2007

Estimating health utilities and quality adjusted life years in seasonal affective disorder research

Michael C. Freed
American University

Kelly J. Rohan
Uniformed Services University of the Health Sciences

Brian T. Yates
American University

Research report

Estimating health utilities and quality adjusted life years in seasonal affective disorder research

Michael C. Freed [a,*], Kelly J. Rohan [b,1], Brian T. Yates [a]

[a] *Department of Psychology, American University, Washington, DC, United States*
[b] *Department of Medical and Clinical Psychology, Uniformed Services University of the Health Sciences, Bethesda, MD, United States*

Abstract

Background: Seasonal affective disorder (SAD) episodes will recur annually without effective intervention. Effectiveness of such interventions is traditionally measured with depression-specific tools (e.g., Beck Depression Inventory 2nd Edition; BDI-II). In a climate of potentially scarce resources, generic outcomes, such as Quality Adjusted Life Years (QALYs), are recommended for cost-effectiveness research. For treatments to be deemed cost-effective, they must show effectiveness relative to each other and relative to interventions across other disorders. To date, QALYs have not been used to determine effectiveness of SAD treatments. Given the recurrent nature of SAD, QALYs, which weight quality of life with time, are an ideal SAD treatment outcome.

Method: A method to assess QALYs for SAD was developed using pilot clinical trial data. The method estimated health utilities, a measure of quality of life for a QALY, by anchoring pilot BDI-II data from the SAD clinical trial with previously derived health utilities for nonseasonal depression.

Results: Relative to no treatment, median QALYs gained ranged from 0.11–0.18 over 1 year, depending on the intervention assessed. Discussion: Any treatment for SAD must compete with spontaneous spring remission, as illness severity attenuates in the spring.

Limitations: Health utilities were estimated from the depression literature, and potential side effects from SAD treatments were not included in the estimates. The clinical trial time horizon was limited to 1-year.

Conclusions: The proposed method offers researchers a tool to transform SAD efficacy data into a generic outcome for use in cost-effectiveness analysis of SAD treatments.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Quality adjusted life year; Cost-effectiveness; Depression; Seasonal affective disorder; Health utility

1. Introduction

1.1. Seasonal affective disorder and quality adjusted life years

Seasonal affective disorder (SAD) is defined by DSM-IV-TR (APA 2000) as a recurrent pattern of major depressive episodes with a temporal relationship between the episodes and time of year. SAD affects up to 9.7% of the population (Birtwistle and Martin, 1999; Rosen et al.,

1990) and prevalence increases with latitude in the United States (Mersch et al., 1999). Left untreated, SAD episodes recur annually (Sakamoto et al., 1995). Although light therapy (LT; Terman et al., 1989) and group cognitive behavioral therapy (CBT; Rohan et al., 2004) improve SAD symptoms, the outcomes used to demonstrate improvement are depression-specific. That is, measures such as the Beck Depression Inventory, 2nd Edition (BDI-II; Beck et al., 1996) can be used only to compare the effectiveness of interventions for depression. In a climate of potentially scarce resources, effective interventions must also be cost-effective (Yates, 1994). A more generic outcome is needed to compare the cost-effectiveness of depression-related interventions against the effectiveness of interventions for other disorders (Pyne et al., 2003).

The U.S. Public Health Service panel on cost-effectiveness recommended the use of quality adjusted life years (QALYs) as that generic measure of effectiveness to inform healthcare resource allocation decisions (Gold et al., 1996). A QALY measures the health impact of an intervention by capturing quality (morbidity) and length of life (mortality) in a single dimension (Torrance and Feeny, 1989; Weinstein and Stason, 1977). Quality of life is weighted by time such that a year in perfect health is worth 1 QALY and death is worth 0. QALYs can be used to assess treatment outcomes in terms of absolute outcomes, without consideration of pre-treatment health. Alternatively, QALYs can be used to assess what is gained (the incremental effectiveness) over and above no treatment (the so-called, do-nothing approach) or over and above an existing treatment. For example, if the QALY for a usual care intervention is 0.3 and the QALY for an enhanced care intervention is 0.9, then the difference between usual care and enhanced care, 0.6, is the QALY gained (QALYG). Thus, 0.6 QALYs are gained from implementing enhanced care over usual care.

In a review of the cost-effectiveness literature for nonseasonal depression treatment, Barrett et al. (2005) reported that only 12 of the 58 articles reviewed (21%) assessed cost per QALY, a cost-utility analysis. The recurrent nature of SAD makes QALYs an attractive treatment outcome. Unlike other chronic conditions, SAD remits in the spring and summer (DSM-IV-TR, 2000) for many sufferers, and thus, even without any treatment, the change of seasons attenuates SAD symptoms. QALYs weight quality of life with time, and therefore can account for seasonal changes in symptom severity. SAD treatments must compete with these seasonal changes. However, no literature to date reported QALYs for SAD.

1.2. Estimating health utilities

The quality of life component of a QALY can be measured by units called health utilities, although other similar units (generically called preference scores) of quality of life exist. Health utilities are ratings of outcome desirability in the face of uncertainty, where outcomes are specific health conditions, like severe depression (Drummond et al., 1997; Torrance and Feeny, 1989). Multiple health conditions (e.g., mild depression and severe depression) are then rated relative to each other and anchored by perfect health (equal to 1) and death (equal to 0). Utilities are elicited through a time consuming method called the standard gamble (SG; Drummond et al., 1997). Because administering the SG is not always feasible, researchers have relied on previously published utilities. For example, Lave et al. (1998) and Katon et al. (2005) used the average of six published depression-related utilities scores. Participants considered depressed (via the researchers' depression outcome) were assigned a utility of 0.51 and participants considered not depressed were assigned a utility of 1. Raw scores from the depression-specific measure were linearly interpolated to reflect utilities between 0.51 and 1.0. However, the utilities that Lave et al. (1998) and Katon et al. (2005) averaged were not all elicited from patients with depression. Utilities derived from patients (vs. providers or nonpatients) are preferred (Hlatk, 1995; Montgomery and Fahey, 2001).

For clinical untreated depression, researchers have elicited utilities as low as 0.3 from participants diagnosed with depression (Revicki and Wood, 1998). That is, participants were indifferent between the following two choices: choice 1) being severely depressed and choice 2) receiving a depression-curing intervention with the probability of success being 0.3 and death 0.7. The probability of success at the point of indifference is the health utility. Utilities for conditions such as treated remitted depression with no pharmacological treatment are 0.895 (Revicki et al., 1995; Revicki et al., 1997) and 0.86 (Revicki and Wood, 1998). Revicki and Wood (1998) also elicited utilities from depressed participants for mild depression (0.64–0.73) and moderate depression (0.55–0.63). However, the mild and moderate depression conditions included the experience of medication side effects along with symptoms of depression.

1.3. Study aims

We present a method to capture QALYs and QALYGs for patients with SAD using Revicki and Wood's (1998)

patient-elicited utility anchors and raw scores from the BDI-II (Beck et al., 1996). We demonstrate our method using pilot data from a clinical trial examining treatments for SAD. As described earlier, a paucity of research exists on QALYs for depression, and no research exists on QALYs for SAD.

2. Method

Data from 20 participants residing in the greater Washington DC area who had enrolled in and completed a pilot investigation examining the efficacy of a novel SAD-tailored cognitive–behavioral therapy (CBT) for patients with SAD were used. Participants entered treatment once they met Structured Interview Guide for the Hamilton Depression Rating Scale–Seasonal Affective Disorder Version (SIGH–SAD; Williams et al., 1992) criteria for a SAD episode. In a 3-condition design with 3 repeated measures, participants were randomized to 1) CBT alone (12, group sessions each 1.5 h long held twice weekly for 6 weeks), 2) Light Therapy alone (LT; 10,000-lux standard light box viewed in participants' homes in 45-minute doses, twice daily for 6 weeks), or 3) CBT+LT (all elements of CBT and LT concurrently). Each participant completed the SIGH-SAD and BDI-II at pre-treatment, immediate post-treatment, and one-year following the end of treatment. The two outcome measures used in study were the SIGH-SAD (Williams et al., 1992) and the BDI-II (Beck et al., 1996.) Both measures have been used extensively in depression research and have adequate reliability and validity.

Institutional Review Boards of the authors' affiliated universities approved the current project. The pilot data showing the efficacy of CBT for SAD can be found in Rohan et al. (2004). To summarize, Rohan et al. found no difference among the conditions at post-treatment,

but BDI-II scores were significantly more favorable for participants in CBT and CBT+LT, when compared to LT at the one-year winter follow up.

3. Results

3.1. Health utility estimation

For the purposes of estimating health utilities, a BDI-II conversion method was developed with consultation from Dr. Scott Parker (personal communication, 2004). In order to convert BDI-II scores to the estimated health utilities, BDI-II scores of 63 (indicating maximal severe depression) were matched to the published health utility for untreated depression (0.3; Revicki and Wood, 1998). Then, BDI-II scores of 0 (indicating no depression) were matched to the published health utility for depression in remission, not in treatment (0.9; Revicki et al., 1995; Revicki and Wood, 1998; rounded to the nearest 10th). Between these published anchors, equally spaced estimated health utility ranges were determined for each level of depression severity as demarcated by BDI-II scores. (Of note, Beck et al., 1996, derived optimal cut scores using receiver operating characteristic (ROC) curves to determine ordinal levels of depression severity.) This was accomplished by dividing the difference between the maximum health utility of 0.9 and minimum utility of 0.3 (difference equal to 0.6) by 5, the number of severity ranges on the BDI-II: remission, sub-clinical, mild, moderate, and severe. Therefore, range of each severity level was determined to be 0.12 health utility units. The conversion method and the resulting transformation tool can be seen in Table 1 and Fig. 1. The estimated health utilities were assigned to each participant at pre-treatment, post-treatment, and 1-year follow-up. The estimated health utilities for nonseasonal depression should be comparable to SAD because

Table 1
Health utility estimation method

Severity	BDI-II	Rationale for Health State Rating	Utility
Remission*	<9	Revicki et al. (1995) and Revicki et al. (1997) used utilities of 0.895 and 0.89, respectively. Revicki and Wood (1998) elicited utilities from patients of 0.86.	0.78 to 0.9
Sub-clinical	9–13	Not an official BDI-II category; however, participants were still symptomatic. These utilities were estimated.	0.66 to <0.78
Mild*	14–19	No published utilities exist for this severity. Health utilities were therefore estimated.**	0.54 to <0.66
Moderate*	20–28	No published utilities exist for this severity. Health utilities were therefore estimated.**	0.42 to <0.54
Severe*	29–63	Revicki et al. (1995) and Revicki et al. (1997) used utilities of .31 and .306, respectively, for untreated depression. Revicki and Wood (1998) elicited 0.3 for untreated severe depression.	0.3 to <0.42

*BDI-II ranges from Beck et al. (1996).
**Revicki and Wood (1998) elicited utilities ranging from 0.64–0.73 for mild depression and 0.55–0.63 for moderate depression. However, their health state descriptions included side effects from pharmacological interventions.
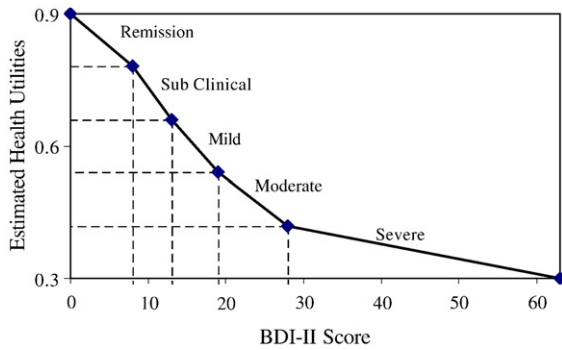
Fig. 1. BDI-II scores to health utilities transformation tool.

SAD is regarded as a seasonal subtype of Major Depression (DSM-IV-TR, 2000).

3.2. QALYs and QALYGs in SAD

To establish QALYs and QALYGs for patients with SAD, the estimated utilities were weighted by time and then summed over 1 year. This time weighting process helped account for the fact that individuals with SAD do not experience equal symptom severity throughout the entire year. By definition, individuals with SAD experience the most severe symptoms during the three winter months, with milder symptom profiles occurring during fall and spring months. In a given year, a single health utility cannot be representative of overall health status.

In consultation with a recognized SAD expert (Dr. Kelly Rohan, personal communication, 2004), a method was created to best fit the pre-treatment, post-treatment,

and 1-year follow-up data into a one-year course of SAD (Table 2). Pre-treatment BDI-II scores were used to estimate health utilities for the 1.5 months (i.e., 6 weeks) of treatment (January 1st to mid-February). Assuming that SAD patients were symptomatic until mid-May, post-treatment scores were used to estimate health utilities for next 3 months. The expected spontaneous and complete remission period from mid-May through the end of September was assigned the most favorable utility 0.9). Because SAD symptoms begin in the early fall, 1-year follow-up scores were used to estimate utilities for the remaining 3 months of the year (i.e., from early October through December). Thus, a QALY over 1 year was calculated by weighting time with a) the median pre-treatment estimated utilities (pre), b) post-treatment estimated utilities (post), c) one-year follow-up estimated utilities (1-year) and d) spring and summer (spr sum) estimated remission. The following formula was used:

$$\frac{[(\text{pre} \times 1.5\,\text{months}) + (\text{post} \times 3\,\text{months}) + (0.9\,\text{spr sum} \times 4.5\,\text{months}) + (1 - \text{year} \times 3\,\text{months})]}{12\,\text{months}}$$

(1)

The calculations assumed no change in symptom severity across the 6-week treatment phase and that any residual SAD symptoms following treatment persisted through mid-May. Also, the one-year follow-up data back estimated SAD severity during the preceding months of October through December, months where symptoms are presumably less severe. And, the one-year assessment occurred when SAD symptoms are the most severe. Thus, the three month period using this one-year data likely estimates symptom severity to be overly severe. The natural course of SAD symptom

Table 2
QALY and QALYG calculation method

QALY method		
Timeframe	**Number of months**	**Data used**
Early January to mid-February	1.5	Pre-treatment
Mid-February to mid-May	3	Post-treatment
Mid-May to end of September	4.5	Spontaneous remission [a]
Beginning of October to end of December	3	1-year follow-up
QALYG method		
Step 1: Calculate no-treatment QALY		
Timeframe	**Number of months**	**Data used**
Early January to mid-February; beginning of October to end of December	7.5	Pre-treatment scores
Mid-May to end of September	4.5	Spontaneous remission [b]
Step 2: Subtract no-treatment QALY from QALY		

[a] Assumed all participants spontaneously remit during summer.
[b] Assumed all participants spontaneously remit during summer. Because there is no change during the summer months, nothing is gained from receiving treatment for SAD during these months.
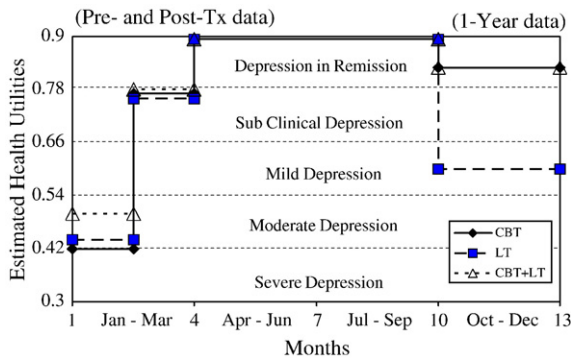
Fig. 2. Median SAD health utilities over 1-year.

onset, escalation, and offset is obviously more gradual than is reflected in the formula.

Fig. 2 graphically displays the median estimated utilities from the pilot dataset (Rohan et al., 2004). Here, a QALY is represented as the area under the curve for a particular treatment, and in this example, is defined as the average monthly utility. Time is an important dimension for a QALY, as theoretical remission during the spring and summer lessens the overall severity captured in a single QALY score. Median, mean, and parenthesized standard deviations are as follows: CBT=0.79, 0.78, (0.03); LT=0.73, 0.74, (0.07); and CBT+LT=0.80, 0.81, (0.05). The Kruskal–Wallis test (Mehta and Patel, 2000) showed no difference in QALYs among the treatment conditions, $p=$ns.

To assess QALYGs, pre-treatment utilities were substituted for post-treatment and 1-year utilities, in the formula above, resulting in a no-treatment QALY. The no-treatment QALY was subtracted from the treatment QALY, resulting in a QALYG. For example, the median QALYG for CBT is 0.18. To calculate this no-treatment QALY, the pre-treatment CBT health utility was 0.42 (seen in Fig. 2) to represent quality of life in the symptomatic months. The median no-treatment CBT QALY, 0.60, was subtracted from the QALY for CBT, 0.78, resulting in 0.18 QALYG. QALYGs greater than 0.12 indicated categorical improvement from a severity level (e.g., moderate to mild depression). Median, mean, and parenthesized standard deviations are as follows: CBT=0.18, 0.18, (0.06); LT=0.11, 0.11, (0.06); and CBT+LT=0.14, 0.14, (0.05). The Kruskal–Wallis test (Mehta and Patel) showed no difference in QALYGs among the treatment conditions, $p=$ns.

4. Discussion

We developed a method to measure a QALY and QALYG for individuals suffering from a subtype of depression: SAD. This method estimated health utilities from BDI-II scores using published depression-related utilities as anchors. The estimated utilities were weighted by time and summed. The resulting outcomes, QALYs and QALYsG, can be used to assess both clinical and statistical significance from clinical trial data. That is, QALYs can be analyzed statistically, but they also have qualitative meaning. For example, a QALY of 0.9 would indicate depression in remission. Similarly, a QALYG of greater than 0.12 would indicate improvement from one BDI-II-determined severity level to another. Using QALYs allows for comparison among treatments for a single disorder or for comparison among treatments of differing disorders (Drummond et al., 1997). Thus, policy and decision makers now have a common metric, albeit constructed with pilot data, on which to compare treatments for SAD relative to each other or relative to treatments for other disorders (e.g., CBT for depression relative to CBT for anxiety). Use of the proposed transformation tool gives clinical researchers a method for generalizing disorder-specific treatment outcomes to a wider audience. Particularly, this wider audience may include health economists and epidemiologists, who are involved in resource allocation.

We recognize several limitations of the current paper. First, we did not administer the standard gamble (SG) to our participants. We assumed that the phenomenological experience of being depressed from nonseasonal depression would parallel an episode of SAD. And, we assumed that the severity categories of the BDI-II would match the severity descriptions used by the researchers from whom we estimated our health utilities. It is not known whether our estimated utilities preserved the preferences inherent in the original utilities elicited from patients diagnosed with depression. Second, we chose to apply health utilities developed from researchers who administered the SG to patients, even though other methods existed. For example, Hatziandreu et al. (1994) elicited preferences from a physician panel rather than from actual patients.

Third, we used data from a pilot clinical trial (Rohan et al., 2004), which was underpowered and not designed to inform researchers about the course of SAD. The lack of significant differences among treatments using the QALY(G) scores may have reflected lack of power and not lack of treatment differences, as Rohan et al., reported significant treatment group differences at her 1-year follow-up. The limited number of symptomatic months for SAD patients reduced the overall QALY(G) severity, thereby masking gains attributable to treatment. Moreover, participants were assessed three times. Symptoms of

M.C. Freed et al. / Journal of Affective Disorders 100 (2007) 83–89

SAD may fluctuate between assessments (e.g., assumed spontaneous remission during spring and summer months), thereby impacting our calculations of QALY (G). Fourth, QALYs are often (but not always) modeled using Bayesian data analytic procedures (O'Hagan and Luce, 2003), where parameters of the model are varied in sensitivity analyses to examine possible changes on outcomes. For example, a utility for severe depression could range from 0.3 (Revicki and Wood, 1998) to 0.51 (Lave et al., 1998), thus impacting the overall QALY or QALYG. Fifth, our estimate of a do-nothing approach used within-group pre-treatment scores. Participants were deliberately not enrolled in treatment until they met criteria for a SAD episode. The pre-treatment scores reflected the most severe participant scores. It is possible that participants would experience some improvement even without treatment, thus reducing the difference between an actual do-nothing control and presumed improvement from treatment. And finally, we looked at QALY(G)s over 1 year, rather than over a participant's lifetime, as is the typical method for assessing QALYs. Despite these limitations, incorporating QALYs into treatment-outcome research for SAD moves us one step closer to completing the research agenda argued for by Yates (1994): the assessment of both costs and outcomes to allow for more complete theoretical and empirical models of mental health treatment.

Acknowledgements

References

American Psychiatric Association (APA), 2000. Quick Reference to the Diagnostic Criteria from DSM-IV-TR. Author: Washington, DC.

Barrett, B., Byford, S., Knapp, M., 2005. Evidence of cost-effective treatments for depression: a systematic review. J. Affect. Disord. 84 (1), 1–13.

Beck, A.T., Steer, R.A., Brown, G.K., 1996. Beck Depression Inventory®—Second Edition Manual. The Psychological Corporation, Harcourt, Brace and Co, United States.

Birtwistle, J., Martin, N., 1999. Seasonal affective disorder: its recognition and treatment. Br. J. Nurs. 8 (15), 1004–1009.

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

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

Hatziandreu, E., Brown, R.E., Revicki, D.A., Turner, R., Martindale, J., Levine, S., Siegel, J.E., 1994. Cost utility of maintenance treatment of recurrent depression with sertraline versus episodic treatment with dothiepin. Pharmacoeconomics 5, 249–268.

Hlatk, M., 1995. Patient preferences and clinical guidelines. JAMA 273, 1219–1220.

Katon, W.J., Schoenbaum, M., Fan, M., Callahan, C.M., Williams, J., Hunkeler, E., et al., 2005. Cost-effectiveness of improving primary care treatment of late-life depression. Arch. Gen. Psychiatry 62, 1313–1320.

Lave, J.R., Frank, R.G., Schulberg, H.C., Kamlet, M.S., 1998. Cost-effectiveness of treatments for major depression in primary care practice. Arch. Gen. Psychiatry 55, 645–651.

Mehta, C., Patel, N., 2000. StatXact 4 for Windows User Manual. CYTEL Software Corp., MA.

Mersch, P.P.A., Middendorp, H.M., Bouhuys, A.L., Beersma, D.G.M., van den Hoofdakker, R.H., 1999. Seasonal affective disorder and latitude: a review of the literature. J. Affect. Disord. 53 (1), 35–48.

Montgomery, A.A., Fahey, T., 2001. How do patients' treatment preferences compare with those of clinicians? Qual. Saf. Health Care 10, i39–i43.

O'Hagan, A., Luce, B.R., 2003. A Primer on Bayesian Statistics in Health Economic and Outcomes Research. MEDTAP International Inc., Bethesda, MD.

Pyne, J.M., Sieber, W.J., David, K., Kaplan, R.M., Rapaport, M.H., Williams, D.K., 2003. Use of the quality of well-being self-administered version (QWB-SA) in assessing health-related quality of life in depressed patients. J. Affect. Disord. 76, 237–247.

Revicki, D.A., Wood, M., 1998. Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications. J. Affect. Disord. 48, 25–36.

Revicki, D.A., Brown, R.E., Palmer, W., 1995. Modeling the cost effectiveness of antidepressant treatment in primary care. Pharmacoeconomics 8, 524–540.

Revicki, D.A., Brown, R.E., Keller, M.B., 1997. Cost-effectiveness of newer antidepressants compared with tricyclic antidepressants in managed care settings. J. Clin. Psychiatry 58, 47–58.

Rohan, K.J., Lindsey, K.T., Roecklein, K.A., Lacy, T.J., 2004. Cognitive–behavioral therapy, light therapy, and their combination in treating seasonal affective disorder. J. Affect. Disord. 80, 273–283.

Rosen, L.N., Targum, S.D., Terman, M., Bryant, M.J., Hoffman, H., Kasper, S.F., Hamovit, J.R., Docherty, J.P., Welch, B., Rosenthal, N.E., 1990. Prevalence of seasonal affective disorder at four latitudes. Psychiatry Res. 31, 131–144.

Sakamoto, K., Nakadaira, S., Kamo, K., Kamo, T., Takahashi, K., 1995. A longitudinal follow-up study of seasonal affective disorder. Am. J. Psychiatry 152 (6), 862–868.

Terman, M., Terman, J.S., Quitkin, F., McGrath, P., Stewart, J., Rafferty, B., 1989. Light therapy for seasonal affective disorder: a review of efficacy. Neuropsychopharmacology 2, 1–22.

Torrance, G.W., Feeny, D., 1989. Utilities and quality-adjusted life years. Int. J. Technol. Assess Health Care 5, 559–575.

Weinstein, M.C., Stason, W.B., 1977. Foundations of cost-effectiveness analysis for health and medical practices. N. Engl. J. Med. 296 (13), 716–721.

Williams, J.B., Link, M.J., Rosenthal, N.E., Amira, L., Terman, M., 1992. Structured Interview Guide for the Hamilton Depression Rating Scale–Seasonal Affective Disorder Version (SIGH–SAD). New York State Psychiatric Institute, New York.

Yates, B., 1994. Toward the incorporation of costs, cost-effectiveness analysis, and cost-benefit analysis into clinical research. J. Consult. Clin. Psychol. 62 (4), 729–736.