Health Psychology
2008, Vol. 27, No. 6, 669–675

0278-6133/08/$12.00 DOI: 10.1037/0278-6133.27.6.669

Are They Really That Happy? Exploring Scale Recalibration in Estimates of Well-Being

Heather P. Lacey
Bryant University

Angela Fagerlin
University of Michigan and VA Health Services Research & Development Center for Practice Management and Outcomes Research

George Loewenstein
Carnegie Mellon University

Dylan M. Smith
University of Michigan and VA Health Services Research & Development Center for Practice Management and Outcomes Research

Jason Riis
New York University

Peter A. Ubel
University of Michigan and VA Health Services Research & Development Center for Practice Management and Outcomes Research

Objective: The authors addressed a lingering concern in research on hedonic adaptation to adverse circumstances. This research typically relies on self-report measures of well-being, which are subjective and depend on the standards that people use in making judgments. The authors employed a novel method to test for, and rule out, such scale recalibration in self-reports of well-being. ***Design:*** The authors asked patients with chronic illness (either lung disease or diabetes) and nonpatients to evaluate quality of life (QoL) for the patients'disease. In addition, the authors also asked them to rank and rate the aversiveness of a diverse set of adverse circumstances, allowing examination of both the numerical ratings and ordering among items. ***Main Outcome Measures:*** The authors compared patients' and nonpatients' ratings and rankings for the patients' disease and other conditions. ***Results and Conclusion:*** The authors found that patients not only assigned higher numerical QoL ratings to their own disease than did nonpatients but also ranked it higher among the broad set of conditions. These results suggest that scale recalibration cannot account for discrepant QoL ratings between patients and nonpatients. More generally, this study presents a new approach for measuring well-being that is not subject to the problem of scale recalibration.

Keywords: quality of life, subjective well-being, chronic disease, scale recalibration

What makes people happier? Money and success? Youth and beauty? Health and strength? Many people go to great lengths to gain or maintain these things, presumably on the basis of expectations of enhanced well-being and happiness. However, a good deal of research suggests that people overestimate the influence of a wide range of life circumstances on well-being.

Over the past few decades, evidence has accumulated that most life circumstances do not affect well-being as dramatically as people typically think. People have been shown to mispredict the emotional impact of finances (Brickman, Coates, & Janoff-Bulman, 1978; Kermer, Driver-Linn, Wilson, & Gilbert, 2006), aging (Lacey, Smith, & Ubel, 2006), health (Boyd, Sutherland,

---

Heather P. Lacey, Department of Applied Psychology, Bryant University. Angela Fagerlin and Dylan M. Smith, Center for Behavioral and Decision Sciences in Medicine and Division of General Internal Medicine, University of Michigan; VA Health Services Research & Development Center for Practice Management and Outcomes Research, Ann Arbor. George Loewenstein, Department of Social and Decision Sciences and Department of Psychology, Carnegie Mellon University. Jason Riis, Department of Marketing, Stern School of Business, New York University. Peter A. Ubel, Center for Behavioral and Decision Sciences in Medicine, Division of General Internal Medicine, and Department of Psychology, University of Michigan; VA Health Services Research & Development Center for Practice Management and Outcomes Research, Ann Arbor.

Jason Riis is now at Marketing Unit, Harvard Business School.

We thank the staff at the Center for Behavioral and Decision Sciences in Medicine for research and administrative assistance. This work was funded by Grants RO1 HD40789-01, RO1 CA87595-01A1, and RO1 HD38963-02. Heather P. Lacey was supported by a Health Services Research & Development postdoctoral fellowship from the Department of Veterans Affairs. Angela Fagerlin and Dylan M. Smith received support from Merit Review Entry Program early career development awards from the Department of Veterans Affairs.

Correspondence concerning this article should be addressed to Heather Lacey, Department of Psychology, Bryant University, 1150 Douglas Pike, Smithfield, RI 02917. E-mail: hlacey@bryant.edu

Heasman, Tritchler, & Cummings, 1990; Brickman et al., 1971; Riis et al., 2005; Sieff, Dawes, & Loewenstein, 1999; Smith, Sherriff, Damschoeder, Loewenstein, & Ubel, 2006; Ubel, Loewenstein, & Jepson, 2005), environment (Loewenstein & Frederick, 1997; Schkade & Kahneman, 1998), hiring and promotion decisions (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998), election outcomes (Gilbert et al., 1998), and even sporting event outcomes (Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000). One possible reason for these misconceptions about circumstances and well-being is that people may overlook or underestimate the human capacity for adaptation to negative events. For example, Gilbert and colleagues (1998) have argued that a "psychological immune system" accounts for adaptation, and that it functions outside of awareness, accounting for affective forecasting errors due to "immune neglect."

Early evidence that well-being adapts to circumstances came from a classic study by Brickman, Coates, and Janoff-Bulmann (1978). They found that lottery winners' happiness levels matched those of control participants within a year of winning, suggesting that happiness levels adapt to the increased wealth. They also found that accident victims who experienced paraplegia also adapted to a large extent (although happiness did not recover completely to the level experienced by control participants). This striking study suggested that intensely joyous or traumatic experiences sometimes have only a limited effect on long-term well-being. Since then, additional evidence for adaptation has emerged, along with research about its mechanisms. There is evidence that happiness is largely heritable (Lykken & Tellegen, 1996) and that well-being is related to stable personality traits (Diener & Lucas, 1999). Members of our own research team have explored the mechanisms of adaptation to physical disability and have found that adaptation to a disability is more gradual for individuals with few financial resources (Smith, Langa, Kabeto, & Ubel, 2005), and that the prospect of future health improvement may inhibit adaptation to a disability (Smith, Loewenstein, Jankovic, & Ubel, 2007).

Despite this evidence, recent questions have been raised about the ubiquity and importance of adaptation in well-being (Diener, Lucas, & Scollon, 2006; Lucas, 2007) on the basis of evidence that adaptation is not always complete (Lucas, 2005), does not occur to the same extent for all life circumstances (Lucas, Clark, Georgellis, & Diener, 2004), and does not occur to the same extent for all individuals (Lucas, Clark, Georgellis, & Diener, 2003). Although this evidence does not contradict the basic finding that adaptation can and often does occur, it does indicate that adaptation may have limits.

The controversy over adaptation is further complicated by a nagging doubt about the validity of well-being measures. There is no truly objective way to measure constructs like happiness, life satisfaction, or quality of life (QoL), all of which rely on self-reports, and thus on personal interpretations of the construct itself and the scale used to measure it. For example, 70 of 100 on a QoL scale may mean something different to a patient with a chronic disease than it does to a healthy person (Ubel, Loewenstein, & Jepson, 2003). Much of the evidence for adaptation is based on the discrepancy between patients' and nonpatients' evalutions of the patients' disease, and this evidence is called into question if patients' and nonpatients' ratings are not comparable.

There is both theoretical and empirical reason to believe that patients and nonpatients may use different standards when evaluating their health. Early research on context effects in attitude measurement highlighted the distinction between how a stimulus is perceived and how it is described. Ostrom and Upshaw (1968) argued that the ratings or response language participants use to describe a stimulus may depend on its comparison to other stimuli, even if the objective perception of the stimulus is unaffected. Biernat, Manis, and Nelson (1991) demonstrated such a distinction between objective perceptions and subjective reports in social judgments. When asked to compare a set of male or female targets on the basis of limited information, participants estimated that the women earned objectively fewer dollars than the men, but they nevertheless described the women as subjectively more financially successful than the men, arguably because gender stereotypes invoked lower standards for women. Similarly, Highhouse, Brooks-Laber, Lin, and Spitzmueller (2003) demonstrated that job-seekers rate a salary as more satisfactory when comparing it with a lower range of alternative salaries. In the health domain, Schwartz and Sprangers (1999; Sprangers & Schwartz, 1999) argued that similar shifts in internal standards may occur as a result of experience with health problems. Consistent with this argument, Dar, Ariely, and Frenk (1995) found that pain threshold increases after a severe injury. Schwarz and Scheuring (1992) found higher health satisfaction ratings when respondents used a rating standard that implied a lower than average frequency of symptoms. In self-ratings of health, Ubel, Jankovich, Smith, Langa, & Fagerlin (2005) demonstrated that older adults rated their own health higher on a rating scale when the endpoint was *perfect health for someone your age* than when the endpoint was *perfect health for a 20-year-old*. When the standard was unspecified (i.e., *perfect health*), participants apparently used *for someone your own age* as their standard. When evaluating well-being for a health state, a financial state, or some other life circumstance, participants may similarly recalibrate the rating scale, implicitly using *for someone in my situation* as their standard, resulting in higher or lower ratings than would be expected on a more broadly interpreted rating scale.

In the health literature, debate about the validity of subjective well-being measures has centered on the concept of *response shift*. In this literature, response shift has been defined as a change in the meaning of a subjective outcome (such as QoL or happiness) as a result of "(a) a change in the respondent's internal standards of measurement (scale recalibration, in psychometric terms); (b) a change in the respondent's values (i.e., the importance of component domains constituting the target construct); or (c) a redefinition of the target construct" (Sprangers & Schwartz, 1999, p. 1508).

This definition of response shift does not distinguish between changes in response that occur because of true adaptation with those that result from measurement error. For example, a person might adapt to her disability by reducing the value she places on physical activity and increasing the value she places on spiritual pursuits. Through this shift in values, she may truly increase her QoL and happiness. By contrast, another person may report an increase in QoL over time simply because she changes the way she interprets concepts like "70 of 100" on a QoL scale. Both of these people have demonstrated a response shift, according to the above definition, but for different reasons. Our interest in this study was to help separate the latter person, who unknowingly exaggerated her QoL due to scale recalibration, from the former, whose QoL score potentially demonstrates true emotional adaptation.

For the current study, we drew on QoL judgments in the health domain to address whether scale recalibration accounts for the impact bias that is often observed in well-being estimates. We elicited QoL estimates for a broad set of conditions from two groups of patients suffering from chronic disease (lung disease or diabetes) and a group of people with neither disease (hereafter *nonpatients*). The rated conditions included both lung disease and diabetes, and 23 other conditions, some health-related and some not. By asking people to evaluate multiple conditions, we were able to compare not only the numerical ratings provided by each group but also the relative ranking of each condition for each group of participants, comparisons that should enable us to identify scale recalibration if it occurs.

If patients perceive their own health condition the same way nonpatients do but use rating scales differently, we would expect patients and nonpatients to give the same rank position to that condition, despite giving different ratings. If patients are using the rating scales differently from nonpatients, we would also expect them to rate not only their own condition higher than nonpatients but other conditions as well.

At one extreme, if scale recalibration accounted for the entire discrepancy between patients' and nonpatients' ratings of health conditions, then both groups should give the same rank orders to a set of conditions, but the patients' ratings (of all conditions) should be higher than nonpatients' ratings. At the opposite extreme, if the entire difference between patients' and nonpatients' ratings results from patients truly perceiving their own disease differently than nonpatients do, then patients should give their own disease a higher rating and ranking than the nonpatients do, but the ratings and rankings of all other diseases should be similar between the groups. In other words, if scale recalibration plays no role in the rating discrepancy—if the discrepancy is not an artifact of scale recalibration—then both ratings and rankings for lung disease should be higher for patients than for nonpatients, and both ratings and rankings for diabetes should be higher for diabetes patients than for nonpatients. In addition, the groups should show no differences in scale usage, with similar range and variance of responses for other items.

Method

Participants

Study participants were drawn from a panel of Internet users who volunteered to receive invitations to complete questionnaires, administered by Survey Sampling International (SSI). Upon completion, participants were entered into a drawing administered by SSI for cash prizes totaling $10,000. E-mail invitations were sent to a random sample of panel members who had previously identified themselves to SSI as (a) suffering from chronic obstructive pulmonary disease, (b) suffering from Type I or Type II diabetes, or (c) suffering from neither of these health conditions. We used demographic information supplied by SSI about their panel members to estimate the makeup of the patient groups in terms of age, sex, and ethnicity, and we used these estimates to establish target response levels for demographic subsamples within the nonpatient group. The number of e-mail invitations was dynamically adjusted until all quotas were completed for each group.

To confirm the patient status of participants recruited through SSI, we asked each participant who logged onto the survey to indicate whether they had ever experienced lung disease or diabetes. Participants were designated as lung disease patients if they indicated experience with lung disease only, as diabetes patients if they indicated experience with diabetes only, and as nonpatients if they indicated experience with neither.[1]

The final pool of participants included 265 lung disease patients, 366 diabetes patients, and 456 nonpatients. Patients in our sample were significantly older ($M = 54.4$ years, $SD = 12.0$) than nonpatients ($M = 49.3$ years, $SD = 12.6$), $t(1{,}069) = 6.75$, $p < .001$, and were significantly less likely to have attended at least some college (37.8%) than nonpatients (45.4%), $\chi^2(1, N = 1{,}072) = 6.31$, $p = .01$. These differences reflect population differences between these patient groups and nonpatients. Riis and colleagues (2005) have previously demonstrated that neither of these characteristics can account for discrepancies in patients' and nonpatients' well-being estimates. Our patients and nonpatients did not differ in sex, $\chi^2(1, N = 1{,}068) = 0.09$, $p = .77$ (70.5% female across groups) or ethnicity, $\chi^2(1, N = 1{,}055) = 0.65$, $p = .42$, (90.5% Caucasian across groups).

Questionnaire Design

All participants saw a list of 23 conditions spanning a wide range of severity (e.g., from having dandruff to being paralyzed from the neck down), including 11 health-related conditions and 12 conditions unrelated to health (see Table 1). Participants were then asked to estimate the QoL of each condition on a scale from 0 (*a quality of life that is no better than death*) to 100 (*an ideal quality of life*). The items had been generated through pilot testing with the goal of developing a full set that would be rated with a relatively even distribution across the 0 to 100 scale. An initial set of items was generated and tested with pilot participants. Items were removed if they clustered too closely with adjacent items, and new items were added to fill large gaps between adjacent items. This process was repeated four times, until an evenly distributed set of items emerged.

Next, participants read two brief scenarios, one describing a lung condition and the other describing diabetes, and were asked to rate it on the same 0 to 100 scale. The scenarios gave brief descriptions of some of the difficulties associated with each disease (see Appendix). Participants in each patient group rated the scenario corresponding to their own health condition first, and nonpatients were randomly assigned to rate either lung disease or diabetes first.

The full set of 25 rated conditions (the initial set of 23, plus lung disease and diabetes) was presented again with the participants' QoL ratings, arranged in rank order from highest to lowest, and participants were given an opportunity to alter their ratings (which could lead to a change in rankings).

Finally, all participants were asked to rate their own QoL using the same rating scale. Participants from each patient group were also asked to compare their own health condition with the scenario describing their disease. Patients indicated whether their own

---

[1] An additional 175 participants indicated experience with both lung disease and diabetes and were not included in analyses reported here.

Table 1
Mean Ratings and Rankings for Each Condition, Across Participant Groups

Condition	Rating *M*	Rating *SD*	Ranking *M*	Ranking *SD*
Eyeglasses required for driving	83.3	25.3	4.1	5.8
Dandruff, controlled with shampoo	80.9	27.1	4.5	6.0
Long, unpleasant commute	64.7	26.3	8.7	5.6
Severe allergies, weekly shots	60.6	24.6	9.7	5.5
Frequent nightmares	60.0	22.8	10.3	5.5
Neighbors frequently interrupt sleep	59.9	25.2	10.4	5.2
Caring for elderly parent	56.7	27.4	10.7	5.5
30 lbs overweight	58.5	26.1	10.7	5.3
Infertility	57.6	25.7	10.8	5.8
Periodic depression	53.4	26.1	12.1	5.5
Frequent insomnia	54.5	25.9	12.2	5.5
Working for a boss you hate	53.6	26.1	12.5	5.3
Annual flooding of home	51.8	24.9	13.1	5.1
Migraines, 3 times per week	46.2	26.9	14.3	5.4
Living in a high crime neighborhood	43.0	25.8	15.4	5.8
Bad marriage/relationship	42.8	26.2	15.6	5.3
Unemployment with few prospects	43.1	25.6	15.8	5.5
Child born with severe disability	40.5	25.2	16.5	5.0
Colostomy	37.6	26.0	17.2	5.3
Paraplegia	32.3	25.3	18.5	5.8
HIV diagnosis	27.2	25.6	19.7	5.6
Child dies painlessly in car accident	24.4	26.1	19.9	6.2
Quadraplegia	15.5	22.2	21.8	6.0

Note. Ratings were made on a 0 to 100 scale, with higher numbers representing better quality of life. Rankings range from 1 to 25, reflecting 25 items rated by each participant. Lower numbers represent better ranking.

condition was better than the scenario, about the same as the scenario, or worse than the scenario.

Results

As noted, rating scores ranged from 0 to 100, with higher numbers indicating more positive evaluations. QoL ratings for all scenarios were compared across groups using one-way analyses of variance with planned contrasts. Ranking scores range from 1 (*best*) to 25 (*worst*), with lower numbers representing more positive evaluations. Rankings were calculated for each participant by assigning a rank of 1 to the item with the highest rating, a rank of 2 to the next highest rating, and so on. The mean rank score was used to resolve ties when multiple items received the same rating. Rankings were compared across groups using the nonparametric Mann–Whitney *U* test. Self-reported QoL was compared across groups, and compared with scenario ratings, using independent samples *t* tests. An alpha level of .05 was used for all comparisons. Because the existing literature and our own pilot testing allowed us to make directional predictions for the patients' disease (i.e., diabetes patients ratings and rankings should be higher than nonpatients'), we used one-tailed significance tests to make these comparisons. We used two-tailed significance tests for all other comparisons.

Do Patients Rate Their Own Health Condition Differently from Nonpatients?

Figure 1 presents the ratings for lung disease and diabetes. As predicted, lung disease patients rated their own condition significantly higher than did nonpatients, $t(1{,}084) = 1.79$, $p = .04$. Diabetes patients also rated their own condition significantly higher than did nonpatients, $t(1{,}084) = 5.86$, $p < .001$.

Do Patients and Nonpatients Use the Rating Scale Differently?

If the discrepancies in ratings resulted from scale recalibration, then the higher ratings that patients gave to their own disease should be accompanied by higher ratings for all items. Patients would be expected to shift all of their responses to a higher level on the scale, and two consequences would follow: (a) The relative position of the patients' disease among other conditions (i.e., its rank) would not differ for patients and nonpatients, and (b) the mean rating across all items would be higher for patients than for nonpatients. Every condition's rating would be increased, but the patients' own disease would not shift any more than other conditions.

We found no evidence that scale recalibration occurred. First, as Figure 2 shows, lung disease patients ranked lung disease significantly higher than did nonpatients, $U = 52937$, $p = .003$, and diabetes patients ranked diabetes significantly higher than did nonpatients, $U = 54145$, $p < .001$. Thus, the ranking analysis revealed the same discrepancies between patients and nonpatients that we observed in the numerical ratings; patients' perceptions of lung disease or diabetes led them to assign a higher QoL to their own illness, relative to other conditions. These results are incon-

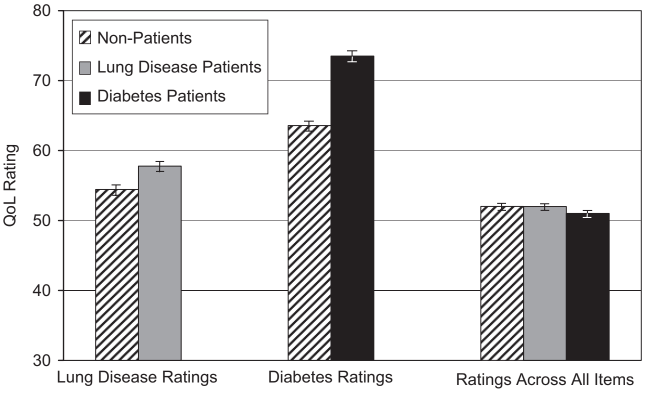

Figure 1. Quality of life (QoL) ratings for lung disease, diabetes, and all conditions, for nonpatients, lung disease patients, and diabetes patients. Error bars represent standard error for the difference in means for each comparison.

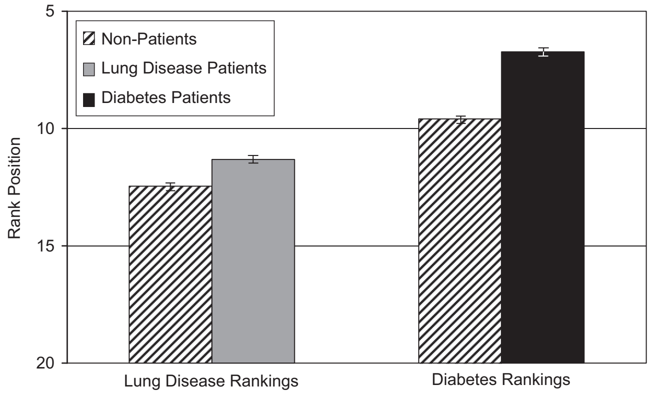

Figure 2. Rank positions for lung disease and diabetes for nonpatients, lung disease patients, and diabetes patients. Error bars represent standard error for the difference in means for each comparison.

sistent with a scale recalibration explanation for the QoL discrepancy, which predicts a similar ratings shift for all items and no reordering among them.

Second, comparisons across the full set of rated items suggest that the patient groups did not use the rating scale differently from the nonpatients. As shown in the right-hand group of bars in Figure 1, all three groups gave equivalent ratings, on average, across the full set of items, $F(2, 1084) = 0.48$, $p = .62$. We further compared the three groups on a number of descriptive statistics and found no differences that would suggest different scale usage (see Table 2). In addition to equivalent mean ratings across items, the groups did not differ in their variance in ratings across items, $F(2, 1084) = 0.96$, $p = .38$; their mean range across ratings, $F(2, 1084) = 0.17$, $p = .85$; their mean maximum ratings, $F(2, 1084) = 0.07$, $p = .93$; or their mean minimum ratings, $F(2, 1084) = 0.34$, $p = .71$.

Finally, we found that each patient group differed from nonpatients in ratings and rankings for their own health condition, but not for each others' condition. Lung disease patients neither rated ($M = 65.1$, $SD = 24.3$), $t(1{,}084) = 0.83$, $p = .41$, nor ranked ($M = 9.1$, $SD = 4.8$), $U = 57485$, $p = .28$, diabetes differently than did nonpatients. Similarly, diabetes patients neither rated ($M = 54.8$, $SD = 24.2$), $t(1{,}084) = 0.23$, $p = .82$, nor ranked ($M = 11.9$, $SD = 5.4$), $U = 78755$, $p = .17$, lung disease differently than did nonpatients. These results provide further evidence that experience with a particular health condition results in more positive evaluations of that health condition but does not influence evaluations of other conditions.

Although patients rated and ranked their own health condition more positively than did nonpatients, it is important to note that when asked about their own current QoL, nonpatients ($M = 82.7$, $SD = 18.1$) rated themselves higher than both lung disease patients ($M = 73.0$, $SD = 21.7$), $t(1{,}084) = 6.45$, $p < .001$, and diabetes patients ($M = 78.5$, $SD = 19.0$), $t(1{,}084) = 3.08$, $p = .002$. However, nonpatients' ratings for the lung disease scenario ($M = 54.4$, $SD = 23.9$) were significantly lower than the self-reported QoL for the lung disease group as a whole, $t(1{,}087) = 10.20$, $p < .001$, and for the 115 lung disease patients who indicated that their own health condition was similar to the lung scenario ($M = 72.3$, $SD = 21.0$), $t(569) = 7.32$, $p < .001$. Similarly, nonpatients' ratings for the diabetes scenario ($M = 63.6$, $SD = 24.0$) was significantly lower than the self-reported QoL for the diabetes group as a whole, $t(1{,}086) = 9.39$, $p < .001$, and for the 80 lung disease patients who indicated that their own health condition was similar to the diabetes scenario ($M = 75.9$, $SD = 20.5$), $t(534) = 4.33$, $p < .001$. In other words, the self-report QoL ratings indicate that QoL does suffer as a result of chronic disease, but not as much as nonpatients imagine.

In summary, we found no evidence that scale recalibration accounts for discrepant QoL ratings between patients and nonpatients. Not only did lung disease patients and diabetes patients rate their own health condition more favorably than did nonpatients, but they also ranked them higher among a broader set of life conditions. Patient versus nonpatient discrepancies were not found when patients were rating a health condition other than their own (i.e., when lung disease patients rated diabetes, or vice versa), nor were discrepancies found for the general set of rated items. There was no evidence to suggest that patients used the rating scale differently than did nonpatients, neither shifting their set of responses up or down the rating scale nor stretching or shrinking their responses across a different range on the rating scale.

Discussion

Although many studies support the claim that well-being adapts to circumstance, the strength of the evidence has been limited by the failure of previous studies to rule out scale recalibration. Self-reports of happiness, life satisfaction, and QoL are all potentially subject to individual interpretation and shifting standards, depending on the experiences of the rater, casting doubt on the conclusions that have been drawn about well-being and its relationship to life circumstances.

This methodological vulnerability has inspired efforts to create objective, systematic standards for well-being measures, such as Hsee and Tang's (2007) modulus-based measure in which participants are given some universal standard (e.g., a sunny day) and asked to evaluate their happiness relative to this standard, an approach that may be less prone to scale recalibration. We were similarly motivated to develop a new measure that would allow us to detect scale recalibration in QoL measures if it occur. We used ranking comparisons to explore the gap between patients' and nonpatients' QoL estimates for health conditions, and we found no evidence that this gap results from scale recalibration. Patients

Table 2
Scale Usage Indicators for Nonpatients, Lung Disease Patients, and Diabetes Patients

Indicator of scale usage	Nonpatients ($n = 456$)	Lung disease patients ($n = 265$)	Diabetes patients ($n = 366$)	p
Mean variance across ratings	661.1	677.2	693.6	.39
Mean range across ratings	84.4	85.1	85.7	.85
Mean maximum rating	93.3	93.2	93.5	.93
Mean minimum rating	8.9	8.1	8.8	.71

Note. Ratings were made on a 0 to 100 scale, with higher numbers representing more positive evaluations.

with chronic lung disease or diabetes evaluated their own disease more positively than did nonpatients, without evaluating other diseases and life conditions any differently and without using the rating scale differently.

The hybrid ranking–rating approach outlined here could easily be applied to other domains. For example, it is notorious that jury awards for pain and suffering tend to be capricious, with amounts awarded displaying both "horizontal" inequities in which people with similar injuries receive vastly different awards and "vertical" inequities in which people with injuries of obviously different severity receive similar awards (Abel, 2006). The problem of determining a money value for a victim's pain and suffering is no easier, and very possibly more difficult, than the problem of evaluating the QoL associated with a medical condition on a 0–100 scale, and a hybrid ranking–rating task of the type introduced here could help to simplify it. Elsewhere, two of the authors (Ubel & Loewenstien, in press) have argued that juries in torte cases, instead of deciding on monetary reward values, should evaluate noneconomic damages by ranking the severity of the plaintiff's damages relative to a standardized graded list of possible damages. Money values could then be determined by the position within the list.

Likewise, there is considerable evidence that "contingent valuation" measures of the value that people place on environmental amenities display a variety of undesirable properties, such as scope insensitivity (in which environmental amenities of vastly different scales are valued similarly). Again, a ranking task can provide a superior measure of environmental value or damage. Thus, for example, although people might state about the same willingness to pay to clean up a single polluted lake or to clean up all the lakes in a state, they will rank the pollution of a single lake as less severe than some third environmental problem, and rank the pollution of multiple lakes as worse than the third problem (Fischhoff et al., 1993). It would be easy to design a tool to evaluate the severity of environmental problems, much as we have done for health problems, by having people position their severity relative to a graded list of environmental problems ranging from the most trivial (e.g., debris in a city park) to the most momentous (e.g., the destruction of an ecosystem).

Further research is necessary to determine whether this approach translates successfully to these issues. If so, the benefits of the methodology may go beyond simply showing that the discrepancy between the patient and nonpatient evaluations of the QoL of patients is not due to scale recalibration.

References

Abel, R. (2006). General damages are incoherent, incalculable, incommensurable, and inegalitarian (but otherwise a great idea). *DePaul Law Review, 55,* 253–329.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of social judgment. *Journal of Personality and Social Psychology, 60,* 485–499.

Boyd, N. F. Sutherland, H. J., Heasman, K. Z., Tritchler, D. L., & Cummings, B. J. (1990). Whose utilities for decision analysis? *Medical Decision Making, 10,* 58–67.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology, 36,* 917–927.

Dar, R., Ariely, D., & Frenk, H. (1995). The effect of past-injury on pain threshold and tolerance. *Pain, 60,* 189–193.

Diener, E., & Lucas, R. E., (1999). Personality and subjective well-being. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 213–229). New York: Russell Sage Foundation.

Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *American Psychologist, 61,* 305–314.

Fischhoff, B., Quadrel, M. J., Kamlet, M., Loewenstein, G., Dawes, R., Fischbeck, P., et al. (1993). Embedding effects: Stimulus representation and response modes. *Journal of Risk and Uncertainty, 6,* 211–234.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. D. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75,* 617–638.

Highhouse, S., Brooks-Laber, M. E., Lin, L., & Spitzmueller, C. (2003). What makes a salary seem reasonable? Frequency context effects on starting-salary expectations. *Journal of Occupational and Organizational Psychology, 76,* 69–81.

Hsee, C. K., & Tang, J. (2007). Sun and water: On a modulus-based measurement of happiness. *Emotion, 7,* 213–218.

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological Science, 17,* 649–653.

Lacey, H. P., Smith, D. M., & Ubel, P. A. (2006). Hope I die before I get old: Mispredicting happiness across the adult lifespan. *Journal of Happiness Studies, 7,* 167–182.

Loewenstein, G., & Frederick, S. (1997). Predicting reactions to environmental change. In M. Bazerman, D. Messick, A. Tenbrunsel, & K. Wade-Benzoni (Eds.), *Environment, ethics, and behavior* (pp. 52–72). San Francisco: New Lexington Press.

Lucas, R. E. (2005). Time does not heal all wounds: A longitudinal study of reaction and adaptation to divorce. *Psychological Science, 16,* 945–950.

Lucas, R. E. (2007). Adaptation and the set-point model of subjective well-being: Does happiness change after major life events? *Current Directions in Psychological Science, 16,* 75–79.

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology, 84,* 527–539.

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science, 15,* 8–13.

Lykken, D., & Telligen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7,* 186–189.

Ostrom, T. M., & Upshaw, H. S. (1968). Psychological perspective and attitude change. In A. Greenwald, T. Brock, & T. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press.

Riis, J., Loewenstein, G., Baron, J., Jepson, C., Fagerlin, A., & Ubel, P. A. (2005). Ignorance of hedonic adaptation to hemo-dialysis: A study using ecological momentary assessment. *Journal of Experimental Psychology, General, 134,* 3–9.

Schkade, D., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9,* 340–346.

Schwartz, C. E., & Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science & Medicine, 48,* 1531–1548.

Schwarz, N., & Scheuring, B. (1992). Frequency reports of behaviors and symptoms: What respondents learn from the response alternatives of the questionnaire. *Zeitschrift für Klinishce Psychologie Forschung und Praxis, 21,* 197–208.

Sieff, E. M., Dawes, R. M., & Loewenstein, G. (1999). Anticipated versus actual reaction to HIV test results. *American Journal of Psychology, 112,* 297–311.

Smith, D. M., Langa, K. M., Kabeto, M. U., & Ubel, P. A. (2005). Health, wealth, and happiness: Financial resources buffer subjective well-being after the onset of a disability. *Psychological Science, 16,* 663–666.

Smith, D. M., Loewenstein, G., Jankovic, A., & Ubel, P. A., (2007). *The dark side of hope: It's harder to adapt when the adversity is temporary.* Unpublished manuscript, University of Michigan.

Smith, D. M., Sherriff, R. L., Damschroder, L., Loewenstein, G., & Ubel, P. A. (2006). Misremembering colostomies? Former patients give lower utility ratings than do current patients. *Health Psychology, 25,* 688–695.

Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science & Medicine, 48,* 1507–1515.

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, 10,* 1054–1057.

Ubel, P., & Loewenstein, G. (in press). Pain and suffering awards: It shouldn't be (just) about pain and suffering. *Journal of Legal Studies.*

Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research, 12,* 599–607.

Ubel, P. A., Loewenstein, G., & Jepson, C. (2005). Disability and sunshine: Can predictions be improved by drawing attention to focusing illusions or emotional adaptation? *Journal Experimental Psychology: Applied, 11,* 111–123.

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78,* 821–836.

Appendix

Health Conditions Scenarios

Lung Disease Scenario

Please consider the following condition: Having a lung condition that causes you to become short of breath after walking briskly for 2 blocks or walking up one flight of stairs.

Diabetes Scenario

Please consider the following condition: Having Diabetes that requires you to check your blood sugar level by sticking your finger with a small needle several times a day, and requires you to use a syringe to give yourself insulin shots in the abdomen three times a day. Imagine that you have no complications of your diabetes. That means your diabetes has not caused problems with your eyes, heart, kidneys, blood vessels, or nerves.