Running Head: Ignorance of Hedonic Adaptation

Ignorance of Hedonic Adaptation to Hemo-Dialysis: A Study Using Ecological Momentary Assessment

Jason Riis

University of Michigan

George Loewenstein

Carnegie Mellon University

Jonathan Baron

University of Pennsylvania

Christopher Jepson

University of Pennsylvania

Angela Fagerlin & Peter A. Ubel

University of Michigan and Veterans Affairs Medical Center, Ann Arbor, Michigan

Corresponding Author:
Jason Riis
Center for Health and Wellbeing
Woodrow Wilson School, Princeton University
Princeton, NJ 08544
Tel: 609-258-7085
Email: jriis@princeton.edu

UNDER 2ND REVIEW AT JEP:G

Word Count: 6,193; References: 28

Abstract

Healthy people generally underestimate the self-reported well-being of people with disabilities and serious illnesses. The cause of this discrepancy is in dispute, and the present study reveals two causes. First, healthy people fail to anticipate hedonic adaptation to poor health. Using an Ecological Momentary Assessment (EMA) measure of mood we fail to find evidence that hemo-dialysis patients are less happy than healthy non-patients, suggesting that they have largely, if not completely, adapted to their condition. In a forecasting task, healthy people failed to anticipate this adaptation. Second, while controls understated their own mood in both an estimation task and a recall task, patients were quite accurate in both tasks. Such a difference in the manner in which patients and healthy people make overall or summary reports of well-being may also account for some of the discrepancy.

Ignorance of Hedonic Adaptation to Hemo-Dialysis: A Study Using Ecological Momentary Assessment

In the same year that Brickman, Coates and Janoff-Bulman (1978) published the famously counterintuitive result that paraplegics are not that much less happy than lottery winners, Sackett and Torrance (1978) demonstrated that there are other serious health conditions which do not seem to be as badly experienced by the people suffering from them as healthy people would expect. Dialysis patients, for example, rated their quality of life as .56 (on a scale from 0 to 1) while healthy people estimated that the quality of life of a dialysis patient would be just .39. Similar discrepancies have been demonstrated for other serious health conditions (Boyd, Sutherland, Heasman, Tritcher & Cummins, 1990; Buick & Petrie, 2002) and for some less serious health conditions (Baron, Asch, Fagerlin, Jepson, Loewenstein, Riis, Stineman & Ubel, 2003).

Although the existence of the discrepancy is well established at this point, its cause is not (Ubel, Loewenstein & Jepson, 2003). The most common explanation for the discrepancy is that healthy people think that health impairments are worse than they are. Research in diverse domains has documented a general tendency for people to underestimate their own and others' speed of adaptation to negative as well as positive outcomes (Gilbert, Pinel, Wilson, Blumberg & Wheatle, 1998; Sieff, Dawes & Loewenstein, 1999). A related explanation is that healthy people overestimate the impact of patients' sickness on their overall well-being, ignoring the many aspects of life that are unchanged (or even improved) by the health condition. Again, research in diverse domains has documented such a 'focusing illusion' (Schkade & Kahneman, 1998; Ubel, Loewenstein, Hershey, Baron, Mohr, Asch & Jepson, 2001; Wilson, Wheatley, Meyers,

Gilbert, Axsom, 2000) – a general tendency to exaggerate the impact of anything one's attention is focused on.[1]

Other explanations for the discrepancy implicate patients as well, rather than just the healthy respondents. For a variety of reasons, patients may exaggerate their well-being. In surveys, they may report a high quality of life even though they experience more moments of misery and fewer moments of joy than healthy people. Patients may be in the habit of exaggerating their reports of well-being for the benefit of those who provide care for them (Diener, Suh, Lucas & Smith, 1999), or they may simply need to avoid being perceived as complainers so as to ensure continued support from family, friends and professional caregivers. They may also have developed a manner of coping with their hardship whereby they tend to focus on more positive aspects of their experiences.

Even if these motivational accounts do not lead to exaggeration, basic memory processes may have the same effect. In general, summary reports of extended experiences tend to overweight certain salient features, such as the peak, end, or trend (for reviews see Ariely & Carmon, 2002; Kahneman, 1999, 2000), and they tend to be insensitive to duration. Patients may experience frequent, lengthy periods of moderately negative mood, but these periods may not be well represented in an overall happiness judgment if, say, the patients experience normal, positive peak moods. Evidence suggests that the peak moods, however brief, will have a greater weight in an overall judgment than will the lengthy periods of mildly negative mood (Kahneman, Frederickson & Schreiber, 1993). Patients may, thus, have very low average moods, but they may remember and report their experience as quite positive, leading to exaggerated reports of well-being.

If, for any of these reasons, patients do exaggerate their well-being, then the low well-being ratings that healthy people estimate for patients would correctly reflect the difference in day to day mood experienced by patients and healthy people. The discrepancy in well-being ratings would reflect reality.

A related explanation for the discrepancy comes from evidence that healthy people tend to underestimate their own mood. Thomas and Diener (1990) have suggested that people may generally tend to recall negative times more readily than positive times. Indeed they found that (healthy) people tend to underestimate the percentage of time that they are in positive relative to negative moods. The same mechanisms that may lead patients to exaggerate their well-being, may simply attenuate this negative bias in their mood memory. For example, a coping strategy that leads them to focus on positive aspects of life, would make them less likely than healthy people to remember their more negative moments. The tendency of controls to recall more negative experience could make them more likely than patients to understate their own well-being. If controls understate their own well-being, then they would be likely to understate the well-being of other people (including patients) as well, and this could contribute to the discrepancy.

The cause of the discrepancy matters. If healthy people misperceive the quality of life associated with different health conditions, they are likely to make sub-optimal decisions, both when it comes to protecting their health and when it comes to deciding between treatments. For example, someone suffering from a gastrointestinal disorder who overestimates how miserable it would be to have a colostomy would be likely to be too reluctant to obtain one. The discrepancy could also lead policy makers to misallocate scarce health care funds if different conditions are subject to different degrees of bias. To

make informed decisions, laypersons and policy makers need to form accurate perceptions of how various health conditions would affect their lives and their well-being.

The goal of the research reported here is to specifically test for the three possible sources of the discrepancy discussed above – misestimation by healthy people, overstatement of mood experiences by patients, and understatement of mood experiences by healthy people. This is accomplished by assessing the real mood of patients and controls, and by comparing the reality to various estimates and recollections of mood. To measure mood, we use the method of Ecological Momentary Assessment (EMA) (Kahneman, 1999; Stone and Shiffman & DeVries, 1999).

In EMA, subjects are given personal digital assistants ("PDAs", e.g. Palm Pilots) to carry with them, wherever they go for a period of several days or more. The method is designed to minimize the influence of biased recall. The PDA prompts the subject to answer questions at several random times throughout the day. In studies of well-being, subjects are asked how they feel at that very moment. Robinson and Clore (2003) found that subjects are faster to answer questions about their immediate mood than about moods in the past. They argue that momentary mood reports are, thus, less likely to reflect biases of episodic or semantic memory. In this sense, momentary mood questions, asked repeatedly and in the subject's normal environment, are less likely to be edited by a subject who wants to represent himself in a certain way. By asking people to make such momentary assessments several times a day on a PDA, researchers can then aggregate these assessments to get an overall picture of an individual's experience, in a way that avoids the biases of the individual's own recall or aggregation process.

In this study, we compared mood estimates (made during interviews) and EMA mood responses (made over the course of a week on a PDA), reported by a sample of chronically ill hemo-dialysis patients and a sample of healthy controls. If the discrepancy in well-being ratings occurs because patients tend to exaggerate their own mood, then their mood estimates given at the beginning of the week and their mood recollections made at the end of the week should overstate their actual mood experience, as indicated by their aggregated EMA responses. If this effect is large, then we would expect the estimates and recollections of the patients to be similar to those of the controls, but their actual mood should be significantly lower. And the controls estimates of patients' moods may accurately reflect this lower mood.

Alternatively, if it is controls' underestimation of their own well-being that accounts for the discrepancy, then we would expect their estimations and recollections of mood to underestimate their actual mood. Patients, in contrast, may be quite accurate.

On the other hand, if patients have largely adapted to their condition, then we would expect their aggregated EMA responses to he high, perhaps as high as those of controls. If the discrepancy results from controls' inability to appreciate this adaptation, then the controls' estimates of the mood they would experience as dialysis patients should understate the patients' actual mood experience (as measured by EMA)

Finally, to the extent that adaptation does occur, we hoped to assess the extent to which patients have insight into its occurrence. We did so by asking them to estimate how happy they would be if they had never had kidney trouble and had never needed dialysis treatment. If patients have adapted, and are aware that they have adapted, then they should not expect good health to have made them any happier than they are now. Also, their

estimates of happiness-if-healthy should not reflect a higher level of happiness than is reported by actual healthy non-patients.

Method

Subjects

Subjects were 49 end stage renal patients receiving hemo-dialysis treatment 3 times per week, and 49 healthy controls who were matched to the patients on age, race, gender and education. Subjects in thirty-two of the matched pairs were paid $30 to participate. The other subjects were paid $50 to participate.[2]

End-stage renal disease is a condition in which the kidneys fail to perform their normal function of cleaning and filtering the blood. Treatment consists of a procedure called hemo-dialysis in which a patient's blood is filtered through a machine. Most patients require treatment 3 times per week for about 3 hours each time. While discomfort and nausea are possible, they are usually minor and the patient can read, write, talk, eat, sleep or watch TV. The patient's lifestyle can include most normal activities, including work, exercise, and leisure; however, the patient will feel fatigued after missing treatment for several days. Also, the patient must follow a strict diet which usually involves reduction of sodium intake, consumption of relatively little meat, and drinking only small amounts of fluids.

Patients were selected from a pool of 299 dialysis patients at 9 dialysis centers in the Ann Arbor area. Forty-five of the patients were deemed ineligible because records or a preliminary interview showed that they were either blind, deaf, illiterate, non-English speaking, or were experiencing dementia. The remaining patients were approached at the

centers while they were undergoing the treatment. After hearing the description of the study, 90 of the remaining patients (35%) agreed to participate.

Of the 90 patients who began the study, 82 completed all three phases (to be described later). Sixty-nine of these patients responded to at least 40% of the prompts during the EMA phase of the study. Healthy, matched controls were sought for these 69 patients. Control subjects were recruited from advertisements placed in a local paper. Individuals responding to the advertisements were first screened in a telephone interview to ensure that they did not suffer from any major health conditions. Controls were sought to match individual patients on gender, race, age (within 6 years), and years of education (within 4 years). Fifty-six controls were recruited. Seven did not complete all phases of the study (including 2 whose response rate during the EMA phase of the study was less than 40%). Forty-nine were successfully matched to a patient.

Of the 49 matched pairs, 31 were female, 30 were Caucasian, 18 were African American, and 1 was Hispanic. The average age of the pairs was 49 years (SE = 2.0) and the average length of education was 15 years (SE = .26). As intended, the controls' self-reported health was better than that of the patients (M = 2.2, SE = .18 and M = 3.8, SE = .16) respectively, on a 5-point scale from 1 (Excellent) to 5 (Poor), $t(29) = 6.8$, $p < 0.001$, $d = 1.7$.[3]

Procedure

Non-participant Questionnaire

To test for the possibility of selection bias among the patients, the "non-participants" (i.e., patients who were asked to participate, but refused), were given a questionnaire to assess their average mood. Specifically, they were asked to rate their average mood during a typical

week, on a 5-point response scale with the following response options: +2 (“very pleasant”), +1 (“slightly pleasant”), 0 (“neutral”), -1 (“slightly unpleasant”), and -2 (“very unpleasant”).

The questionnaire was given to 126 of the 164 non-participants (Due to an administration error, the questionnaire was not given to the first 38 non-participants). These patients were asked to complete the questionnaire while they were at the center and undergoing treatment, and many reported being simply too tired from the (concurrent) treatment to complete the questionnaire. Still the response rate was quite high (79%).

Three Stages

There were three stages to the study: 1) the entry interview, 2) the EMA week, and 3) the exit interview. The mood estimation tasks used in the entry and exit interviews were designed to allow direct comparison to the EMA responses.

Entry Interview

During the entry interview, subjects were first shown how to use the PDA. They answered several sample questions to familiarize themselves with how the screen could be tapped to enter responses. They were also shown how the PDA could be put to “sleep” so that it would not disturb them at night, or at other inappropriate times (e.g., while attending a movie).

Next, subjects completed a questionnaire in which they were asked to estimate their typical mood:

Entry. Subjects estimated the percentage of time during a typical week that they would experience each of 5 mood levels. The percentages were to add to 100. The mood levels were the same ones that would be offered as response options during the EMA

session, and they were the same ones that were used in the non-participant questionnaire (i.e., from +2 (“very pleasant”) to -2 (“very unpleasant”)).

Ecological Momentary Assessment (EMA)

Subjects carried the PDAs for 7 days, beginning immediately after the entry interview.[4] The first 7 pairs of subjects received PDAs scheduled to beep randomly once within each 2 hour interval of the day (e.g., 8:00 – 10:00, 10:00-12:00, etc.). The remaining 42 pairs received PDAs that were scheduled to beep randomly within each 90-minute period.[5]

The first question was a single item mood measure, which instructed subjects to think back to the mood they were feeling just before the PDA beeped, and to tap the button on the screen that best described that mood. There were five response buttons, representing the same 5 mood levels that have been described previously (i.e., from +2 (“very pleasant”) to -2 (“very unpleasant”)).

There were nine additional mood measures. Subjects indicated the extent to which they were experiencing 9 different emotions. These items, previously used by Thomas and Diener (1990), were happy, joyful, pleased, enjoyment/fun, depressed/blue, unhappy, frustrated, angry/hostile, worried/anxious. For each item, there were 7 response options, labeled 0 to 6, and anchored at “Not at all” and “Extremely much”.

Two additional questions asked subjects about the extent to which they were feeling 1) pain or physical discomfort, and 2) tired/fatigued. There were five response buttons, each with a number on it, and a descriptor beside it. These were as follows: 1) Not at all, 2) A little, 3) Moderately, 4) Quite a bit, 5) Extremely.

Finally, on ten percent of the prompts, an additional question was asked, and on these trials, this additional question was presented first, before the twelve questions described above. This was an overall life satisfaction question, and since participants usually answered the momentary mood question first, an introduction screen was shown before the question screen was shown, to emphasize the difference. The introduction screen said "The next question will ask you about how you feel about your life as a whole." The question on the following question screen was then simply "How do you feel about your life as a whole?" There were seven response buttons corresponding to a 7-point scale ranging from –3 to +3, and anchored at "Very Unsatisfactory" and "Very Satisfactory".

Exit Interview

When the PDAs were returned at the end of the week, subjects were asked to make several mood estimates. These estimates were in the same format as the estimate that was made during the entry interview (i.e., estimates of the percentage of time spent in each of the 5 mood levels). The different estimation instructions are described below:

Recall. Subjects estimated the percentage of time spent in each mood level during the previous week (during which they had carried the PDA).

Typical. Subjects estimated the percentage of time spent in each mood level during a typical week. This was identical to the estimation task given at the entry interview.

Dialysis. All subjects were presented with a scenario describing the experience of a dialysis patient. Patients and controls alike were asked to imagine that they were the patient in the scenario. Controls were asked to imagine that they had been dialysis patients for either one year, or for as long as their matched patient had been on dialysis (whichever was

greater). All subjects then estimated the amount of time they would spend in each mood level if their experience were the same as that of the patient in the scenario.

Healthy. Patients were asked to estimate the percentage of time they would spend in each mood level if they had never had kidney problems and if they had never needed dialysis treatment.

Results

Analyses are reported for the 49 pairs of matched patients and controls. All t-tests are paired, unless otherwise specified.

Actual Mood: EMA Responses

The average EMA response to the overall mood question (on the +2 to –2 scale) was 0.70 (SE = .07) for the patients and 0.83 (SE = .07) for the controls. This difference is not significant, $t(48) = 1.4$, $p = 0.16$. We can, of course, not rule out the possibility that there is, in fact, a difference in average mood between the groups. With this sample size and a critical value of .05, we had the statistical power (.80) to detect differences of .23 (on the 5-point scale) between the groups. It appears likely then, that if there is a difference in mood between the groups, the difference is small. This supports the suggestion that the patients have adapted quite well to their condition. [6]

Further evidence of a lack of difference in mood comes from responses to other EMA questions. For each subject, the responses to the 4 positive emotion questions were averaged (i.e., happy, joyful, pleased, enjoyment/fun), as were the responses to the 5 negative emotion questions (i.e., depressed/blue, unhappy, frustrated, angry/hostile, worried/anxious). The average positive responses were not significantly different between groups ($M = 3.1$, $SE = .17$ for patients, and $M = 3.3$, $SE = .16$ for controls), $t(48) < 1$, nor

were the average negative responses (M = 0.9 , SE = .13 for patients, and M = 1.0 , SE = .12 for controls), t (48) < 1. There were also no significant differences between groups on any of the 9 individual emotions.

There were no significant differences between patients and controls in their responses to the questions about pain (M = 1.4, SE = .19 and M = 1.1, SE = .19, respectively), t(48) < 1, tiredness (M = 2.1, SE = .19 and M = 2.1, SE = .17, respectively), t(48) < 1, or overall life satisfaction (M = 1.1, SE = .18 and M = 1.3, SE = .16, respectively), t(45) < 1.

The response rates for both groups were high. Patients responded to 72% of the PDA prompts (SE = 2.1) for an average of 39 responses over the 7 days, while the controls responded to 78%, (SE = 2.1) for an average of 43 responses. The response rates were significantly different, t (48) = 2.3, p = 0.025, d = 0.43, however EMA response rate was not a significant predictor of average EMA mood when entered alone or when entered simultaneously into a regression with the group variable (i.e., patient vs. control). If either group were systematically not responding during times of peak high or low emotion, thus biasing the average of self-reported mood, then that average should be correlated with response rate.

Average Mood Estimates

For each of the estimation tasks (i.e., entry, recall, typical, dialysis, healthy), subjects estimated 5 mood percentages. For some subjects, these percentages did not add to 100, so all percentages were scaled, by dividing each percentage by the sum of the 5 percentages, and multiplying the quotient by 100. These scaled percentage estimates corresponded to a scalar estimated average mood, and those averages are shown in Table 1.

The averages were computed by dividing each scaled percentage by 100 and multiplying the quotient by the respective mood level (i.e., +2, +1, 0, -1, or -2). The sum of these 5 values was the estimated average mood for that particular estimation task.

Estimation of One's Own Mood

For each subject, the difference between the entry estimate and actual EMA average was computed. The patients' entry estimates slightly overestimated their EMA average mood ($M = +0.08$, $SE = .08$), although not significantly so, $t(48) = 1.0$, $p = .32$, while the controls significantly underestimated their EMA average mood ($M = -0.16$, $SE = .08$), $t(48) = 2.1$, $p = .042$. A t-test comparing these estimation errors was significant, $t(48) = 2.1$, $p = 0.042$, $d = 0.44$.

A similar pattern was observed for the recall estimates. While the patients' recall estimates accurately reflected their EMA average mood ($M = 0.0$, $SE = .06$), the controls again showed significant underestimation ($M = -0.23$, $SE = .07$), $t(48) = 3.3$, $p = .002$. Again, these errors were significantly different between groups, $t(84) = 2.7$, $p = 0.009$, $d = 0.49$.[7]

In sum, there was little indication that patients exaggerate their mood. In fact, their expectations and recollection were quite accurate. On the other hand, the controls' expectations were worse than their actual experience, and, consistent with prior research (Thomas and Diener, 1990), their memories showed the same pattern. (Alternative explanations of this finding are refuted in the appendix.) If healthy people tend to recall a more negative experience, then this should lead them to give deflated judgments of well-being when they evaluate their lives. If their judgments of their own well being are deflated in this sense, then it is likely that their estimations of others' well-being would be deflated

as well. And this may be at least part of the reason why they underestimate the well-being of patients.

Imagining Dialysis

For control subjects, the average estimate of mood while imagining life under the dialysis scenario was -0.38 (SE = .11). This was a very large and significant underestimation of the actual EMA average mood of the dialysis patients (M = 0.70, SE = .07), $t(48) = 7.6$, $p < 0.001$, $d = 1.7$, and of the patients' own estimation of the dialysis scenario (M = 0.63, SE = .12), $t(48) = 5.6$, $p < .001$, $d = 1.3$.[8] The controls' estimates were significantly negative, $t(48) = 3.43$, $p < .001$, while the patients' experience was significantly positive, $t(48) = 9.95$, $p < .001$. Importantly, the patients' average for a typical week (at exit, M = 0.61, SE = .10) and for the dialysis scenario (M = 0.63, SE = .12) were very similar, $t(48) < 1$, suggesting that the scenario was a fair representation of their condition.

Mood if Healthy

For patients, the mood estimates when imagining that they had never had kidney trouble or needed dialysis treatment (M = 1.16, SE = .09) were higher than their mood estimates for a typical week (at exit, M = 0.61, SE = .10), $t(48) = 5.8$, $p < 0.001$, $d = 0.84$, and higher than their actual EMA response average (M = 0.70, SE = .09), $t(48) = 4.9$, $p < 0.001$, $d = 0.82$. Furthermore, their healthy estimate was higher than the controls' estimate of a typical week (at exit, M = 0.67, SE = .09), $t(48) = 4.0$, $p < 0.001$, $d = 0.79$, and higher than the controls' actual EMA response average (M = 0.83, SE = .07), $t(48) = 3.1$, $p = 0.003$, $d = 0.58$. These findings support the suggestion that patients are themselves not aware of the extent to which they have adapted to their condition.

Discussion

In short, we failed to find evidence that patients experienced lower moods than healthy controls. Both patients and controls, however, predicted that the difference in mood experienced under health and illness would be large. We also failed to find evidence that patients exaggerate their mood, although we did find that healthy people seem to understate their own mood. These findings are discussed below.

It appears that dialysis patients do, largely at least, adapt to their condition. Although they report their health as being much worse than that of healthy controls, they do not appear to be much, if at all, less happy than people who do not suffer from kidney disease or from any other serious health condition. The EMA procedure greatly reduced the likelihood of response biases. Subjects were asked about their mood repeatedly, at different times of the day, throughout the normal routine of their lives. If the patients really did spend a great deal of time in a depressed mood, then this procedure should have picked it up. The previously observed tendency of healthy people to underestimate the reported quality of life of people with various health conditions does seem to be due, in large part, to their misperception of the extent to which people can adapt to such conditions.

We cannot rule out the possibility of scale renorming, that is, the possibility that the EMA response options meant something different to the patients than to the healthy controls. For example, what a patient reports as a “very pleasant” feeling may only be reported as “slightly pleasant” by a healthy person, because the patients’ standards may have lowered. We doubt very much, however, that this is the case. We have investigated the same phenomenon in quality of life evaluations, and, contrary to the hypothesis of scale renorming, have found that the discrepancy is actually greater when quality of life is

measured with scales that have well-defined demarcations (Baron et al., 2003), and that the discrepancy persists even when we use scales that are not susceptible to recalibration (Pond, Fagerlin, Loewenstein, Smith, Riis & Ubel, 2004).

Headey and Wearing (1992) have argued that individuals have a baseline mood level to which they return after events move them temporarily above or below that baseline. Consistent with this account are findings from a twins study, which suggests that genetic variation, and not variance in life circumstances, accounts most of the variance in well-being across individuals (Lykken & Tellegen, 1996). The current finding is also consistent with this "baseline" account. Although we do not have measures of well-being for people who recently became sick, we do find that dialysis patients, who have endured illness and uncomfortable circumstances for months or years, are experiencing normal (or at least close to normal), positive mood levels.

Interestingly, the dialysis patients themselves seem unaware of the extent to which they have adapted. They believe that they would be happier if they had never been sick, yet they appear to be incorrect in this belief, as they are already about as happy as healthy people. In imagining a life that had always been free of illness, they may instead imagine the initial mood increase that would follow the transition from their current state to one of good health (Kahneman, 1999), and they may assume that the feeling from such a transition would result, not from the transition, but from the better quality of experience in the healthy state.

Healthy people are clearly unaware of the extent to which adaptation to dialysis occurs. Their estimates of the moods that they would experience if they were on dialysis were much lower than the actual moods experienced by the dialysis patients. In fact, they

estimated that their average mood would be negative if they were on dialysis, while the patients themselves actually experienced positive average moods. This is a rare case where people incorrectly estimate even the valence of a different life circumstance (Wilson & Gilbert, 2003). The discrepancy in well-being ratings reported by other researchers (e.g., Sackett & Torrance, 1978), thus, does not seem to be a mere artifact of different response processes used by patients and healthy people when they answer questions about their overall well-being. The surprisingly high ratings that are often given by patients seem, at least in the case of dialysis patients, to reflect a genuinely high frequency and intensity of positive mood. To our knowledge, this is the first study to show that healthy people grossly underestimate sick people's actual quality of emotional experience.

That said, we do find some evidence that part of the discrepancy may be accounted for by differences in the manner in which patients and healthy people make overall or summary reports of well-being. Consistent with prior research (Thomas & Diener, 1990), healthy people tended to slightly underestimate their own actual average mood. This was the case for the mood estimations (of a typical week) that the participants made at the entry interview, and for the recall estimate at the end of the week. Patients, on the other hand, did not underestimate their actual mood. It is possible that in coping with their hardship, patients have developed a tendency to focus more on positive experience. This may be a crucial part of the adaptation process and should be an area of future investigation.

This difference in recall tendencies could account for some of the discrepancy between patients' self-reports of well-being and healthy persons' predictions of patients' well-being. If healthy people tend to recall a more negative experience, then this should lead them to give slightly deflated judgments of well-being. If their judgments of their own well being

are deflated in this sense, then it is likely that their estimations of other's well-being would be deflated as well. This may contribute to their underestimation of the well-being of patients. Still, although significant, the recall difference effect was not large. Most of the discrepancy appears to be due to healthy patients simply not recognizing how much positive affect a typical dialysis patient would experience.

There are undoubtedly some circumstances to which people cannot adapt, but we seem to underestimate the severity of such limiting circumstances. For most of us, it would take a lot more than we think to make us permanently miserable. The current study provides what is, to date, the most convincing demonstration of this fact. Healthy people think that dialysis would lead to a much more miserable life than it, in fact, does. But this misperception will be a difficult one to correct. Even dialysis patients, who have themselves experienced adaptation, seem not to appreciate the extent of their own adaptation. Getting others to appreciate it will surely be more difficult.

Concluding Comments

Ignorance of adaptation can have negative consequences for decision-making. It can cause individuals to opt for unnecessarily risky surgeries, and policy makers to invest in programs that have a minimal impact on people's well-being, possibly at the expense of programs that really do prevent misery. This is not to say that research and treatment of kidney disease should not continue to be priorities. Indeed, dialysis treatments keep kidney patients alive. But, in making difficult policy decisions, consideration of the actual moods experienced by patients may influence priorities between serious conditions, such as, for example, paraplegia and depression.

Further investigation of the relationship between mood and retrospective reports of well-being is warranted. The relationship seems to differ between different national populations. For example, Oishi (2002) found that when American and Asian subjects had similar levels of mood, the Americans tended to recall more positive levels of mood than the Asians (see also Riis & Kahneman, 2004). Our finding that mood recall is more accurate among dialysis patients than among healthy patients is the first evidence that the relationship between mood and retrospective reports may differ for different health populations as well. This line of research will be aided by developments in EMA aimed at reducing subject burden (Kahneman, Krueger, Schkade, Schwarz & Stone, 2004) and at improving and validating its accuracy (Kahneman & Riis, in press).

When evaluating their quality of life, and when making decisions about how to improve that quality of life, people certainly think about dimensions other than mood (Fredrickson, 2000). Meaning, achievement, and identity are some of the other things that people value, and these may be quite independent of mood. Healthy people may fear illness, not just because of its influence on mood, but because of its influence on these other dimensions. But, insofar as mood is an important dimension of quality of life, healthy people's apparent exaggeration of the influence of illness on mood will lead to incorrect perceptions of how illness will influence quality of life.

References

Ariely, D. & Carmon, Z. (2002). Summary Assessment of Experiences: The Whole is different from the Sum of its Parts. In G. Loewenstein, D. Read, and R. Baumeister (Eds.) *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice.* New York: Russell Sage Foundation Press.

Baron, J., Asch, D. A., Fagerlin, A., Jepson, C., Loewenstein, G., Riis, J., Stineman, M. G., & Ubel, P. A. (2003). Effect of assessment method on the discrepancy between judgments of health disorders people have and do not have: A web study. *Medical Decision Making, 23(5),* 422-434.

Boyd, N., Sutherland, H. J., Heasman, K. Z., Tritcher, D. L., & Cummings, B. (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 & Social Psychology, 36(8),* 917-927.

Buick, D.L., & Petrie, K. J. (2002). “I know just how you feel": The validity of healthy women's perceptions of breast cancer patients receiving treatment. *Journal of Applied Social Psychology, 32(1),* 110-123.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125(2),* 276-302.

Fredrickson, B.L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion, 14(4),* 577-606.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. Journal of *Personality & Social Psychology, 75(3),* 617-638.

Headey, B., & Wearing, A. (1992). *Understanding happiness: A theory of subjective well-being.* Melbourne, Australia: Longman Cheshire.

Kahneman, D., Fredrickson, D.L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4*, 401-415.

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well being: The foundations of hedonic psychology* (pp. 3-25). New York, NY, US: Russell Sage Foundation.

Kahneman, D. (2000). Evaluation by moments: Past and future. In D. Kahneman and A. Tversky (eds.), *Choices, Values and Frames* (pp. 293-308). New York: Cambridge University Press and the Russell Sage Foundation.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwartz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: the day reconstruction method (DRM). Manuscript under review.

Kahneman, D., & Riis, J. (in press). Living, and thinking about it: Two contrasting views of life. In F. Huppert, B. Keverne, & N. Baylis (Eds.), *The Science of Well-Being.*

Lykken, D. & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7(3),* 186-189.

Oishi, S. (2002). The experiencing and remembering of well-being: A cross-cultural analysis. *Personality and Social Psychology Bulletin*, 28(10), 1398-1406.

Pond, H. M., Fagerlin, A., Loewenstein, G., Smith, D., Riis, J., & Ubel, P. A. (2004). Unbelievably happy? Exploring whether scale recalibration accounts for the happiness gap. Manuscript in preparation.

Riis, J. & Kahneman, D. (2004). Cross-national differences in well-being: Are the Americans really happier than the French? Manuscript under review.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin,* 128(6), 934-960.

Sackett, D. J., & Torrance, G. W. (1978). The utility of different health states as perceived by the general public. *Journal of Chronic Diseases, 31,* 697-704.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9(5),* 340-346.

Sieff, E.M., Dawes, R.M. & Loewenstein, G. (1999). Anticipated versus actual reaction to HIV test results. *American Journal of Psychology, 112(2),* 297-311.

Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well being: The foundations of hedonic psychology* (pp. 26-39). New York, NY, US: Russell Sage Foundation.

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality & Social Psychology, 59(2),* 291-297.

Ubel, P.A., Loewenstein, G., Hershey, J., Baron, J., Mohr, T., Asch, D., Jepson, C. (2001). Do nonpatients underestimate the quality of life associated with chronic health states because of a focusing illusion? *Medical Decision Making, 21(3),* 190-199.

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(6),* 599-607.

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 & Social Psychology, 78(5),* 821-836.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. Zanna (Ed.), *Advances in experimental social psychology (pp. 345-411).* New York: Elsevier.

Appendix

For the last 17 pairs of subjects, an additional recall task was added to test for a possible alternative explanation of this result. Since the subjects did miss some EMA prompts during the week, it is possible that the EMA responses presented a biased representation of weekly mood, and that this representation was differently biased for patients and controls. Perhaps, say, patients tended to miss EMA prompts when they were in a bad mood, and controls tended to miss EMA prompts when they were in a good mood. In the original recall estimation task though, subjects were asked to estimate the percentage of time spent in each mood category in the past week. So any discrepancies between the recall estimate and the actual EMA responses could be due, not to memory distortion, but to biases in EMA reporting. Specifically, subjects may have accurately recalled their actual mood, but if EMA did not reflect actual mood, then EMA would diverge from recalled mood too (as in fact it did for controls).

To allow for a test of this possibility, an additional recall task was added for the last 17 pairs of subjects. After making the original recall estimation, and the rest of the exit interview estimations, they were also asked to recall the percentage of time that they made each of the 5 different EMA responses during the past week. While the original recall question asked them to recall their actual mood, this question just asked them to recall their EMA responses. A discrepancy between the two recall measures would suggest that any difference in recall bias could be an artifact of EMA response rate bias. But no such discrepancy was found. The difference between the original mood recall measure and the EMA response recall measure was calculated for each subject in the 17 pairs for which the data were available. The difference was virtually 0 for both groups of subjects ($M = 0.03$,

SE = 0.07 for patients, and M = 0.00, SE = 0.05 for controls) and the difference was not significant, t (16) < 1. Furthermore, the same pattern of recall error that was observed in the full set of 49 pairs was observed for both recall measures in this sub-sample. For the patients, both recall measures were accurate – the difference between mood recall and average EMA response was 0.0 (SE = .11), and the difference between their EMA recall and average EMA response was -0.03 (SE = .14). For the controls, both recall measures tended to understate EMA average. For the controls, the difference between mood recall and average EMA response was -0.20 (SE = .13), t(16) 1.53, p = 0.14, and the difference between their EMA recall and average EMA response was also -0.20 (SE = 13), t(16) = 1.52, p = .15. While the sample size is small, none of these results suggests any support for the hypothesis that the differential recall error between patients and controls that was observed in the full set of 49 pairs occurred because of differences in EMA bias.

A regression analysis on the full set of 49 pairs tests the same hypotheses. If a low EMA response rate indicates a bias in EMA average, then people who miss more beeps should have a systematic bias in the recall estimation. Furthermore, if this response rate bias is different for patients and controls, then response rate should predict recall error differently for patients and controls. To test this, we regressed the recall error (i.e., recall estimate minus EMA average) on response rate (i.e., proportion of EMA prompts the subject responded to), group (i.e., patient or control), and the interaction term. The interaction term was not significant, beta < 0.001 (t < 1), which does not support the hypothesis that the relationship between error rate and response rate was different for the different groups. When the interaction term was removed, and Response Rate and Group were entered simultaneously, only Group was significant, beta = 0.12 (t = 2.5, p = 0.015).

This does not support the hypothesis that subjects with low response rates made more biased recall estimates

Author Note

Jason Riis, Department of Psychology, University of Michigan; George Loewenstein, Department of Social and Decision Sciences, Carnegie-Mellon University; Jonathan Baron, Department of Psychology, University of Pennsylvania; Christopher Jepson, Division of General Internal Medicine, University of Pennsylvania School of Medicine; Angela Fagerlin, Division of General Internal Medicine, University of Michigan, and Veterans Affairs Medical Center, Ann Arbor, Michigan; Peter A. Ubel, Division of General Internal Medicine, University of Michigan, and Veterans Affairs Medical Center, Ann Arbor, Michigan.

The research was supported by N.I.H. Grant AG16258 (P. Ubel, Principal Investigator). We thank Barbara Fredrickson and Lisa Feldman Barrett for discussion of methodological issues, and Daniel Kahneman, Norbert Schwarz and Dylan Smith for commenting on a previous draft. We thank Linda Nyquist for assistance in recruiting control subjects, and we thank the patients and staff at the participating dialysis centers. We also thank Laura Damschroder, Christine Goldstein, Jennifer Heckendorn, Susan Metosky, Molly Miklosovic, Jennifer Ream, Todd Roberts, Brianna Sarr, Ryan Sherriff, and Penn Stabler for research assistance.

Correspondence concerning this article should be addressed to Jason Riis, Center for Health and Wellbeing, Princeton University, Princeton, NJ 08544. Email: jriis@princeton.edu.

Footnotes

[1] Our group has two studies, however, suggesting that of these two mechanisms, underestimating the speed of adaptation is more plausible than the focusing illusion (Baron et al., 2003; Ubel et al., 2003).

[2] The increased payment was intended to increase recruitment rates, but it had no effect.

[3] Only the first 30 pairs were asked this health question. In an effort to reduce participant burden, a questionnaire pertaining to health which was administered at the end of the study was cut. This question was part of that questionnaire. Our recruiting methods were similar for the remaining pairs, so there is no reason why this very large difference in health would not also be observed for the remaining pairs.

[4] The Palm model was the IIIxe. It ran a program called ESP Blue, which was developed by Chip Jensen. The program can be downloaded at http://www.med.umich.edu/pihcd/esp/esp.htm. ESP Blue is based on a program called ESP, which was developed at Boston College by Lisa Feldman Barrett and programmed by Daniel J. Barrett. While the program was running, users were not able to use the Palm's other programs or functions.

[5]There is not a standard prompt rate in EMA. The rate was increased because, for some analyses (not reported here), we intend to examine the course of mood changes through a patient's weekly dialysis cycle. Increasing the prompt rate in this study did not change the results.

[6] Since our ultimate patient sample was not randomly selected, we examine the possibility that they are not representative, with respect to mood, of the dialysis population.

First, we compare them to the "non-participants" from the dialysis centers, that is, patients who did not want to participate in the study, but who did complete a preliminary mood questionnaire. That questionnaire was on the same +2 to -2 scale as the EMA mood question. The average questionnaire response was .74 (SE = .11) which is similar to the EMA average reported by the 49 matched patients (M = .70, SE = .07), t (146)< 1.

Furthermore, the 20 patients who responded to at least 40% of the EMA prompts but who were not matched with a control subject, were very similar in EMA average to the 49 matched patients (M = .69, SE = .08 and M = .70, SE = .07, respectively), t (67) < 1. Finally, of the 21 patients who did not complete the study, 17 had at least one EMA response and their EMA average (M = .69, SE = .12) was also similar to that of the 49 matched pairs, t (64) < 1. Among all patients, response rate was not correlated with EMA average, r = -.04, ns.

None of these results supports the hypothesis that the matched patients were unrepresentative of the dialysis population.

[7]In an alternate test, EMA average was regressed simultaneously on entry average and on group, and then, separately, on recall average and on group. In both cases, the group coefficient was significant, beta = 0.09 (t = 2.0, p = 0.046) and beta = 0.09 (t=2.7, p = 0.008), respectively, consistent with the results of the t-tests. This rules out the possibility that the significant t-tests were artifacts of the use of difference scores. The significant group coefficients also suggest that the lack of difference between patients and controls in EMA average cannot be attributed to low power.

[8]For the dialysis estimate, 7 of the 49 control subjects were not asked to imagine that they had had the condition for a period of time. They were only asked to imagine that

they were dialysis patients. When these 7 controls are left out, the dialysis average for the 42 remaining controls was -0.35 and the tests were still significant.

Table 1

Means of Actual and Estimated Mood for Hemo-dialysis Patients and for Healthy Controls

	Patients (n = 49)	Controls (n = 49)
Actual Mood (EMA)	0.70	0.83
Estimated Mood		
Typical week (at entry)	0.78	0.67
Recall of EMA week	0.70	0.60
Typical week (at exit)	0.61	0.67
Imagining dialysis scenario	0.63	-0.38
Imagining never having had kidney disease	1.16	NA

Note. All means are on a +2 to –2 scale.