DOI: 10.1111/j.1475-6773.2009.00974.x
METHODS ARTICLE

Methods

How Bad Is Depression? Preference Score Estimates from Depressed Patients and the General Population

Jeffrey M. Pyne, John C. Fortney, Shanti Tripathi, David Feeny, Peter Ubel, and John Brazier

Objective. To compare depression health state preference scores across four groups: (1) general population, (2) previous history of depression but not currently depressed, (3) less severe current depression, and (4) more severe current depression.

Data Sources. Primary data were collected from 95 general population, 163 primary care, and 83 specialty mental health subjects.

Study Design. Stratified sampling frames were used to recruit general population and patient subjects. Subjects completed cross-sectional surveys. Key variables included rating scale and standard gamble scores assigned to depression health state descriptions developed from the Patient Health Questionnaire-9 (PHQ-9) and SF-12.

Data Collection/Extraction Methods. Each subject completed an in-person interview. Forty-nine subjects completed test/retest reliability interviews.

Principal Findings. Depressed patient preference scores for three of six SF-12 depression health states were significantly lower than the general population using the rating scale and two of six were significantly lower using standard gamble. Depressed patient scores for five of six PHQ-9 depression health states were significantly lower than the general population using the rating scale and two of six were significantly lower using standard gamble.

Conclusions. Depressed patients report lower preference scores for depression health states than the general population. In effect, they perceived depression to be worse than the general public perceived it to be. Additional research is needed to examine the implications for cost-effectiveness ratios using general population preference scores versus depressed patient preference scores.

Key Words. Depression, rating scale, standard gamble, cost–utility, health-related quality of life

Health state preference scores assign a quantitative measure of value to specific health states constrained by death (given a score of 0) and perfect health (given a score of 1 or 100). The specific health states used in this context can be an individual's current health or a description of a hypothetical health state. Health

state preference scores are obtained using a variety of methods (Drummond et al. 1997). Health state preference scores form the basis for calculating quality-adjusted life-years (QALYs). Cost per QALY ratios are increasingly used to inform health care resource allocation decisions (National Institute for Clinical Excellence 2004). However, important methodological issues remain regarding the measurement of health state preferences, including who should be the source of the health state preferences used in cost per QALY calculations.

The Panel on Cost-Effectiveness in Health and Medicine recommended using the general population as the source of health state preferences for the reference case analysis (Gold et al. 1996). The Panel's rationale for making this recommendation was based on fairness and minimizing bias, that is, the general population is blind to its own self-interest (unaware of future health problems) and therefore able to provide a less biased assessment of health state preferences. However, in practice, researchers use many sources to generate health state preferences (Brauer et al. 2006). A recent review of cost–utility analyses published between 1998 and 2001 found that 30.3 percent of preference scores were derived from the community, 23.3 percent from patients, 21.0 percent from clinicians, and 18.7 percent from the authors (Brauer et al. 2006). While distinctions are drawn between utility, value, and preference scores (Gold et al. 1996), for simplicity, this paper will use the term "preference score" for each.

Health state preferences obtained from different groups are often similar but can vary widely (Ubel, Loewenstein, and Jepson 2003). Specifically, health state preference scores obtained from patients who have experienced the condition may differ from preference scores obtained from groups who have not experienced the condition. Individuals with the condition may incorporate a greater range of experiences associated with a health state, may accommodate to their current state of health, or may change the way they rate their health in comparison with others (scale recalibration) (Ubel, Loewenstein, and Jepson 2003; Ubel et al. 2005). Within-group differences also exist. For example, the

---

Address correspondence to Jeffrey M. Pyne, M.D., Center for Mental Healthcare and Outcomes Research (152/NLR), Central Arkansas Veterans Healthcare System, North Little Rock, AR 72114; e-mail: jmpyne@uams.edu. John C. Fortney, Ph.D., and Shanti Tripathi, M.S., are with the Center for Mental Healthcare and Outcomes Research (152/NLR), Central Arkansas Veterans Healthcare System, North Little Rock, AR. Jeffrey M. Pyne, M.D., and John Fortney, Ph.D., are also with the Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock, AR. David Feeny, Ph.D., is with the Center for Health Research, Northwest/Hawai'i/Southeast, Kaiser Permanente Northwest Region, Portland, OR; Peter Ubel, M.D., is with the Center for Behavioral and Decision Sciences in Medicine, Ann Arbor VAMC and University of Michigan, Ann Arbor, MI. John Brazier, Ph.D., is with the Health Economics and Decision Science, School of Health and Related Research, Sheffield, UK.

severity of illness (Badia et al. 1996; Lenert, Treadwell, and Schwartz 1999; Insinga and Fryback 2003) and the length of time since a health event (Adang et al. 1998; Smith et al. 2006) may impact health preference scores.

A number of studies have compared health state preference scores generated by different groups. Some of these studies have found differences based on health experience (Gabriel et al. 1999; Lenert, Treadwell, and Schwartz 1999; De Wit, Busschbach, and De Charro 2000; Postulart and Adang 2000; Insinga and Fryback 2003; Rashidi, Anis, and Marra 2006) while others have not (Balaban et al. 1986; Revicki, Shakespeare, and Kind 1996; Dolders et al. 2006). In general, studies that compare patient and general population health state preferences find that patients assign preference scores to less than perfect health states that are equal to or greater than the preference scores assigned by members of the general population (Sackett and Torrance 1978; Balaban et al. 1986; Froberg and Kane 1989b; De Wit, Busschbach, and De Charro 2000; Dolders et al. 2006). A conclusion that could be drawn from these studies is that using general population health state preferences might result in more favorable cost per QALY ratios than using patient preferences, except in cases of life-saving interventions (Brazier et al. 2005). For example, if the general population assigns a lower preference score than patients to a less than perfect health state, then using general population preferences for an intervention that restores perfect health would result in a larger QALY difference and a more attractive cost per QALY ratio. Conversely, a life-saving intervention for unhealthy patients could appear less cost-effective using general population preference scores because the patient would return to a health state the general population assigned a lower preference score to.

Our study explored whether depression experience influenced depression health state preferences and how this might affect cost per QALY calculations. We chose depression because depression is often misunderstood and stigmatized by the general population (Link et al. 1999; Barney et al. 2006; Perry et al. 2007). The objective of this study was to compare depression health state preferences across four groups: (1) general population, (2) patients with past depression but not currently depressed, (3) patients with mild to moderate depression, and (4) patients with moderate to severe depression.

METHODS

Design

Our study was a cross-sectional, face-to-face survey of individuals sampled from the following recruitment sites: general population, primary care clinics, and

specialty mental health clinics. Our recruitment target for the general population sample was 100, and we recruited 95. From the clinic sites, we attempted to recruit subjects with a broad range of depression severity (see Table 1). Our recruitment target from the clinic sites was 300, and we recruited 246. We also collected test–retest reliability data within 2 weeks of the baseline interview from 49 randomly selected subjects (15 from the general population and 34 from the clinic sites).

Subjects

Eligibility criteria for all groups included (1) age 18–70 years, (2) able to read and understand English, (3) negative screen for significant cognitive impairment as evidenced by diagnosis of dementia or a score >8 on the Blessed Orientation–Memory–Concentration test, (4) no history of schizophrenia diagnosis, (5) negative screen for bipolar disorder, (6) no life-threatening condition, (7) residence within 60 miles of downtown Little Rock, and (8) access to a telephone. Subjects were compensated US$30 to complete the interview. The University of Arkansas for Medical Sciences (UAMS) Institutional Review Board approved the research protocol.

The general population group was recruited from Central Arkansas (Little Rock and surrounding areas) using a commercially available phone list. The Central Arkansas area was selected because the location corresponded with the clinic sites. The phone list included phone numbers, addresses, age, gender, and ethnicity. Potential subjects were selected from the phone list using a stratified random sampling plan to approximate the age, gender, and ethnicity demographic characteristics of Central Arkansas residents. The general population sampling plan did not include depression severity. Potential participants were mailed a postcard stating that they would receive a phone call in 2 weeks about the research study unless they called a toll-free telephone number to decline participation.

From the primary care and specialty mental health clinic sites affiliated with the UAMS, we recruited three patient groups: patients who had past but not current depression, patients with current mild to moderate depression, and patients with moderate to severe depression. Current depression severity was based on reported Patient Health Questionnaire-9 (PHQ-9) severity cut-off scores (Kroenke, Spitzer, and Williams 2001). Patients with history of depression were only recruited from primary care sites, reported that a clinician had made a diagnosis of depression in the past, and had a current PHQ-9 score <5. The groups with current depression were recruited from primary care and specialty mental health care sites. The mild to moderate depression group

Table 1: Demographic, Clinical Characteristics, and Depression Health State Preference Scores across Four Study Groups

Variable	*General Population (N= 95)*	*Depression History But Not Currently Depressed (N= 61)*	*Mild to Moderate Depression (N= 97)*	*Moderate to Severe Depression (N= 88)*
Recruitment *N* (target) by clinic				
Mental health clinic			25 (25)	58 (75)
Primary care clinic		61 (100)	72 (75)	30 (25)
Age				
Mean (SD)	43.1 (13.7)	41.3 (11.3)	42.9 (11.4)	42.5 (11.0)
Gender[3]				
Male	45%	28%	19%	14%
Female	55%	72%	81%	86%
Race				
Caucasian	61%	51%	62%	70%
Other	39%	49%	38%	30%
Marital[1]				
Married/live together	40%	51%	34%	27%
Never married/live alone	60%	49%	66%	73%
Education				
HS grad or less	41%	36%	44%	57%
More than HS	59%	64%	56%	43%
PHQ-9 score[a1,b3,c3]				
Mean (SE)	4.4 (0.60)	2.0 (0.75)	9.5 (0.60)	21.2 (0.63)
Chronic depression[3]				
Yes	11%	11%	34%	64%
No	89%	89%	66%	36%
Current depression treatment[3]				
Yes	20%	52%	82%	89%
No	80%	48%	18%	11%
Ever treated for depression?[3]				
Yes	41%	84%	95%	98%
No	59%	16%	5%	2%
No. of depression episodes[3]				
0	61%	0%	1%	1%
1	16%	11%	12%	16%
2	2%	17%	15%	17%
3 or more	21%	72%	72%	66%
Physical health comorbidity[a1,b3,c3]				
Mean (SE)	2.0 (0.24)	3.0 (0.30)	3.5 (0.24)	4.5 (0.25)
SF-12 preference scores				
Rating scale, mean (SE)				
Mild[c1]	89.5 (1.6)	86.1 (2.0)	88.1 (1.6)	83.6 (1.7)
Moderate[c2]	72.2 (1.8)	69.2 (2.3)	67.1 (1.8)	62.7 (1.9)
Severe[c1]	50.7 (2.2)	50.4 (2.7)	45.9 (2.2)	42.3 (2.3)

continued

Table 1. *Continued*

Variable	*General Population (N= 95)*	*Depression History But Not Currently Depressed (N= 61)*	*Mild to Moderate Depression (N= 97)*	*Moderate to Severe Depression (N= 88)*
Standard gamble, mean (SE)				
Mild[c1]	0.87 (0.02)	0.89 (0.02)	0.87 (0.02)	0.79 (0.02)
Moderate[c1]	0.77 (0.02)	0.80 (0.03)	0.74 (0.02)	0.69 (0.02)
Severe	0.63 (0.02)	0.68 (0.03)	0.63 (0.02)	0.58 (0.03)
PHQ-9 preference scores				
Rating scale, mean (SE)				
Mild[c1]	74.7 (1.9)	75.4 (2.3)	70.8 (1.8)	67.2 (1.9)
Moderate[b2,c3]	62.6 (1.9)	58.6 (2.4)	53.6 (1.9)	49.5 (2.0)
Severe[b2,c3]	43.5 (2.1)	35.8 (2.6)	33.5 (2.0)	30.7 (2.1)
Standard gamble, mean (SE)				
Mild[c1]	0.78 (0.02)	0.83 (0.03)	0.78 (0.02)	0.70 (0.02)
Moderate[c1]	0.70 (0.02)	0.74 (0.03)	0.68 (0.02)	0.63 (0.02)
Severe	0.54 (0.03)	0.59 (0.03)	0.54 (0.02)	0.51 (0.03)
Current health preference scores				
Rating scale[b3,c3]				
Mean (SE)	85.2 (2.0)	80.1 (2.6)	71.0 (2.0)	49.0 (2.1)
Standard gamble[b1,c3]				
Mean (SE)	0.83 (0.02)	0.87 (0.03)	0.74 (0.02)	0.60 (0.02)
Without depression rating scale				
Mean (SE)	90.1 (1.30)	93.0 (1.6)	91.2 (1.3)	86.0 (1.3)
Without depression standard gamble[a1]				
Mean (SE)	0.86 (0.02)	0.94 (0.02)	0.86 (0.02)	0.82 (0.02)

Comparisons between groups was done only for continuous variables; therefore, superscripts using letters are only included for continuous variables: a, general population versus depression history but not currently depressed; b, general population versus mild to moderate depression; c, general population versus moderate to severe depression.

[1]$p<.05$; [2]$p<.01$; [3]$p<.001$.

PHQ-9, Patient Health Questionnaire-9.

had a current PHQ-9 score of 5–14, and the moderate to severe depression group had a PHQ-9 score of 15 or more.

Health state preference scores assign value to health state descriptions on a scale from 0 (equivalent to death) to 1 or 100 (equivalent to perfect health). The next two sections describe the methods we used to generate depression health state descriptions and preference scores.

Depression Health State Descriptions

To create hypothetical depression descriptions, we chose to use the format of two existing, well-validated, and widely used instruments: the PHQ-9 from the

PRIME-MD and the Medical Outcomes Study SF-12. To create the PHQ-9 health state descriptions, we reviewed the PHQ-9 responses of 3,000 primary care subjects that were previously used to validate the PHQ-9 (Kroenke, Spitzer, and Williams 2001). A distribution of the responses (0–3) was generated for each of the nine items within each category of overall severity (none or minimal, mild, moderate, moderately severe, and severe). For example, the most frequent response for subjects in the overall depression severity category of "none or minimal depression" for item #1 (little interest or pleasure in doing things) was 0 (not at all), and the most frequent response for those subjects with severe depression was a rating of 3 ("nearly every day"). Using these item distributions, a modal depression health state description was created for mild, moderate, and severe depression (see Appendix SA2 for depression health state descriptions).

The SF-12 is a 12-item general measure of health status (Ware, Kosinski, and Keller 1996). The SF-12 contains one or two items for the following eight health dimensions: physical functioning, role functioning physical, bodily pain, general health perception, energy/vitality, social functioning, role emotional functioning, and mental health. To create the SF-12 outcome descriptions, we modified the depression descriptions previously reported using cluster analysis methods (Sugar et al. 1998). The modifications included (a) using single responses for each item rather than response ranges, (b) using six dimensions of the SF-12 developed by Brazier and colleagues (Brazier et al. 1998; Brazier, Roberts, and Deverill 2002), and (c) adding a severe depression description. The modifications were needed to facilitate the mapping of valuations to individual items in the SF-12 and to include a severe depression description more consistent with specialty mental health subjects. The result was mild, moderate, and severe depression health state description based on SF-12 items (see Appendix SA2).

Procedures for Eliciting Preference Scores

The preference scoring procedures included simple ranking, rating scale, and standard gamble, in this order. The rating scale preceded the standard gamble to avoid the anchoring effect induced by the standard gamble (Llewellyn-Thomas et al. 1984; Froberg and Kane 1989a, b). Subjects were introduced to the preference score procedures using practice health states and then moved to the depression health state descriptions. The practice health states were "wearing glasses" and "blindness in both eyes." The interviewers were trained to use hard copy rating scale and standard gamble props based on McMaster

University specifications (Furlong et al. 1990). Interviewers randomly started with either the PHQ-9 or the SF-12 depression descriptions and randomly presented the three severity descriptions from each instrument.

Simple ranking of health states used hard copy index cards with the health state described on one side. The subject placed the PHQ-9 and SF-12 cards in order from most to least desirable. The simple rank order of health states was used as a validity check for the rating scale and standard gamble ratings.

The rating scale was presented as a 100 mm line divided into five unit intervals with end points defined as death (0) and perfect health (100). For a given health state, the respondent assigned a number between 0 and 100, which corresponded to the preference score.

The standard gamble method is consistent with von Neumann-Morgenstern expected utility axioms. The standard gamble incorporates choice and risk by setting up a choice between two alternatives: choice A—living in a particular health state with certainty, or choice B—a gamble on a hypothetical treatment for which the outcome is uncertain. The subject was told that a hypothetical treatment will lead to perfect health with a probability of p, or immediate death with a probability of $1 - p$. The subject was then asked to choose between choice A (depression outcome with certainty) or B (the gamble). The probability (p) is varied until the subject is indifferent between choices A and B and the preference score for health state A equals p. We used a ping-pong search procedure where gamble probabilities alternate between high and low values in an iterative search that closes in on the indifference point (Llewellyn-Thomas et al. 1984).

Clinical Characteristics of Subjects

Chronic depression was defined as feeling down, depressed, or hopeless most of the time over the past 2 years without feeling depression free for a period of 2 months or more during this time. Current depression treatment or ever being treated for depression included antidepressant medication or counseling. Current physical health comorbidity was determined from a list of 18 physical health problems.

Statistical Analysis

Categorical demographic and clinical variables were compared using a χ^2 test. Continuous demographic and clinical variables were compared using the general population as the reference category and a general linear model pro-

cedure with the Dunnett post hoc test to adjust for multiple comparisons. Because none of the depression health states were rated worse than death, no adjustments for this response were needed. Similar methods were used to explore the potential influence of current depression on preference scores assigned to hypothetical health states. To do this we examined the preference scores assigned to each subject's current health state and their current health without taking into account the effects of depression.

Test–retest preference scores were obtained on 14.1 percent (49/341) of the total sample: 15 from the general population and 34 from the patient groups. Test–retest reliability was determined using two approaches. First, we calculated the intraclass correlation coefficient. Second, we calculated the difference in hypothetical depression health state preference scores. Differences between subjects from the general population and patient groups were compared using Wilcoxon test and Mann–Whitney U test, respectively.

RESULTS

Table 1 presents a demographic and clinical description of the general population and depression groups. Reflecting the epidemiology of depression, the percent of females in the depression groups was greater than the general population group ($\chi^2 = 27.1$, $p < .001$). Increasing depression severity was also associated with a lower percentage of being married or living together compared with the general population sample ($\chi^2 = 9.3$, $p = .03$).

As expected, depression severity and number of depression episodes were greater in the groups with current depression than the general population group. The group with depression history but no current depression was constrained to have PHQ-9 scores < 5, resulting in this group having a lower depression score than the general population sample ($p = .02$). We did not stratify the general population sample by depression severity and seven subjects (7.4 percent) in the general population sample had PHQ-9 scores of 15 or greater, indicating moderate to severe depression. Depression chronicity ($p < .001$) and history of current ($p < .001$) or any ($p < .001$) depression treatment increased with depression severity. Physical health comorbidity also increased with greater depression severity with the general population reporting significantly less physical health comorbidity than all other groups.

Table 1 reports the preference scores associated with the three depression health states. The overall trend was for a decrease in preference scores as the depression severity of the respondent increased. The comparisons

reported here are between the general population and the other groups because the general population is the recommended source for health state preferences.

Using the SF-12 health states (Table 1), we found significant differences between the general population and moderate to severe depression groups. More specifically, for all SF-12 depression health states (mild, moderate, and severe), the general population rating scale scores were significantly higher than the moderate to severe depression group scores (89.5 versus 83.6, $p = .04$; 72.2 versus 62.7, $p = .001$; 50.7 versus 42.3, $p = .02$, respectively). In addition, the mean sample standard gamble scores for the mild and moderate SF-12 depression health states were significantly higher in the general population group than the moderate to severe depression group scores (0.87 versus 0.79, $p = .02$ and 0.77 versus 0.69, $p = .01$, respectively).

Using PHQ-9 health states (Table 1), five out of six general population rating scale scores were significantly higher than patient groups with current depression. The proportionate differences between the general population and patient groups with current depression also appeared to increase with hypothetical depression health state severity. For example, the proportionate differences between the severe depression group and the general population group increased from 10 percent (67.2/74.7) for the mild PHQ-9 health state to 21 percent (49.5/62.6) for the moderate PHQ-9 health state to 29 percent (30.7/43.5) for the severe PHQ-9 health state. No significant differences were found between the general population group and the depression history only group except for a trend for the severe hypothetical depression health state (43.5 versus 35.8, $p = .05$). Standard gamble comparisons resulted in more limited differences. The mean general population standard gamble scores for the mild and moderate PHQ-9 depression health states were significantly higher than the mean moderate to severe depression group scores (0.78 versus 0.70, $p = .03$ and 0.70 versus 0.63, $p = .03$, respectively).

As expected, patients with current depression rated their current health lower than the general population using the rating scale and standard gamble (see Table 1). For example, the general population rating scale score for current health was significantly higher than the mild to moderate and moderate to severe depression groups (85.2 versus 71.0, $p < .001$ and 85.2 versus 49.0, $p < .001$, respectively), and the general population standard gamble scores were also significantly higher than the mild to moderate and moderate to severe depression groups (0.83 versus 0.74, $p = .02$ and 0.83 versus 0.60, $p < .001$, respectively). However, when subjects were asked to rate their current health without taking into account the effects of depression there were no

significant differences between the general population and depression group scores. There was a significant difference between the general population and the depression history only group standard gamble score for current health without taking into account the effects of depression (0.86 versus 0.94, $p = .049$, respectively).

Test–retest preference score results for the hypothetical depression health states were obtained from 15 general population subjects and 34 patient subjects. The intraclass correlation coefficient for all subjects completing the test–retest procedure was in the fair to good range: 0.519 for visual analog scale (VAS) and 0.522 for standard gamble (SG) scores (Fleiss 1986). We examined the mean rank difference for each group across the 12 different hypothetical depression health states using a nonparametric Mann–Whitney *U* test and found no statistical differences between the general population and patient groups. The absolute difference of mean differences for the VAS scores ranged from 0.18 to 3.80 using the 1–100 scale and from 0.01 to 0.04 for the SG scores using the 0–1 scale.

DISCUSSION

Studies of patient and nonpatient hypothetical health state preferences typically report either no difference or patient preferences exceeding nonpatient preferences (Balaban et al. 1986; Froberg and Kane 1989b; Boyd et al. 1990; Tsevat et al. 1998; Gabriel et al. 1999; De Wit, Busschbach, and De Charro 2000; Ubel et al. 2005; Dolders et al. 2006). Thus, general population preference scores will result in similar if not more favorable cost-effectiveness ratios compared with patient preference scores, except in the case of life-saving interventions. To our knowledge, this is the first study to report patient health state preferences that are consistently equal to or lower than general population preferences. Specifically, individuals with current depression reported lower depression health state preference scores than a general population sample—they perceived depression to be worse than the general public perceived it to be. This finding is most pronounced using the PHQ-9 depression health state descriptions and the rating scale preference method.

The data in this study do not allow us to determine whether discrepancies between patient and public preferences resulted because patients overestimated how bad depression is, or because the general public underestimated how bad it is, or whether both phenomena contributed. Individuals with depression might overestimate the impact of depression through negative

cognitive distortions that are commonly associated with depression. For example, cognitive distortions such as all-or-nothing thinking or overgeneralizing negative events and rejecting positive events are common problems addressed in cognitive-behavioral therapy for depression. Extending this argument to current health state preferences, we would expect individuals with current depression to assign low ratings to their own health with or without depression, and we did not find evidence for this. Instead, depressed patient preference scores for current health state without depression were indistinguishable from the general population or patients with a history of depression only. Other investigators have coined the term "sadder but wiser" to describe depressed subjects' view of reality, while nondepressed subjects view their circumstances as more favorable than they really are (Alloy and Abramson 1979; Seligman 1998). At the very least, these results lend credibility to depressed subject preference scores.

The general population might underestimate the impact of depression because of stigma associated with the disease—the idea that depression is a personal weakness and depressed persons need to pull themselves up by their bootstraps like everyone else. A measure of public stigma was not included in this study, but recent studies suggest that public stigma associated with depression continues to exist (Link et al. 1999; Barney et al. 2006; Perry et al. 2007). Public stigma may result in the general population being less sympathetic to the suffering of individuals with depression and less willing to validate the impact of depression symptoms.

In general, the preference scores for the group with past but not current depression were not significantly different from the general population. The effects of depression on depression health state preference scores appear to be greater for subjects experiencing current depression than for those with a history of depression only. Because we did not conduct debriefing interviews with subjects, explanations for this observation are unclear. However, based on theoretical considerations, there could be a role for coping and appraisal related to current depression severity whereby depressed patients utilize more emotion-focused and less problem-focused coping strategies than subjects with a history of depression only when considering the preference scores for hypothetical depression health states (Lazarus and Folkman 1984; Matheson and Anisman 2003). Future research is needed to better understand the depression health state preference differences between currently depressed subjects and the other groups (general population and history of depression only).

More significant differences between the preference scores of the general population and individuals with depression were noted when using the

rating scale versus standard gamble method (8 significant differences versus 4, respectively). There remains considerable debate about which preference score method is the gold standard (Gold et al. 1996; Green, Brazier, and Deverill 2000; Sherbourne et al. 2001). A concern raised about depressed patients assigning preference scores to health states is that suicidal ideation (a common symptom of depression) would result in depressed patients choosing death over any other outcome. Methods exist for assigning preferences to health states considered worse than dead (Macran and Kind 2001); however, we did not find evidence of this among depressed subjects in this study.

More differences in depressed patient versus general population preference scores were noted when using the PHQ-9 versus SF-12 depression health state descriptions (7 versus 5, respectively). All PHQ-9 depression health state preference scores were lower than SF-12 preference scores, especially for the mild PHQ-9 depression health state. Five of six PHQ-9 depression health state descriptions were significantly lower for depressed patients versus the general population using the rating scale preference method. The depression health state descriptions using the PHQ-9 included the DSM-IV depression symptoms and a generic description of functional impairment associated with work, home, and relationships, whereas the SF-12 descriptions are based on a more generic measure of functioning. Therefore, the PHQ-9 descriptions were more depression specific than the SF-12, and it appears that depressed patients assigned lower preference scores to the more depression-specific health state descriptions.

Overall, there are two implications of these findings. First, the use of general population preference scores for depression interventions could result in less favorable cost-effectiveness ratios compared with using depressed patient preference scores because preference score differences between depressed health states and perfect health are smaller for the general population than depressed patients. A cost-effectiveness ratio with a smaller denominator would result in a larger (less favorable) ratio. For example, a patient with moderate depression restored to perfect health would result in an SF-12 rating scale preference score change of 0.37 using the more severely depressed patient preference scores and 0.28 using the general population preference scores. Similarly, the same patient would have an SF-12 standard gamble preference score change of 0.30 using the more severely depressed patient preference scores and 0.23 using the general population preference scores. It is important to note that these potential change scores, 0.09 and 0.07, both exceed the minimally important clinical difference (approximately 0.03) reported for preference scores (Walters and Brazier 2003; Kaplan 2005). Thus,

many of the differences in change score estimates based on depressed patient versus the general population preferences were both clinically and statistically significant. If these preference score differences were part of an incremental cost per QALY analysis, the ratio would be approximately 30 percent greater using general population versus depressed patient preference scores. Preference weights for the SF-12 or PHQ-9 that are derived from depressed patients are not available at this time; therefore, we were not able to reanalyze existing cost-effectiveness analysis datasets using depressed patient preference weights. Second, these findings may contribute to our understanding of the observation that mental health treatment resources are not keeping pace with physical health treatment resources (Beck et al. 2003; Schomerus, Matschinger, and Angermeyer 2006). If the general population underestimates the impact of depression, then there may be less motivation to invest health care resources for depression treatment (McKie and Richardson 2003).

This study had several limitations. Fewer standard gamble comparisons were statistically significant between depressed patients and the general population compared with rating scale comparisons. This is important because the standard gamble tends to be the preferred preference elicitation method among at least some health care economists. However, the connection between the standard gamble preference scores and how patients make health care decisions has been the subject of debate, and rating scale methods have been widely accepted as the most practical of the preference elicitation methods (Brazier et al. 1999). In addition, more statistically significant differences between depressed patients and the general population were noted for the depression-specific health states (PHQ-9) than the generic health states (SF-12). This is important because the Panel on Cost Effectiveness in Health and Medicine recommends the use of generic measures (Gold et al. 1996). However, there are several high-profile economic evaluations of depression interventions that converted depression-specific symptom severity into generic QALYs (Schoenbaum et al. 2001; Simon et al. 2001; Katon et al. 2005), and there is some evidence to support the validity of these conversion formulas (Pyne et al. 2007). Subjects were recruited from a convenience sample and from a single state and therefore may not be representative of the universe of depressed patients or the general population.

In conclusion, depressed patients report lower depression health state preference scores than the general population. Given this finding, cost-effectiveness ratios using general population preference scores may result in less favorable cost-effectiveness ratios compared with ratios using depressed patient preferences. At the very least, we recommend replication of our findings

and consideration of depressed patient preference scores to calculate QALYs in sensitivity analyses for cost-effectiveness analyses of depression interventions.

ACKNOWLEDGMENT

Joint Acknowledgment/Disclosure Statement: This work was supported by NIMH R21 MH64681-01A1 and Veterans Affairs Research Career Development Award (Dr. Pyne). The authors acknowledge Christian Lynch and Silas Williams for their tremendous data collection efforts.

Disclosures: None.

Disclaimers: None.

REFERENCES

Adang, E. M., G. Kootstra, G. L. Engel, J. P. van Hooff, and H. L. Merckelbach. 1998. "Do Retrospective and Prospective Quality of Life Assessments Differ for Pancreas–Kidney Transplant Recipients?" *Transplant International* 11 (1): 11–5.

Alloy, L. B., and L. Y. Abramson. 1979. "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?" *Journal of Experimental Psychology: General* 108 (4): 441–85.

Badia, X., A. Diaz-Prieto, M. Rue, and D. L. Patrick. 1996. "Measuring Health and Health State Preferences among Critically Ill Patients." *Intensive Care Medicine* 22 (12): 1379–84.

Balaban, D. J., P. C. Sagi, N. I. Goldfarb, and S. Nettler. 1986. "Weights for Scoring the Quality of Well-being Instrument among Rheumatic Arthritics." *Medical Care* 24 (11): 973–80.

Barney, L. J., K. M. Griffiths, A. F. Jorm, and H. Christensen. 2006. "Stigma about Depression and Its Impact on Help-Seeking Intentions." *Australian and New Zealand Journal of Psychiatry* 40 (1): 51–4.

Beck, M., S. Dietrich, H. Matschinger, and M. C. Angermeyer. 2003. "Alcoholism: Low Standing with the Public? Attitudes towards Spending Financial Resources on Medical Care and Research on Alcoholism." *Alcohol and Alcoholism* 38 (6): 602–5.

Boyd, N., H. Sutherland, K. Heasman, D. L. Tritchler, and B. J. Cummings. 1990. "Whose Utilities for Decision Analysis?" *Medical Decision Making* 10 (1): 58–67.

Brauer, C. A., A. B. Rosen, D. Greenberg, and P. J. Neumann. 2006. "Trends in the Measurement of Health Utilities in Published Cost–Utility Analyses." *Value in Health* 9 (4): 213–18.

Brazier, J., R. Akehurst, A. Brennan, P. Dolan, K. Claxton, C. McCabe, M. Sculpher, and A. Tsuchyia. 2005. "Should Patients Have a Greater Role in Valuing Health States?" *Applied Health Economics and Health Policy* 4 (4): 201–8.

Brazier, J., M. Deverill, C. Green, R. Harper, and A. Booth. 1999. "A Review of the Use of Health Status Measures in Economic Evaluation." *Health Technology Assessment* 3 (9): i–iv, 1–164.

Brazier, J., J. Roberts, and M. Deverill. 2002. "The Estimation of a Preference-Based Measure of Health from The SF-36." *Journal of Health Economics* 21 (2): 271–92.

Brazier, J., T. Usherwood, R. Harper, and K. Thomas. 1998. "Deriving a Preference-Based Single Index from the UK SF-36 Health Survey." *Journal of Clinical Epidemiology* 51 (11): 1115–28.

De Wit, G. A., J. J. Busschbach, and F. T. De Charro. 2000. "Sensitivity and Perspective in the Valuation of Health Status: Whose Values Count?" *Health Economics* 9 (2): 109–26.

Dolders, M. G., M. P. Zeegers, W. Groot, and A. Ament. 2006. "A Meta-Analysis Demonstrates No Significant Differences between Patient and Population Preferences." *Journal of Clinical Epidemiology* 59 (7): 653–64.

Drummond, M. F., B. O'Brien, G. L. Stoddart, and G. W. Torrance. 1997. *Methods for the Economic Evaluation of Health Care Programmes.* 2nd Edition. New York: Oxford University Press.

Fleiss, J. L. 1986. *The Design and Analysis of Clinical Experiments.* New York: John Wiley & Sons.

Froberg, D. G., and R. L. Kane. 1989a. "Methodology for Measuring Health-State Preferences—II: Scaling Methods." *Journal of Clinical Epidemiology* 42 (5): 459–71.

———. 1989b. "Methodology for Measuring Health-State Preferences III: Population and Context Effects." *Journal of Clinical Epidemiology* 42 (6): 585–92.

Furlong, W., D. Feeny, G. W. Torrance, R. Barr, and J. Horsman. 1990. *Guide to Design and Development of Health-State Utility Instrumentation.* Hamilton, ON, Canada: McMaster University.

Gabriel, S. E., T. S. Kneeland, L. J. Melton III, M. M. Moncur, B. Ettinger, and A. N. Tosteson. 1999. "Health-Related Quality of Life in Economic Evaluations for Osteoporosis: Whose Values Should We Use?" *Medical Decision Making* 19 (2): 141–48.

Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein. 1996. *Cost-Effectiveness in Health and Medicine.* New York: Oxford University Press Inc.

Green, C., J. Brazier, and M. Deverill. 2000. "Valuing Health-Related Quality of Life. A Review of Health State Valuation Techniques." *PharmacoEconomics* 17 (2): 151–65.

Insinga, R. P., and D. G. Fryback. 2003. "Understanding Differences between Self-Ratings and Population Ratings for Health in the EuroQOL." *Quality of Life Research* 12 (6): 611–9.

Kaplan, R. M. 2005. "The Minimally Clinically Important Difference in Generic Utility-Based Measures." *Chronic Obstructive Pulmonary Disease* 2 (1): 91–7.

Katon, W. J., M. Schoenbaum, M. Y. Fan, C. M. Callahan, J. Williams Jr., E. Hunkeler, L. Harpole, X. H. Zhou, C. Langston, and J. Unutzer. 2005. "Cost-Effectiveness of Improving Primary Care Treatment of Late-Life Depression." *Archives of General Psychiatry* 62 (12): 1313–20.

Kroenke, K., R. L. Spitzer, and J. B. Williams. 2001. "The PHQ-9 Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine* 16 (9): 606–13.

Lazarus, R. S., and S. Folkman. 1984. *Stress, Appraisal, and Coping.* New York: Springer.
Lenert, L. A., J. R. Treadwell, and C. E. Schwartz. 1999. "Associations between Health Status and Utilities Implications for Policy." *Medical Care* 37 (5): 479–89.
Link, B. G., J. C. Phelan, M. Bresnahan, A. Stueve, and B. A. Pescosolido. 1999. "Public Conceptions of Mental Illness: Labels, Causes, Dangerousness, and Social Distance." *American Journal of Public Health* 89 (9): 1328–33.
Llewellyn-Thomas, H., H. J. Sutherland, R. Tibshirani, A. Ciampi, J. E. Till, and N. F. Boyd. 1984. "Describing Health States: Methodologic Issues in Obtaining Values for Health States." *Medical Care* 22: 543–52.
Macran, S., and P. Kind. 2001. " 'Death' and the Valuation of Health-Related Quality of Life." *Medical Care* 39 (3): 217–27.
Matheson, K., and H. Anisman. 2003. "Systems of Coping Associated with Dysphoria, Anxiety and Depressive Illness: A Multivariate Profile Perspective." *Stress* 6 (3): 223–34.
McKie, J., and J. Richardson. 2003. "The Rule of Rescue." *Social Science and Medicine* 56 (12): 2407–19.
National Institute for Clinical Excellence. 2004. "Guide to the Methods of Technology Appraisal" [accessed on January 2, 2008]. Available at http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf
Perry, B. L., B. A. Pescosolido, J. K. Martin, J. D. McLeod, and P. S. Jensen. 2007. "Comparison of Public Attributions, Attitudes, and Stigma in Regard to Depression among Children and Adults." *Psychiatric Services* 58 (5): 632–5.
Postulart, D., and E. M. M. Adang. 2000. "Response Shift and Adaption in Chronically Ill Patients." *Medical Decision Making* 20: 186–93.
Pyne, J. M., S. Tripathi, D. K. Williams, and J. Fortney. 2007. "Depression-Free Day to Utility-Weighted Score: Is It Valid?" *Medical Care* 45 (4): 357–62.
Rashidi, A. A., A. H. Anis, and C. A. Marra. 2006. "Do Visual Analogue Scale (VAS) Derived Standard Gamble (SG) Utilities Agree with Health Utilities Index Utilities? A Comparison of Patient and Community Preferences for Health Status in Rheumatoid Arthritis Patients." *Health and Quality of Life Outcomes* 4: 25.
Revicki, D. A., A. Shakespeare, and P. Kind. 1996. "Preferences for Schizophrenia-Related Health States: A Comparison of Patients, Caregivers and Psychiatrists." *International Clinical Psychopharmacology* 11 (2): 101–8.
Sackett, D. L., and G. W. Torrance. 1978. "The Utility of Different Health States as Perceived by the General Public." *Journal of Chronic Diseases* 31: 697–704.
Schoenbaum, M., J. Unutzer, C. Sherbourne, N. Duan, L. V. Rubenstein, J. Miranda, L. S. Meredith, M. F. Carney, and K. Wells. 2001. "Cost-Effectiveness of Practice-Initiated Quality Improvement for Depression: Results of a Randomized Controlled Trial." *Journal of the American Medical Association* 286 (11): 1325–30.
Schomerus, G., H. Matschinger, and M. C. Angermeyer. 2006. "Preferences of the Public Regarding Cutbacks in Expenditure for Patient Care: Are There Indications of Discrimination against Those with Mental Disorders?" *Social Psychiatry and Psychiatric Epidemiology* 41 (5): 369–77.
Seligman, M. E. P. 1998. *Learned Optimism: How to Change Your Mind and Your Life.* New York: A.A. Knopf.

Sherbourne, C. D., J. Unutzer, M. M. Schoenbaum, N. Duan, L. A. Lenert, R. Sturm, and K. B. Wells. 2001. "Can Utility-Weighted Health-Related Quality-of-Life Estimates Capture Health Effects of Quality Improvement for Depression?" *Medical Care* 39 (11): 1246–59.

Simon, G. E., W. J. Katon, M. VonKorff, J. Unutzer, E. H. Lin, E. A. Walker, T. Bush, C. Rutter, and E. Ludman. 2001. "Cost-Effectiveness of a Collaborative Care Program for Primary Care Patients with Persistent Depression." *American Journal of Psychiatry* 158 (10): 1638–44.

Smith, D. M., R. L. Sherriff, L. Damschroder, G. Loewenstein, and P. A. Ubel. 2006. "Misremembering Colostomies? Former Patients Give Lower Utility Ratings Than Do Current Patients." *Health Psychology* 25 (6): 688–95.

Sugar, C. A., R. Sturm, T. T. Lee, C. D. Sherbourne, R. A. Loshen, K. B. Wells, and L. A. Lenert. 1998. "Empirically Defined Health States for Depression from the SF-12." *Health Services Research* 33 (4): 911–28.

Tsevat, J., N. V. Dawson, A. W. Wu, J. Lynn, J. R. Soukup, E. F. Cook, H. Vidaillet, and R. S. Phillips. 1998. "Health Values of Hospitalized Patients 80 Years or Older. HELP Investigators. Hospitalized Elderly Longitudinal Project." *Journal of the American Medical Association* 279 (5): 371–5.

Ubel, P. A., G. Loewenstein, and C. Jepson. 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.

Ubel, P. A., G. Loewenstein, N. Schwarz, and D. Smith. 2005. "Misimagining the Unimaginable: The Disability Paradox and Health Care Decision Making." *Health Psychology* 24 (4, suppl.): S57–62.

Walters, S. J., and J. E. Brazier. 2003. "What Is the Relationship between the Minimally Important Difference and Health State Utility Values? The Case of the SF-6D." *Health and Quality of Life Outcomes* 1: 4.

Ware, J. Jr., M. Kosinski, and S. D. Keller. 1996. "A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity." *Medical Care* 34 (3): 220–33.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: PHQ-9 and SF-12 Depression Health State Descriptions.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.