interviews were reviewed to identify common FoR themes, which were grouped into broad categories of FoRs with more detailed subcategories. Individual participant answers were then coded into one or more FoR categories through an iterative process, throughout which categories were continuously refined or added. Data were coded and analyzed using ATLAS.ti software (version 7.5; Friese 2015). We present the FoR categories that were identified along with examples of participant responses.

Our first aim was to examine the frequency of each FoR across all responses and study domains. To accomplish this, we examined the 200 responses provided by the 100 participants, and we present the percentage of responses that reference each FoR. We also indicate the number of participants who contributed to the percentage since each participant rated two domains. Participant responses sometimes included more than one FoR per domain. Commonly mentioned subcategories of the broad FoR categories are also reported.

Our next aim was to explore whether the prevalence of FoRs differed between the study domains, i.e., health, well-being, fatigue, and pain. For this purpose, we examined the responses for each domain (each had about 50 responses) and present the percentage of the responses for each FoR reported within each domain. We highlight differences in FoRs between domains that are of at least medium size as indicated by an Odds Ratio (OR) of >3.0 (Cohen 1992). Our third aim was to explore whether the prevalence of FoRs differed between age groups. To do this, we first examined the percentage of responses combined across all domains that reported a FoR for each age group. We then examined the percentage of responses separately for each of the four domains by age group and highlight major differences between the age groups (OR >3.0).

Results

Sample Characteristics

Table 1 provides an overview of the participant sample ($n = 100$). The mean age of the sample was 51.2 years. Thirty-five participants were in the younger age group (18 to 45 years), thirty-three were in the middle age group (46 to 59 years), and thirty-two were in the older age group (60 years and older). Fifty-eight participants (58%) were female and fifty-six (56%) were employed (full- or part-time). Seventy-eight participants (78%) had some college education or more.

Coding of FoRs Reported by Participants

Four broad FoR categories were identified in interview responses. The first FoR was references to *Other People*. In this category participants made comparisons to another person or a group of people. Five examples follow.

...“Compared to other people my age; especially women my age—not any particular person. My health is probably “fair” within the general population, but within people my age, it’s probably “good.”...” (Domain: Health; Female, Age: 61)

...“At church I’m part of a senior connections group with a lot of older people. Compared to them I’m in pretty good health. When I compare to my daughter or son, I think “age is creeping up.”...” (Domain: Health; Male, Age: 67)

Table 1 Demographic characteristics of the participant sample ($n = 100$)

Demographic characteristics		Participants ($N = 100$)
Age (Years)	Mean (SD)	51.2 (15.4)
Age by group	18 to 45 years	35
	46 to 59 years	33
	60+ years	32
Gender	Male	42
	Female	58
Marital status	Married /Living as Married	25
	Widowed	4
	Separated	1
	Divorced	22
	Never Married	48
Highest level of education completed	Less than High School	1
	High School	21
	Some College	29
	Bachelor's Degree	26
	Graduate or Professional School	23
Current employment status	Employed full-time	38
	Employed part-time	18
	Homemaker	3
	Student	1
	Retired	18
	Out of work	17
	Unable to work	5
Socioeconomic Status (SES): annual household income	≤\$34,999	52
	\$35,000 to \$74,999	31
	\$75,000+	17
Ethnicity	Hispanic or Latino	8
Race	American Indian or Alaska Native	3
	Asian	10
	Black or African American	31
	White	54
	Other	2

...“I compared myself to my coworkers, family, and friends. I thought about what they are enduring at the present. I generally thought of people my age.”... (Domain: Pain; Male, Age: 52)

...“I thought about somebody who is homeless, and I know a lot of women going through a divorce—my life is good compared to theirs.”... (Domain: Well-being; Female, Age: 53)

...“Compared to people my age, I get tired more easily. Some of my friends get home from work and are ready to go back out in minutes but I need to take a nap before I do anything else.”... (Domain: Fatigue; Female, Age: 26)

The second FoR identified was references to an *Earlier Time in Life*. In this category participants made comparisons to how they felt during an earlier time in their life, such as how they felt when they were younger. Four examples follow.

...“Thought about when I was a little bit younger. When I was in my early 30’s through early 40’s I exercised a lot and was in very good shape....” (Domain: Health; Female, Age: 49)

...“I thought about how I was before a year ago when I started gaining weight. I was less fatigued and happier.”... (Domain: Fatigue; Female, Age: 28)

...“When you are young, you are healthy. Being 67 is very different. I thought about 5 years ago when I had fewer problems. Now I have times when I do not have any energy and I also have mood changes now....” (Domain: Health; Female, Age: 67)

...“I compared to an earlier time when I experienced more pain.”... (Domain: Pain; Male, Age: 77)

The third FoR noted in interview responses was reference to an *Important Event in the Past*. In this category, participants made note of a specific significant event in their life, such as retirement or major medical events. Four examples follow.

...“I thought about past major injuries or episodes when I had 9 or 10 level pain; trauma, accidents.”... (Domain: Pain; Female, Age: 57)

...“I compared myself to before I retired; beginning of 2015. I was very stressed out and not happy. I had health issues like headaches, anxiety, and exhaustion. I saw my doctor about once/month with a new complaint....” (Domain: Health; Female, Age: 72)

...“Compared to my best years; 1997-2002, before I got pregnant. I didn’t have any health or financial problems or stress. Things shifted in 2003 when I had a high-risk pregnancy and a death in my family....” (Domain: Well-being; Female, Age: 35)

...“I compared to pain I felt after my back surgery. The pain after surgery was awful and would have been a 10. I use this period of time as a baseline when evaluating all pain.”... (Domain: Pain; Male, Age: 49)

The fourth FoR described by participants was reference to a *Hypothetical (desirable or undesirable) Situation*. In this category, participants indicated that they thought about how they would feel if something about themselves or their life was different;

a comparison with an ideal self or situation, or an undesirable self or situation. Four examples follow.

...“I think my health could be better. I could be physically thinner, I could eat better. It’s not horrible, but not excellent. There is room for improvement. I need to eat better because I’m getting older and there is some diabetes in my family. There is definite room for improvement....” (Domain: Health; Female, Age: 49)

...“I thought of things I should have done years ago, like saving money...” (Domain: Well-being; Female, Age: 58)

...“I thought about my happy relationship, my good job which pays well. The only thing that stopped me from giving a higher number is my weight and the fact that I’m not married. Also, if I were a home-owner I’d rate that a ‘10’....” (Domain: Well-being; Female, Age: 34)

...“It mostly revolves around my financial health, because a lot of the health issues revolve around financial stability. If my financial situation was better, I would have better health insurance, I could afford better food, I could probably afford to join a fitness center, and I could do more for members of my family to help them with their basic needs....”(Domain: Well-being; Male, Age: 58)

Overall agreement between the two raters (DJ and MM) was calculated at the broad FoR level and was satisfactory (mean kappa = 0.71; kappa *Other People* = 0.81; kappa *Past Self/Past Event* = 0.76; kappa *Hypothetical* = 0.52; kappa *no FoR* = 0.75). Most disagreements occurred for participants’ references to a *Hypothetical (desirable or undesirable) Situation* in the well-being domain. For this FoR, the specific hypothetical situation was often not explicitly stated by the participant but rather described in more abstract terms (e.g., “There is room for improvement”, “I’m not where I want to be”). These more abstract descriptions of a participant’s FoR in the well-being domain led to initial disagreements by the two raters that were resolved and reconciled after a joint review and discussion of the participant’s descriptive narrative.

FoR Categories Combined Over All Domains

Across all four domains, 39 (19.5%) of the 200 participant responses indicated that no specific FoR was used when considering the item. A single FoR was reported in 93 (46.5%) responses, two FoRs were reported in 61 (30.5%) responses, and three FoRs were reported in 7 (3.5%) responses.

References to Other People References to another person or group of other people were made in 66 (33%) of the 200 responses. We also classified participants’ responses regarding whom they were thinking about into more detailed subcategories (these could be overlapping because participants could mention multiple types of other people): Members of the participant’s close social network (i.e., family and close friends) were mentioned in 42 (21%) responses, and people outside of participants’ close social networks were mentioned in 24 (12%) responses. Comparisons to people of similar age

occurred in 17 (9%) responses, comparisons to younger people occurred in 5 (3%) responses, and comparisons to older people occurred in 4 (2%) responses. Additional social comparison groups (e.g., based on occupation) were infrequently mentioned by individual respondents.

References to an Earlier Time in Life Participants stated that they thought of how they felt in the past compared to how they are feeling now in 91 (46%) responses. Participants referenced a time in the past when they felt better in 56 (28%) responses, and a time when they felt worse in 32 (16%) responses; the valence was not specified in three responses.

References to an Important Event in the Past Participants reported that they had a specific important past event in their mind in 60 (30%) responses. The most commonly reported events were health-related events, such as injuries, onset of medical conditions, and surgeries, and they were mentioned in 38 (19%) of the responses. Events related to income and occupation, such as retirement or absence of employment, were mentioned in 10 (5%) responses. Events pertaining to relationships, such as marriage, death of a loved one, or divorce, were mentioned in 11 (6%) responses.

References to a Hypothetical (Desirable or Undesirable) Situation Participants indicated that they thought about how they would feel, if something about themselves or their life was different in 19 (10%) of the responses. Participants most commonly mentioned that their rating would have been different if their income, occupation, or educational attainment was different in 10 (5%) responses, if their relationships were different in 7 (4%) responses, and if their health was different in 3 (2%) responses.

Frequency of FoRs by Domain

Next, we examined whether the types of FoRs differed by domain. As shown in Table 2, FoRs were reported in all four domains, but the distribution of FoR categories varied across the domains. The pain domain had the highest number of responses where no specific FoR was reported (28%) compared to the well-being domain that had the fewest responses (10%) where no specific FoR was reported (OR = 3.5). Multiple FoRs were reported most frequently in the well-being domain compared to the pain domain, which had the lowest frequency of multiple FoRs (OR = 3.7).

References to *Other People* were reported at the highest frequency in the health domain compared to the pain domain, which had the lowest frequency (OR = 3.1). The rates of references to an *Earlier Time in Life* and *Important Event in the Past* were generally comparable across the domains. References to a *Hypothetical Situation* were mainly reported for ratings of well-being and to a much lesser extent for health (OR = 11.8), whereas they were not reported at all for ratings of fatigue or pain.

Frequency of FoRs by Age Group Combined Over All Domains

We next compared the frequency of FoRs across all domains between the young, middle, and older age groups. No major age differences were evident (all ORs < 3.0). Across age groups, the frequency of no specific FoR ranged from 17% to 22%, and the

Table 2 Frequency of different frames of reference (FoR) by domain across participants

	Health ^a (n = 50)	Well-being ^a (n = 51)	Fatigue ^a (n = 49)	Pain ^a (n = 50)
Number of FoRs				
No specific FoR	20%	10%	20%	28%
1 FoR	44%	45%	43%	54%
Multiple FoRs	36%	45%	37%	18%
FoR categories				
References to other people	44%	35%	33%	20%
References to an earlier time in life	46%	41%	59%	36%
References to an important event in the past	28%	31%	24%	36%
References to a hypothetical situation	4%	33%	0%	0%

^a Percentages are based on the column total n of responses for each domain

frequency of multiple FoRs ranged from 30% to 41% of the responses. References to *Other People* ranged from 30% to 38% of the responses, references to an *Earlier Time in Life* from 40% to 53%, and references to an *Important Event in the Past* from 24% to 34% across the age groups. The lowest frequency was observed for references to a *Hypothetical Situation* (8% to 11% of the responses across age groups).

Frequency of FoRs by Age Group and by Domain

Finally, we explored the frequency of FoRs for each domain for the young, middle, and older age groups. Table 3 presents the types of FoRs by domain and age group. Given the relatively small samples when the data are parsed this way, we only highlight very distinctive differences among the age groups.

Health Older respondents tended to report FoRs more frequently than younger (OR = 8.4) and middle-age respondents (OR = 3.4). The older respondents also tended to report references to *Other People* more frequently than the youngest respondents (OR = 4.2).

Well-Being Compared to the older age group, the young (OR = 4.5) and middle age groups (OR = 3.7) had a higher number of responses where FoRs were reported. The frequency of no specific FoR responses, however, was quite low in all age groups (see Table 3). References to an *Important Event in the Past* were made more frequently in the young (OR = 3.9) and the older age group (OR = 5.3) compared to the middle age group. All three age groups reported references to a *Hypothetical Situation* at comparable rates.

Fatigue The older age group had the highest number of responses with references to an *Important Event in the Past* compared to the middle age group (OR = 4.1). References to a *Hypothetical Situation* were not reported by any age group.

Table 3 Frequency of different frames of reference (FoR) by domain and age group

		No Specific FoR ^a	FoR Other people ^a	FoR Earlier time ^a	FoR Past event ^a	FoR Hypothetical ^a	Multiple FoRs ^a
	Age group						
Health	18–45 years (n = 17)	35%	29%	35%	29%	0%	24%
	46–59 years (n = 17)	18%	41%	47%	29%	12%	35%
	60+ years (n = 16)	6%	63%	56%	25%	0%	50%
Well-being	18–45 years (n = 19)	5%	26%	37%	37%	37%	42%
	46–59 years (n = 16)	6%	44%	56%	13%	31%	44%
	60+ years (n = 16)	19%	38%	31%	44%	31%	50%
Fatigue	18–45 years (n = 17)	12%	41%	65%	24%	0%	41%
	46–59 years (n = 16)	25%	25%	63%	13%	0%	35%
	60+ years (n = 16)	25%	31%	50%	38%	0%	44%
Pain	18–45 years (n = 17)	29%	29%	24%	35%	0%	18%
	46–59 years (n = 17)	18%	12%	47%	41%	0%	18%
	60+ years (n = 16)	38%	19%	38%	31%	0%	19%

^a Percentages are based on the row total n of responses for each age group

Pain There were no major differences between the three age groups for the frequencies of FoRs in the pain domain. As with fatigue, references to a *Hypothetical Situation* were not reported by any age group.

Discussion

The goal of this qualitative study was to examine which FoRs were naturally used by people from the general US population when providing self-reports of their health, well-being, and two common medical symptoms, fatigue and pain. We found that the qualitative interviews were successful in that participants generated various FoRs that could be reliably classified. Notably, we identified a FoR that has not been mentioned in prior research: references to a *Hypothetical Situation*, and it was especially evident in the well-being domain. This FoR shares similarities with the previously-identified construct of “possible selves” or “ideal selves”;

these selves refer to how people think about their potential and about their future, including their ideas of what they might become, and what they do or do not want to become (Frazier et al. 2000; Hooker and Kaus 1994; Markus and Nurius 1986; Higgins 1987). The FoR that we identified in the present study expands upon this construct in that it also included references to situations or circumstances in the past that participants wished would have been different.

The present study extends prior research that was limited in focus to the use of FoRs for self-rated health and health-related quality of life (Kaplan and Baron-Epel 2003; Fayers et al. 2007). A major study finding is that FoRs are reported in self-report ratings across *several domains*, not just for self-rated health. That is, self-report ratings are likely to evoke comparison standards, even if no such comparisons are intended or made explicit in the instructions or item wording – a phenomenon that generalizes across a variety of self-report content domains commonly used in survey research. Notably, this is the case even for ratings of symptoms like pain and fatigue, which concern more immediate sensory experiences, and not just for more “evaluative” judgments, such as health and well-being.

A second major study finding was that there is much variation among individuals both in terms of how many FoRs (if any) are reported and which types of FoRs are applied in self-report ratings. Depending upon the domain, 10% to 28% of responses did not reference a specific FoR, whereas multiple FoRs were reported in 18% to 45% of responses. The different types of FoRs that were reported involved social comparisons (“other people”), temporal comparisons (“earlier time in life”, “important past life events”), as well as comparisons with ideal states or other imagined states (“hypothetical situation”). Acknowledging this wide variation in the number and types of FoRs is important in that this variation could reduce the standardization of self-report ratings and the comparability of ratings across survey respondents.

Although qualitative research is not usually employed to examine subgroup differences because of small sample sizes, we examined the reporting of FoRs by study domain to generate hypotheses for future research. FoRs were generally reported in all four domains, yet there were some differences. Specifically, the use of at least one FoR was more prevalent for well-being (90%) compared to pain (72%) ratings. Furthermore, participants reported using references to *Other People* at higher rates in the health compared to the pain domain. The health domain had 44% of responses referencing *Other People*, compared to 20% of responses in the pain domain. It is also interesting that the likelihood of references to *Other People* in pain ratings is different than in fatigue (33% reported that FoR), as both are usually transient symptoms. One speculation is that pain is viewed as an experience with a clear anchor of “no pain” and levels of pain are reported in absolute sensory perception levels without regard to considerations of other people. In contrast, fatigue may not have a clear “no fatigue” level and may be more likely to be judged based upon what the respondent thinks is normative within their context. Somewhat consistent with this observation about pain ratings is that 38% of our oldest age group stated that they did not use a specific FoR for that rating, whereas they reported FoRs more frequently for the other domains. Finally, it is noteworthy that well-being was essentially the only domain that evoked references to a *Hypothetical Situation* (33% of responses in this domain). This FoR focused on considerations of what could have been or what might be in one’s life in the future. Well-being and health represent broader and multifaceted constructs (Kaplan and

Baron-Epel 2003) compared to pain and fatigue. Lifestyle, events, and opportunities are examples of how people view the determinants of the well-being and health that they experience. As such, comparisons to hypothetical situations (e.g., if I were retired, if I drove a bigger car, etc.) or behaviors (e.g., if I exercised more often) may be more salient in self-report ratings pertaining to those constructs.

Finally, we explored whether FoRs varied by age group and domain. Few pronounced differences were evident. There were two notable observations for the health domain: The older age group tended to report FoRs more often than the other two age groups. For example, the older age group made greater use of comparisons to *Other People* relative to the young group. These results are in line with prior research on reference group theory, which holds that subjective health assessment is dependent on the comparison group, or frame of reference, utilized (see Heckhausen and Brim 1997). According to this perspective, comparisons with other people increase with age. This is thought to allow individuals to adjust their standards in the process of healthy aging by comparing themselves, most often, to people of a similar age, and thereby continue to view their health in a positive light (Levkoff et al. 1987; Cockerham et al. 1983; Baron-Epel and Kaplan 2001; Fienberg et al. 1985; Kaplan and Baron-Epel 2003). Additionally, people in the middle age group tended to use references to an *Important Event* to a lesser extent compared to both the young and older age groups (this was the case for the well-being and fatigue domains). These results are in line with prior research on life events across adulthood. For example, prior studies have documented that significant life events are more salient at younger ages (e.g., transitioning to parenthood or career entry) and older ages (e.g., retirement and reductions in social networks) of the adult life span (Hatch and Dohrenwend 2007; Wrzus et al. 2013; Carstensen 1992).

The present study has several limitations. First, the non-probability sampling approach creates opportunity for potential participant selection bias. As is often the case with internet survey panels (Hays et al. 2015), our sample was highly educated, with 78% of participants having at least some college education. While we used recruitment quotas to ensure heterogeneity of other key participant characteristics within the interviewed sample, it is unclear the degree to which our results about the frequency of the use of specific FoRs may generalize to other samples or the broader US population as a whole. Future research is planned to further estimate rates of FoRs within larger samples.

Second, the use of more than one FoR for a single question was observed among a portion of participants (range of multiple FoRs by study domain = 18% to 45% of responses). It is possible that participants considered one of those FoRs more or less important. We did not ask participants to think about which of the multiple FoRs was the most influential for their self-report rating. Future research would benefit from trying to ascertain whether certain FoRs are considered more important to participants compared to others. Third, given the qualitative nature of this study and the sample size, we did not examine whether differences in participants' reporting of FoRs influence levels of ratings of health, well-being, fatigue, and pain. We intend to answer this question empirically in a future large-scale quantitative study drawing on the FoRs identified in the present study.

Conclusions and Future Directions

Our findings suggest that participants commonly engage in comparisons using FoRs when providing self-reports of health, well-being, fatigue, and pain even when a FoR is not specified within the text of a survey question. Because these constructs are widely considered through self-report methods in population and health survey research, the potential effects of FoRs may be far-ranging. Notably, the four items utilized in this study are similar to the Patient Reported Outcomes Measurement Information System (PROMIS®) (Cella et al. 2010; Cella et al. 2007; Hays et al. 2009) global health items, and use of PROMIS® by researchers and clinicians is increasing. This highlights the importance of seeking a better understanding of how FoRs may influence bivariate relationships between subgroups of individuals (e.g., different age groups) and health-related outcomes. If people differ in the FoRs they naturally use when responding to survey questions, their ratings may be influenced by the particular FoR they invoke, which may subsequently threaten the comparability of responses across groups. An important question for future research is to delineate systematic differences in the use of FoRs and to examine whether controlling for such differences in questionnaire instructions or item text facilitates unbiased comparisons in survey research.

Acknowledgements This work was supported by a grant from the National Institute on Aging (R01 AG042407, PI: Arthur A. Stone, Ph.D.). The authors would like to thank Mona Martin, Carla Ascoytia, Adam Bailey, Julia Correll, and Beatriz Medina for their valuable contributions to the study data collection and analysis efforts.

Compliance with Ethical Standards

Conflict of Interest A.A.S. is a Senior Scientist with the Gallup Organization. K.P.McC., L.M.S., and S.C.K. are employees of Health Research Associates (HRA), which was contracted as a research partner for the qualitative interviews for this project.

References

- 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*, 1373–1381.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101.
- Carstensen, L. L. (1992). Social and emotional patterns in adulthood: support for socioemotional selectivity theory. *Psychology and Aging*, *7*, 331–338.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al. (2007). The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. *Medical Care*, *45*, S3–S11.
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., et al. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, *63*, 1179–1194.
- Cleeland, C. (1994). Pain assessment: global use of the brief pain inventory. *Annals of Academic Medicine Singapore*, *23*, 129–138.
- Cockerham, W. C., Sharp, K., & Wilcox, J. A. (1983). Aging and perceived health status. *Journal of Gerontology*, *38*, 349–355.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

- 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*, 956–967.
- DeSalvo, K. B., Fan, V. S., McDonnell, M. B., & Fihn, S. D. (2005). Predicting mortality and healthcare utilization with a single question. *Health Services Research*, *40*, 1234–1246.
- 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. *J Gen Intern Med*, *21*, 267–275.
- 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.
- Effiace, F., Baccarani, M., Breccia, M., Cottone, F., Alimena, G., Deliliers, G. L., et al. (2013). Chronic fatigue is the most important factor limiting health-related quality of life of chronic myeloid leukemia patients treated with imatinib. *Leukemia*, *27*, 1511–1519.
- Fayers, P. M., Langston, A. L., Robertson, C., & PRISM trial group. (2007). Implicit self-comparisons against others could bias quality of life assessments. *J Clin Epidemiol*, *60*, 1034–1039.
- 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*, 836–844.
- 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*, 547–564.
- Franks, P., Gold, M. R., & Fiscella, K. (2003). Sociodemographics, self-rated health, and mortality in the US. *Social Science & Medicine*, *56*, 2505–2514.
- Frazier, L. D., Hooker, K., Johnson, P. M., & Kaus, C. R. (2000). Continuity and change in possible selves in later life: A 5-year longitudinal study. *Basic and Applied Social Psychology*, *22*, 237–243.
- Friese, S. (2015). *ATLAS.Ti 7: User guide and reference*. Berlin: ATLAS.ti Scientific Software Development GmbH.
- Hatch, S. L., & Dohrenwend, B. P. (2007). Distribution of traumatic and other stressful life events by race/ethnicity, gender, SES and age: a review of the research. *American Journal of Community Psychology*, *40*, 313–332.
- Hays, R. D., Bjorner, J. B., Revicki, D. A., Spritzer, K. L., & Cella, D. (2009). Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of Life Research*, *18*, 873–880.
- Hays, R. D., Liu, H. H., & Kapteyn, A. (2015). Use of internet panels to conduct surveys. *Behavior Research Methods*, *47*, 685–690.
- Heckhausen, J., & Brim, O. G. (1997). Perceived problems for self and others: self-protection by social downgrading throughout adulthood. *Psychology and Aging*, *12*, 610–619.
- Higgins, E. T. (1987). Self-discrepancy - a theory relating self and affect. *Psychological Review*, *94*, 319–340.
- Hooker, K., & Kaus, C. R. (1994). Health-related possible selves in young and middle adulthood. *Psychology and Aging*, *9*, 126–133.
- Jordan, K., Proskorovsky, I., Lewis, P., Ishak, J., Payne, K., Lordan, N., et al. (2014). Effect of general symptom level, specific adverse events, treatment patterns, and patient characteristics on health-related quality of life in patients with multiple myeloma: results of a European, multicenter cohort study. *Supportive Care in Cancer*, *22*, 417–426.
- Jylha, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, *69*, 307–316.
- Kaplan, G., & Baron-Epel, O. (2003). What lies behind the subjective evaluation of health status? *Social Science & Medicine*, *56*, 1669–1676.
- Levkoff, S. E., Cleary, P. D., & Wetle, T. (1987). Differences in the appraisal of health between aged and middle-aged adults. *Journal of Gerontology*, *42*, 114–120.
- Markus, H., & Nurius, P. (1986). Possible Selves. *American Psychologist*, *41*, 954–969.
- Mendoza, T. R., Wang, X. S., Cleeland, C. S., Morrissey, M., Johnson, B. A., Wendt, J. K., et al. (1999). The rapid assessment of fatigue severity in cancer patients: use of the brief fatigue Inventory. *Cancer*, *85*, 1186–1196.
- Ostbye, T., Krause, K. M., Norton, M. C., Tschanz, J., Sanders, L., Hayden, K., et al. (2006). Ten dimensions of health and their relationships with overall self-reported health and survival in a predominately religiously active elderly population: the cache county memory study. *Journal of the American Geriatrics Society*, *54*, 199–209.
- Ricci, J., Chee, E., & Lorandeanu, A. (2006). Fatigue in the US workforce: prevalence and cost of lost productive work time. *Value in Health*, *9*, A166–A166.
- 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*, 1–10.

- Roberts, G. (1999). Age effects and health appraisal: a meta-analysis. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *54*, S24–S30.
- 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*, 739–761.
- Sargent-Cox, K. A., Anstey, K. J., & Luszcz, M. A. (2010). 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.
- Schwarz, N. (1999). Self-reports: how the questions shape the answers. *American Psychologist*, *54*, 93–105.
- 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*, 662–667.
- Tourangeau, R. (1991). Context effects on responses to attitude questions: Attitudes as memory structures. In N. Schwarz & S. Sudman (Eds.), *Contextual effects in social and psychological research* (pp. 35–48). New York: Springer-Verlag.
- 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*, 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*, 251–257.
- 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*, 1237–1244.
- 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*, 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*, 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.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. New York: Sage.
- Wizus, C., Hanel, M., Wagner, J., & Neyer, F. J. (2013). Social network changes and life events across the life span: a meta-analysis. *Psychological Bulletin*, *139*, 53–80.