Qual Life Res (2013) 22:2761–2768
DOI 10.1007/s11136-013-0418-5

Self-rated health: analysis of distances and transitions between response options

Thomas V. Perneger · Angèle Gayet-Ageron · Delphine S. Courvoisier · Thomas Agoritsas · Stéphane Cullati

Accepted: 15 April 2013 / Published online: 25 April 2013

Abstract
Purpose We explored health differences between population groups who describe their health as excellent, very good, good, fair, or poor.
Methods We used data from a population-based survey which included self-rated health (SRH) and three global measures of health: the SF36 general health score (computed from the 4 items other than SRH), the EQ-5D health utility, and a visual analogue health thermometer. We compared health characteristics of respondents across the five health ratings.
Results Survey respondents ($N = 1.844$, 49.2 % response) rated their health as excellent (12.2 %), very good (39.1 %), good (41.9 %), fair (6.0 %), or poor (0.9 %). The means of global health assessments were not equidistant across these five groups, for example, means of the health thermometer were 95.8 (SRH excellent), 88.8 (SRH very good), 76.6 (SRH good), 49.7 (SRH fair), and 33.5 (SRH poor, $p < 0.001$). Recoding the SRH to reflect these mean values substantially improved the variance explained by the SRH, for example, the linear r^2 increased from 0.50 to 0.56 for the health thermometer if the SRH was coded as poor = 1, fair = 2, good = 3.7, very good = 4.5, and excellent = 5. Furthermore, transitions between response options were not explained by the same health-related characteristics of the respondents.
Conclusions The adjectival SRH is not an evenly spaced interval scale. However, it can be turned into an interval variable if the ratings are recoded in proportion to the underlying construct of health. Possible improvements include the addition of a rating option between good and fair or the use of a numerical scale instead of the classic adjectival scale.

Keywords Self-rated health · Health status measurement · Population surveys · Response scale

T. V. Perneger (✉) · A. Gayet-Ageron · D. S. Courvoisier · T. Agoritsas · S. Cullati
Division of clinical epidemiology, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland
e-mail: Thomas.Perneger@hcuge.ch

A self-rated health (SRH) item is included in many health status questionnaires and population surveys. It is often phrased as "Would you say your health is… excellent/very good/good/fair/poor," but other response scales have been used as well [1, 2]. This item is intuitively appealing, because it enables respondents to communicate their own assessment of their health, as they understand it. Furthermore, SRH predicts mortality [3–11], use of health services [12], and health expenditures [13] in various populations. Thus, the usefulness of this item is firmly established.

Nevertheless, what exactly self-ratings of health mean is not entirely clear [1, 2, 14, 15]. The dimensions included under the umbrella of "health" may vary between individuals. Some respondents may consider absolute health ratings, while others may compare themselves to other people of the same age. Self-rated health may reflect both an enduring self-concept of health that is impervious to temporary health problems and a fluctuating assessment that reflects current health problems [14]. Finally, the meaning of the various health ratings, such as "excellent" or "good," may vary between respondents. Despite these limitations, few studies have examined correlates of various levels of self-rated health, in little detail [4, 5, 10, 16].

Another issue is the selection of appropriate numbers to represent the ordinal response scale. If the groups of respondents who select a given health rating were evenly

spaced in terms of the measured construct—that is, their health—the adjectival scale could be converted into consecutive numerals, such as 1-2-3-4-5, and the resulting scale would be an interval variable [17]. If the distances in terms of health between adjacent health ratings varied in magnitude, then consecutive numbering would result in an ordinal variable. In that case, assignment of numerical values in proportion to the underlying health construct would restore interval properties. This is advantageous, because interval variables allow the derivation of useful statistics, such as differences, means, or variances [17, 18]. While treating summative multi-item scales as interval variables is deemed acceptable, even though they are sums of ordinal items, such tolerance does not extend to individual items such as the self-rated health [18]. The main practical issue is that inferences based on mean values of non-interval variables may be incorrect, since the assigned values are not in proportion to the measured construct.

In this study, we explored two issues related to self-rated health. The first is the spacing of the response options. Are the differences between adjacent ratings of the same magnitude in terms of health or are they uneven? The second is the description of transitions between ratings: How similar or different are people in adjacent categories of self-rated health? We used data from a population-based survey to address these questions.

Methods

Study design

Data were obtained as part of a general population mail survey in western Switzerland, whose primary aim was to obtain reference values for the EuroQol EQ-5D health utility instrument in a French-speaking population [19, 20]. Non-institutionalized residents aged 18 and over were eligible, and the target sample size was 1,600. The study was approved by the research ethics committee at University Hospitals of Geneva. The selection of a random sample from population registries and data collection (initial mailing and up to 2 reminders to non-respondents) were done by an independent survey firm (Infometrics, Le Muids, Switzerland).

Variables

The survey questionnaire included the 5-item general health subscale of the SF36 questionnaire, composed of the SRH item rated between excellent and poor and of four statements regarding the respondent's health (health is excellent, is as healthy as anyone, expects health to get worse, and gets sick easier than others) answered on a 5-point scale (definitely true, mostly true, don't know, mostly false, and definitely false) [21, 22]. SRH was used as the main variable of interest, and the four other items were combined into a separate validation score, scaled between 0 and 100.

The questionnaire also included the EQ-5D health utility instrument [23], which consists of five dimension-specific items—mobility, self-care, usual activities, pain and discomfort, and anxiety and depression—and a "health thermometer," a vertical visual analogue scale (VAS) that measures self-perceived health. The five dimension-specific items are combined into a health utility score. For exploratory purposes, a randomly selected subset completed an expanded version of the EQ-5D, in which 5 additional dimensions were rated—sleep, memory and concentration, energy and fatigue, sight and hearing, and contacts with others [20]. All EQ-5D items, and the additional items as well, were answered on 3-point scales (typically: no problem, moderate problem, and severe problem) [20]. Finally, the respondents answered questions about current treatment for a chronic or acute health problem, doctor visit in the past 6 months, hospital stay in the past 6 months, sex, age, country of birth, and level of education.

Analysis

To examine the spacing between the SRH responses, we examined means of three continuous measures of health: (1) the EQ-5D health utility (computed according to the European algorithm) [24], (2) the thermometer VAS value, and (3) a general health score based on the mean of the four SF36 general health items (the single health status rating was excluded from the original score to avoid overlap and a spuriously inflated correlation). All three continuous health measures were scored between 0 (lowest level) and 100 (highest level). We also computed standardized scores, with a mean of 0 and variance of 1, to facilitate comparisons between the 3 variables.

From this analysis we derived an alternative coding scheme for SRH. Instead of the evenly spaced coding (poor = 1, fair = 2, good = 3, very good = 4, excellent = 5), we assigned the values of 1 and 5 to the ratings of "poor" and "excellent," but set the codes of the middle categories in proportion to the differences between the means of the standardized scores. Three coding schemes were derived, based on the three standardized scores.

To determine to what extent these alternative coding schemes improved the linear correlation between the SRH and the global measures of health, we obtained coefficients of determination (r^2), with the health status being coded either by evenly spaced codes or by the alternative codes. The r^2 represents the proportion of variance in one variable

that is explained by the other. In interpreting this coefficient, it is useful to note that r^2 cannot exceed the product of the reliabilities of the 2 variables if measurement errors are independent [18]; for example, if the reliability coefficients are both 0.8, then r^2 cannot exceed 0.64.

To describe respondents' health status according to self-rated health, we compared the 5 groups defined by the SRH on the following variables: presence of a problem (either moderate or severe) on each of the EQ-5D health dimensions, and on the 5 additional dimensions in the subsample who answered these items, current treatment for an acute or chronic health problem, doctor visit in the past 6 months, and hospitalization in the past 6 months. The comparisons between these proportions were done using a chi-square test for linear trend.

Finally, we obtained four multiple logistic regression models comparing adjacent categories of health (e.g., excellent vs. very good), using the same predictors. The discrimination between adjacent categories of SRH was assessed by *C* statistics. The *C* statistic represents the probability that a person with the higher health rating will have a better health predicted by the regression model, when compared to a person with the lower health rating. A value of 0.5 indicates that the model does not discriminate (no better than a coin toss), and 1.0 represents perfect discrimination. The *C* statistic can also be interpreted as a measure of separation of 2 distributions (here, the distributions of predicted health levels in people with higher and lower health ratings). If the distributions were normal with a standard deviation (SD) of 1, a difference between the means of 0.5 would correspond to *C* statistic of 0.64, a difference of 1.0 to $C = 0.76$, a difference of 1.5 to $C = 0.86$, and a difference of 2.0 to $C = 0.92$ [25]. Analyses were done using SPSS version 17.

Table 1 Characteristics of 1,844 survey respondents who provided a rating of health, *N* (%)

Sex: women	1,035 (56.5)
Age-group	
20–39	568 (31.8)
40–59	652 (36.5)
60–79	474 (26.6)
≥80	90 (5.0)
Country of birth	
Switzerland	1,384 (75.2)
Europe	351 (19.1)
Outside Europe	105 (5.7)
Education: basic or vocational training	1,075 (58.8)
Medical visit in the past 6 months	1,225 (70.2)
Hospitalization in the past 6 months	147 (8.2)
Currently treated for chronic health problem	635 (35.0)
Currently treated for acute health problem	209 (11.8)
Current health status	
Excellent	225 (12.2)
Very good	721 (39.1)
Good	772 (41.9)
Fair	110 (6.0)
Poor	16 (0.9)
Problem with mobility (EQ-5D)	147 (8.0)
Problem with autonomy (EQ-5D)	42 (2.3)
Problem with daily activities (EQ-5D)	160 (8.7)
Problem with pain or discomfort (EQ-5D)	839 (45.9)
Problem with anxiety or depression (EQ-5D)	577 (31.6)
Problem with sleep*	120 (36.7)
Problem with memory or concentration*	115 (35.2)
Problem with fatigue/energy*	174 (53.4)
Problem with vision/hearing*	93 (28.7)
Problem in contacts with others*	23 (7.1)

* Among random subsample of 328 respondents

Results

Study sample

In total, 1,952 persons returned the questionnaire (52.1 % of 3,747 eligible persons contacted). Among respondents 1,844 (49.2 % of eligible persons) provided the requisite health rating and were included in the analysis. Women were in the majority, as were Swiss-born respondents, and those with only basic education (Table 1). The mean age was 52.3 years (SD 16.3). Most had seen a doctor in the past 6 months, and 8.2 % had been hospitalized during that time. About a third were treated for a chronic health problem, fewer for an acute health problem.

About half described their health as excellent or very good (Table 1). Of health domains explored by the EQ-5D, problems with mobility, autonomy, and daily activities were rare, but substantial minorities of respondents reported problems with pain or discomfort, and anxiety or depression. Among the subsample who were asked about additional problems ($N = 328$), problems with sleep, cognition, fatigue, or sensory functions were common as well.

The means and SDs were 69.6 (18.7) for the 4-item general health SF36 score, 81.9 (15.4) for the health thermometer, and 83.1 (14.7) for the EQ-5D health utility (Table 2).

Distances between health ratings

The means of all three continuous health assessments respected the order of the categories of the SRH (Table 2).

Table 2 Global evaluations of health—a 4-item measure of general health from the SF36, the EQ-5D health thermometer, and the EQ-5D health utility—across ratings of health on scale between "excellent" and "poor," and corresponding optimal 1–5 scores, in a general population sample

	Total	Self-rated health				
		Excellent	Very good	Good	Fair	Poor
General health SF36 (based on 4 items)						
Natural units, mean (SD)	69.6 (18.7)	87.5 (12.7)	77.7 (12.3)	62.5 (14.4)	37.0 (13.6)	18.4 (14.0)
Standardized units, mean (SD)	0 (1)	0.96 (0.68)	0.43 (0.66)	−0.38 (0.77)	−1.74 (0.73)	−2.73 (0.75)
Scaling from 1 to 5*		5	4.4	3.5	2.1	1
Health thermometer from EQ-5D						
Natural units, mean (SD)	81.9 (15.4)	95.8 (5.2)	88.8 (8.2)	76.7 (11.8)	49.7 (14.8)	33.5 (19.3)
Standardized units, mean (SD)	0 (1)	0.90 (0.34)	0.45 (0.53)	−0.33 (0.77)	−2.09 (0.96)	−3.13 (1.25)
Scaling from 1 to 5*		5	4.5	3.7	2.0	1
Health utility from EQ-5D						
Natural units, mean (SD)	83.1 (14.7)	94.4 (7.4)	88.8 (10.6)	78.8 (11.7)	59.9 (17.3)	35.3 (18.8)
Standardized units, mean (SD)	0 (1)	0.77 (0.50)	0.39 (0.72)	−0.29 (0.80)	−1.57 (1.17)	−3.25 (1.28)
Scaling from 1 to 5*		5	4.6	3.9	2.7	1

* Values of 5 and 1 set a priori, intermediate values proportional to mean measurements

The general health SF36 score ranged from 87.5 (excellent SWRH) to 18.4 (poor SRH), the health thermometer from 95.8 to 33.5, and the health utility from 94.4 to 35.3. The variance in health utility and thermometer scores was lower at the upper end of SRH than at the lower end; for example, the standard deviation of the health utility rating was 7.4 among those in excellent health but 18.8 among those in poor health. This tendency was less noticeable for the SF36 general health score.

Respondents in excellent health were on average not quite one SD unit above the general mean, and those in poor health were about three SD units below the general mean (Table 2). The rating of "very good" was about a half SD unit or less below the rating of "excellent" for all three scores. The widest gaps were between "good" and "fair" for the general health score (1.36 SD units) and the health thermometer (1.76 SD units), and between "fair" and "poor" for the health utility (1.68 SD units). Recoding the mean values between 1 (poor) and 5 (excellent) showed the same deviation from linearity using a different metric.

Variance explained

When coded evenly as 1-2-3-4-5, the self-rated health item explained half of the variance in the health thermometer score (r^2 0.50) and slightly less for the general health SF36 score and the health utility (Table 3). Using the alternative coding schemes, the variance explained increased for all three global measures of health. The highest value of the coefficient of determination was 0.56, obtained for the health thermometer.

Comparison of health ratings

Virtually, all health-related characteristics had a monotonic association with the SRH (Table 4). However, these associations were typically not linear, and different items discriminated between different health ratings. For instance, problems with mobility, autonomy, and daily activities were similarly low for respondents in excellent and very good health, but distinguished much better between respondents in good, fair, and poor health. Problems with pain or discomfort or fatigue and energy displayed a more progressive gradient, but with evidence of a floor effect, as respondents in fair and poor health were almost unanimous in reporting such problems. Similar floor effects were seen for doctor visits and for the proportion treated for a chronic health problem. On the other hand, the gradient was progressive throughout all transitions for problems with anxiety or depression.

The differences between adjacent health ratings were analyzed using multiple logistic regression models (Table 5). The transition between "excellent" and "very good" was most strongly associated with problems with pain or discomfort, but also with anxiety or depression and with current treatment for chronic and acute health problems. The discrimination of the model was rather weak, with a *C* statistic of 0.66 (95 % confidence interval 0.62–0.69). The transition between "very good" and "good" was also influenced by the prevalence of pain or discomfort, but also by the other predictors. The discrimination was moderate, with a *C* statistic of 0.76 (95 % confidence interval 0.73–0.78). The odds ratios were still

Table 3 Variance explained (r^2) in global measures of health—a 4-item measure of general health from the SF36, the EQ-5D health thermometer, and the EQ-5D health utility—by the single health rating, either coded evenly as 1-2-3-4-5 or coded proportionally to the mean values of the global measures of health (as in Table 2)

Global measures of health:	Even coding	Proportional coding
General health SF36 (based on 4 items)	0.46	0.49
Health thermometer from EQ-5D	0.50	0.56
Health utility from EQ-5D	0.36	0.40

Table 4 Percentages of health problems and use of health services according to self-rated health

Respondent characteristic	Self-rated health					*P* value for linear trend
	Excellent	Very good	Good	Fair	Poor	
Problem with mobility (EQ-5D)	0.4	1.7	8.5	50.0	87.5	<0.001
Problem with autonomy (EQ-5D)	0.0	0.1	1.7	18.5	50.0	<0.001
Problem with daily activities (EQ-5D)	0.4	0.4	9.3	65.1	87.5	<0.001
Problem with pain or discomfort (EQ-5D)	7.6	28.8	64.8	94.5	100	<0.001
Problem with anxiety or depression (EQ-5D)	9.4	21.2	41.8	64.8	93.8	<0.001
Problem with sleep*	5.7	24.8	45.4	81.5	66.7	<0.001
Problem with memory or concentration*	5.7	18.2	50.4	63.0	100	<0.001
Problem with fatigue/energy*	5.7	36.7	70.0	96.4	100	<0.001
Problem with vision/hearing*	5.9	16.5	36.4	65.4	100	<0.001
Problem in contacts with others*	0.0	4.1	7.1	25.7	50.0	<0.001
Doctor visit in the past 6 months	47.9	60.7	80.7	98.1	100	<0.001
Hospitalized in the past 6 months	2.7	5.1	10.2	24.0	28.6	<0.001
Treated for chronic health problem	7.2	19.0	48.7	92.5	93.8	<0.001
Treated for acute health problem	1.4	5.8	14.6	47.0	68.8	<0.001

* Only available for a subset of the respondents

Table 5 Transitions between ratings: logistic regression models predicting the lower of two adjacent ratings

	"Very good" versus "excellent"		"Good" versus "very good"		"Fair" versus "good"		"Poor" versus "fair"	
	Odds ratio	*P* value	Odds ratio	*P* value	Odds ratio	*P* value	Odds ratio	*P* value
Number of problems with mobility, autonomy, or daily activities (0–3)	1.5	0.68	2.8	<0.001	3.5	<0.001	2.3	0.012
Problem with pain or discomfort	4.0	<0.001	2.9	<0.001	3.1	0.023	Infinite	NA
Problem with anxiety or depression	2.1	0.004	2.3	<0.001	2.1	0.009	5.7	0.11
Treated for chronic health problem	2.0	0.019	2.5	<0.001	6.3	<0.001	1.0	0.98
Treated for acute health problem	2.8	0.10	1.7	0.013	2.9	<0.001	1.9	0.29

higher for the transition between "good" and "fair," and the *C* statistic reached 0.90 (95 % confidence interval 0.87–0.93). Last, the transition between "fair" and "poor" was significantly associated only with problems with mobility, autonomy, and daily activities (but this analysis was hampered by limited sample size). The *C* statistic of the model was 0.80 (95 % confidence interval 0.69–0.91).

Discussion

This population-based study confirms that the excellent-to-poor rating of health is not on an evenly spaced interval scale. While the results varied somewhat according to the reference variable used (health thermometer, health utility, or general health scale), the distance was smallest between

the ratings "excellent" and "very good," and about three times as large between "good" and "fair." Furthermore, the transitions between adjacent ratings were not explained by the same health characteristics.

Uneven spacing

The uneven spacing of the self-rated health descriptors is compatible with previous studies that compared respondent characteristics across categories of SRH. In US adults, a functional limitation score was similar for respondents in excellent, very good, and good health, but substantially worse for those in fair and particularly poor health [5]. Among British civil servants, the prevalence of unhealthy behaviors and of several chronic diseases increased only modestly between those in very good and good health and much more steeply for those in average, poor, and very poor health [10]. The strongest evidence comes from the analysis of a large sample of older European adults, in which self-rated health was compared with a standardized health index based on 20 self-reported diagnoses or health attributes, scaled from 0 to 1 [26]. The transition between poor and fair health occurred at the health index of 0.62, between fair and good at 0.80, between good and very good at 0.93, and between very good and excellent at 0.98. The nonlinearity of the self-rated health item is also supported by prognostic studies. When self-rated health is used to predict mortality, the risk ratio is smallest for the comparison of "excellent" with "very good" or "good" health [3–5, 11]. Our findings expand these results to a wider range of health characteristics.

What coding should be adopted for the SRH item is a somewhat arbitrary decision. We recommend the coding scheme based on the values of the health thermometer (e.g., excellent = 5, very good = 4.5, good = 3.7, fair = 2, poor = 1), because this variable was more closely correlated with the SRH than the other scores, and in our opinion it is conceptually the closest to SRH. Our results suggest that the SRH and the health thermometer are practically equivalent in terms of the measured health construct. Their adjusted r square was 0.56, and the coefficient of determination cannot exceed the product of the reliabilities of the 2 variables. Since the reliability of even a very good single item will be 0.7 or 0.8 at the most [27, 28], the correlation between the recoded SRH and the health thermometer would be very close to 1 after correction for attenuation.

Transitions between ratings

While the difference between "excellent" and "very good" health was the smallest of the health transitions, these labels are far from equivalent. People who defined their health as "very good" instead of "excellent" were more than twice as likely to suffer from anxiety or depression, and more than three times as likely to report pain or discomfort, cognitive difficulties, or sensory limitations. They were also more likely to be treated for an acute or chronic health problem and to have used health services in the past 6 months. These findings suggest that the addition of "very good" between "excellent" and "good" yields a meaningful improvement in discrimination. In contrast, the gap between ratings of "good" and "fair" health was wide. An additional health descriptor in this interval may be useful, along the lines of "moderately good" or "average." This additional descriptor may also help capture true variance in the middle of the health spectrum; the middle option "good" was selected by more than 40 % of the respondents, some of whom may have felt less than "good," but found "fair" too severe.

A caveat is that the gap between "good" and "fair" may be larger in French-speaking populations than in other language groups, because the official SF36 translation [22] of "fair" as "médiocre" is understood by many native speakers as "pretty bad," despite an etymology that suggests "average." Similar linguistic issues were noted by other Romance-language authors [11]. It is possible that different coding schemes may be required in different languages to obtain equivalent interval scales. Besides, evidence exists that the meaning of self-rated health varies from one country to another [26, 29]. A further exploration of intercultural equivalence of self-rated health is advisable [30].

That the transitions between health ratings were associated with a different set of predictors is intuitively logical. This finding suggests that the definition of health may not be stable across the spectrum between excellent and poor; for example, the difference between poor and fair health is explained in good part by a loss of autonomy and mobility, whereas these problems are uniformly absent for those in very good and excellent health. At least two phenomena may contribute to this finding. Firstly, some health problems cause a greater deterioration of self-perceived health than others. Secondly, people may redefine what contributes to their health as their level of health changes. For instance, difficulty sleeping may lead people to reevaluate their self-perceived health at the high end of the health spectrum, but if their health deteriorated seriously, the same person might not consider poor sleep reason enough to rate health lower. The latter hypothesis, analogous to a response shift [31], cannot be examined in a cross-sectional study.

Strengths and limitations

The strengths of this study are the general population sampling frame and the wide array of health-related

variables examined in relation to the single-item health rating. The main message was consistent across analyses—the single-item health rating is not linear vis-à-vis an underlying global concept of health.

The main limitation of this study is that the data were purely quantitative. We did not question respondents about their reasons for choosing a given health rating. Secondly, this analysis used an existing dataset that lacked information about some potentially important variables, such as psychosocial constructs (e.g., social support, self-efficacy, or sense of coherence) or specific clinical diagnoses. Also, this was a cross-sectional study, which afforded no opportunity to assess self-rated health over time and through health transitions. As for most population surveys, the response rate was less than optimal, so that selection bias is a possibility.

A methodological objection may be that we did not use item response theory to examine the self-rated health item. This is primarily because we wanted to focus on the item per se and not on the performance of any multi-item scale that would include the self-rated health item. Nevertheless, item response theory would allow the exploration of additional issues, such as differential item functioning across subgroups; this has been done previously for other health scales [32]. Finally, the analysis of transitions between ratings confirms that the self-rated health item probes a complex and multifaceted domain, despite the simplicity of its wording. These diverse facets remain implicit, which may be seen as weakness of this instrument. In situations where specific reasons for health ratings are of interest, multi-item health scales, such as the SF-12 or the EQ-5D, may be preferable.

Perspectives and recommendations

The main implication of our findings is that the SRH item should be coded unevenly, in proportion to the underlying construct of health. The data presented in this paper suggest that the coding of poor health as 1, fair as 2, good as 3.7, very good as 4.5, and excellent as 5 would be a reasonable solution, at least in a French-speaking population. This will improve the interpretation of mean values of SRH in populations or patient groups and facilitate any statistical operations that assume an interval rather than ordinal variable. Recoding is a simple operation, so there is no downside to it.

In addition, the gap between "good" and "fair" health is larger than is desirable, so that the possibility of inserting an intermediate label should be considered. This would improve the relevance and validity of the item for respondents whose health falls in between these labels and who may currently have difficulty selecting an appropriate response.

An alternative may be to forego adjectives altogether and use a numerical scale or a visual analogue scale instead, as is commonly done in pain assessment. In our data, the recoded adjectival scale was closely correlated with the health thermometer as represented in the EQ-5D instrument, that is, a visual analogue scale. Such scales require health descriptors only at the extremities of the scale and may prove more consistent across socioeconomic groups, education levels, and possibly languages.

References

1. Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science and Medicine, 69*, 307–316.
2. Fayers, P. M., & Sprangers, M. A. G. (2002). Understanding self-rated health. *Lancet, 359*, 187–188.
3. DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2005). Mortality prediction with a single general sefl-rated health question. A meta-analysis. *Journal of General Internal Medicine, 20*, 267–275.
4. Kaplan, G. A., & Camacho, T. (1983). Perceived health and mortality: A 9-year follow-up of the human population laboratory cohort. *American Journal of Epidemiology, 117*, 292–304.
5. Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior, 38*, 21–37.
6. Fried, L. P., Kronmal, R. A., Newman, A. B., Bild, D. E., Mittelmark, M. B., Polak, J. F., et al. (1998). Risk factors for 5-year mortality in older adults: The cardiovascular health study. *JAMA, 279*, 585–592.
7. McGee, D. L., Liao, Y., Cao, G., & Cooper, R. S. (1999). Self-reported health status and mortality in a multiethnic US cohort. *American Journal of Epidemiology, 149*, 41–46.
8. Idler, E. L., Kasl, S. V., & Lemke, J. H. (1990). Self-evaluated health and mortality among the elderly in New Haven, Connecticut, and Iowa and Washington Counties, Iowa, 1982–1986. *American Journal of Epidemiology, 131*, 91–103.
9. Idler, E. L., Russell, L. B., & Davis, D. (2000). Survival, functional limitations, and self-rated health in the NHANES I epidemiologic follow-up study, 1992. *American Journal of Epidemiology, 152*, 874–883.
10. Singh-Manoux, A., Gueguen, A., Martikainen, P., Ferrie, J., Marmot, M., & Shipley, M. (2007). Self-rated health and mortality: Short- and long-term associations in the Whitehall II study. *Psychosomatic Medicine, 69*, 138–143.
11. Lima-Costa, F. M., Cesar, C. C., Chor, D., & Proietti, F. A. (2012). Self-rated health compared with objectively measured health status as a tool for mortality risk screening in older adults: Ten-year follow-up of the Bambui cohort study on aging. *American Journal of Epidemiology, 175*, 228–235.
12. DeSalvo, K. B., Fan, V. S., McDonnell, M. B., et al. (2005). Predicting mortality and healthcare utilization with a single question. *Health Services Research, 40*, 1233–1246.
13. DeSalvo, K. B., Jones, T. M., Peabody, J., et al. (2009). Health care expenditure prediction with a single item, self-rated health measure. *Medical Care, 47*, 440–447.
14. Bailis, D. S., Segall, A., & Chipperfield, J. G. (2003). Two views of self-rated general health status. *Social Science and Medicine, 56*, 203–217.

15. Knäuper, B., & Turner, P. A. (2003). Measuring health: Improving the validity of health-assessments. *Quality of Life Research, 12*(Suppl 1), 81–89.
16. Mäntyselkä, P. T., Turunen, J. H. O., Ahonen, R. S., & Kumpusalo, E. A. (2003). Chronic pain and poor self-rated health. *JAMA, 290*, 2435–2442.
17. Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680.
18. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill, Inc.
19. Perneger, T. V., Combescure, C., & Courvoisier, D. S. (2010). General population reference values for the French version of the EuroQol EQ-5D health utility instrument. *Value Health, 13*, 631–635.
20. Perneger, T. V., & Courvoisier, D. S. (2011). Exploration of dimensions of health be included in multi-attribute health utility assessment. *International Journal for Quality in Health Care, 23*, 52–59.
21. Ware, 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.
22. Leplège, A., Ecosse, E., Verdier, A., & Perneger, T. V. (1998). The French SF-36 Health Survey: Translation, cultural adaptation, and preliminary psychometric evaluation. *Journal of Clinical Epidemiology, 51*, 1013–1023.
23. Rabin, R., & De Charro, F. (2001). EQ-5D: A measure of health status from the EuroQoL Group. *Annals of Medicine, 33*, 337–343.
24. Greiner, W., Weijnen, T., Nieuwenhuizen, M., Oppe, S., Badia, X., Busschbach, J., et al. (2003). A single European currency for EQ-5D health states. Results from a six country study. *The European Journal of Health Economics, 4*, 222–231.
25. Austin, P. C., & Steyerberg, E. W. (2012). Interpreting the concordance statistic of a logistic regression model: Relation to the variance and odds ratio of a continuous explanatory variable. *BMC Medical Research Methodology, 12*, 82.
26. Jurges, H. (2007). True health versus response styles: Exploring cross-country differences in self-reported health. *Health Economics, 16*, 163–178.
27. Martikainen, P., Aromaa, A., Heliövaara, M., Klaukka, T., Knekt, P., Maatela, J., et al. (1999). Reliability of perceived health by sex and age. *Social Science and Medicine, 48*, 1117–1122.
28. Zajacova, A., & Dowd, J. B. (2011). Reliability of self-rated health in US adults. *American Journal of Epidemiology, 174*, 977–983.
29. König, H. H., Bernert, S., Angermeyer, M. C., Matschinger, H., Martinez, M., Vilagut, G., et al. (2009). Comparison of population health status in six European countries. Results of a representative survey using the EQ-5D questionnaire. *Medical Care, 47*, 255–261.
30. Salomon, J. A., Tandon, A., & Murray, C. J. L. (2004). World health survey pilot study collaborating group. Comparability of self-rated health: Cross sectional multi-country survey using anchoring vignettes. *BMJ, 328*, 258.
31. Barclay-Goddard, R., Epstein, J. D., & Mayo, N. E. (2009). Response shift: A brief overview and proposed research priorities. *Quality of Life Research, 18*, 335–348.
32. Raczek, A. E., Ware, J. E., Bjorner, J. B., Gandek, B., Haley, S. M., Aaronson, N. K., et al. (1998). Comparison of Rasch and summated rating scales constructed from SF-36 Physical Functioning items in seven countries: Results from the IQOLA project. *Journal of Clinical Epidemiology, 51*, 1203–1214.