



Published in final edited form as:

*Psychiatr Serv.* 2010 February ; 61(2): 160–163. doi:10.1176/appi.ps.61.2.160.

## Advantages of Using Estimated Depression-Free Days for Evaluating Treatment Efficacy

**Dr. Steven D. Vannoy, Ph.D., M.P.H.,**

Department of Psychiatry and Behavioral Sciences, University of Washington, 1959 N.E. Pacific St., Box 356560, BB1533, Seattle, WA 98195-6560

**Dr. Patricia Arean, Ph.D.,** and

Department of Psychiatry, University of California, San Francisco

**Dr. Jürgen Unützer, M.D., M.P.H.**

Department of Psychiatry and Behavioral Sciences, University of Washington, 1959 N.E. Pacific St., Box 356560, BB1533, Seattle, WA 98195-6560

Steven D. Vannoy: svannoy@u.washington.edu

### Abstract

**Objective**—Several common methods for measuring treatment response present a snapshot of depression symptoms. The construct of estimated depression-free days (DFDs) simultaneously captures treatment outcome and estimates the patient's experience of depression over time. The study compared this measure with traditional measures used in depression treatment research.

**Methods**—This secondary data analysis was based on data from the Improving Mood—Promoting Access to Collaborative Treatment trial, a multisite depression treatment study conducted in 18 primary care clinics in five states and representing eight health care systems. The sample of older adults (N=906) had been randomly assigned to receive collaborative care for depression. Participants were aged 60 or older and met criteria for major depressive disorder, dysthymia, or both. Exclusion criteria included severe cognitive impairment, active substance abuse, active suicidal behavior, severe mental illness, and active treatment from a psychiatrist. The Patient Health Questionnaire (PHQ-9) and the Hopkins Symptom Checklist (HSCL-20) were used as outcome measures at four assessment points (baseline, three months, six months, and 12 months). Outcomes were computed for relative change, standardized differences, the proportion of improvement in depression, and DFDs.

**Results**—Using four assessment points improved the agreement between DFDs and the course of symptom change between pre- and posttest measures.

**Conclusions**—The DFD is a valid measure for estimating treatment outcomes that reflects the course of symptom change over time. When multiple assessments were conducted between the pre- and posttest periods, DFDs incorporated additional data yet remained easily interpreted. The DFD should be considered for reporting outcomes in depression research.

Systematic, quantitative assessment of outcomes is a fundamental procedure in depression treatment research. However, the metrics most commonly used in outcome research bear little resemblance to the day-to-day experience of individuals with depression. Although there may be no methodological disadvantage to using abstract statistical constructs in evaluating treatment efficacy, the need to facilitate effectiveness research introduces a broader set of demands on treatment research. Two such demands are the facilitation of

---

**Disclosures:** The authors report no competing interests.

conducting cost-effectiveness analyses to help judge the relative value of an intervention and the ability to communicate outcomes effectively to frontline clinicians who are increasingly interested in incorporating evidence-based practices that have been substantiated through effectiveness research. In this report we illustrate the feasibility and validity of using the concept of estimated depression-free days (DFDs) as an outcome metric that is methodologically sound, easily incorporated into cost-effectiveness analyses, and inherently representative of the lived experience of patients with depression (1).

Comparing response to treatment between groups is most commonly done by transforming two assessment points into an effect size. For example, Cohen's *d* is a standardized effect size measure that indicates the differential change in symptom severity between two groups in terms of standard deviation from the mean (2). This type of effect size is efficient for comparing groups but conveys virtually no clinically relevant information. To help reconcile clinical terminology with outcome metrics, Riso and colleagues (3) established a basis for using a clinically relevant treatment response, commonly defined as a 50% reduction in symptoms between an initial assessment point and a follow-up assessment. Using treatment response (or other clinically relevant metrics, such as remission) offers the advantage of providing clinically relevant information, but this information is presented as a snapshot in time and does not reflect the actual course of change between assessment points and thus the depression-relevant experience of the patient over time. The DFD is an outcome metric that is both easily interpretable and intrinsically more accurate than methods based on simple transformations of two assessment points when multiple assessments are available. The concept of estimating DFDs from depression severity scores was initially used in analyses of a depression treatment trial by Lave and colleagues (4), and it has since been used in several trials of depression treatment (1,4–13). Converting ratings of depression severity over time into DFDs produces a construct with more direct clinical relevancy and minimal loss of precision (1,9). Furthermore, DFDs can be easily translated to quality-adjusted life years (9) to facilitate cost analyses (9,11,13–17).

In this report we present depression outcomes based on two measures of depression symptom severity—the Patient Health Questionnaire (PHQ-9) (18,19) and the Hopkins Symptom Checklist (HSCL-20), a 20-item subset of depression items from the Symptom Checklist–90 (20)—that were used in a large effectiveness trial of collaborative care for depression treatment for older adults. In doing so, we demonstrate two characteristics that make the choice of DFD as a main outcome metric compelling—namely, the clinical relevancy of DFD and the potential improvement in assessment accuracy when multiple assessment points are available.

## Methods

Data were derived from the intervention arm of the Improving Mood—Promoting Access to Collaborative Treatment (IMPACT) study (21). The IMPACT study was a multisite, randomized trial comparing a primary care–based collaborative care model with usual primary care for late-life depression. The study was conducted at seven study sites in five states (California, Indiana, North Carolina, Texas, and Washington) and represented eight health care organizations and 18 primary care clinics. Recruitment occurred between June 1999 and August 2001. Patients were followed for 24 months.

## Sample

Primary care patients aged 60 or older were recruited from 18 diverse primary care clinics. All participants signed written informed consent forms approved by the institutional review boards at the study coordinating center and all study sites. Of the 35,098 patients approached, 1,801 met eligibility requirements (major depression, dysthymia, or both),

consented to treatment, and were randomly assigned to the intervention arm of the study; 906 were randomly assigned to the IMPACT model of collaborative care. Intervention participants were selected for these analyses because, in addition to independent assessments of depression severity using the HSCL-20, they systematically completed PHQ-9 questionnaires at each clinical encounter as an integral part of their treatment (22). Exclusion criteria included severe cognitive impairment, active substance abuse, active suicidal behavior, severe mental illness, and active treatment from a psychiatrist.

### DFD estimation

DFD estimates are calculated by using linear interpolation to estimate daily depression severity across assessment points (1). In this study our standard assessment points were at baseline, three months, six months, and 12 months. Study outcome assessments were conducted with the HSCL-20 via telephone by an independent assessor blind to the study conditions. In addition, the care managers in the study used the PHQ-9 as a clinical assessment tool. The clinical assessments were conducted at each contact point with the patient. For this study we isolated results from PHQ-9s administered within 30 days of the four standardized assessment points. We used PHQ-9 data because of their clinical utility, as well as for the opportunity to look at how a larger number of assessment points, compared with the standard four points, would influence DFD estimates.

### Estimates of DFD based on HSCL-20 data

To estimate DFDs, we assigned a depression level to each day within the assessment period. Days within an assessment period in which the average HSCL-20 score was below .5 (on a scale of 0–4) were characterized as fully depression free and hence were assigned a score of 1. Days with average HSCL-20 scores above 1.7 (the mean score of depressed patients entering the trial) were characterized as fully depressed days and assigned a score of 0. For assessment periods in which the average depression score was between .5 and 1.7, linear interpolation was used to convert daily values into proportions between 0 and 1.

The composite estimate of DFDs was equal to the number of days within the assessment period multiplied by the assigned level of depression. When multiple assessment points were available, the DFD estimate was computed for each intermittent time period, and then the total DFD was computed as a weighted sum (weighted by the relative duration of each period). For example, to compute DFDs using baseline and 12 months, each point has an equal distribution, so the formula is  $DFD = 365 \times [(.5 \times \text{baseline DFD}) + (.5 \times 12\text{-month DFD})]$ . To compute DFDs with the four assessment points distributed at baseline, three months, six months, and 12 months, the formula is  $DFD = 365 \times [(.125 \times \text{baseline DFD}) + (.250 \times 3\text{-month DFD}) + (.375 \times 6\text{-month DFD}) + (.250 \times 12\text{-month DFD})]$ .

As stated above, the weight for each assessment point varies by the amount of time that point contributes to the estimate. Hence in this example, the three-month period beginning at baseline represents 25% of the total period, so baseline contributes to half of the three-month period ( $.5 \times 25\% = .125$ ). The three-month assessment point contributes to the calculation twice, once for the initial period between baseline and three months (weight = .125) and once again for the three-month period between three months and six months (weight = .125), for a total weight of .25. The six-month assessment point contributes to both the three-month interval between three and six months (weight = .125) and to the longer, six-month interval between six and 12 months (weight = .25), for a total weight of .375.

Estimates of DFDs have been reported with the use of the Hamilton Depression Rating Scale, the Beck Depression Inventory, and the HSCL-20. No standards exist for establishing scale cutoffs for the interpolation process. Our HSCL-20 thresholds for computing DFDs

were adapted from work by Simon (23), who used thresholds of .5 and 2.0 as one and zero DFDs, respectively. We used an upper cutoff point of 1.7—the mean baseline HSCL-20 score of IMPACT participants—to better reflect this sample's reported depression severity, all of whom met Structured Clinical Interview for DSM Disorders criteria for major depression or dysthymic disorder at the time of study entry.

### Estimates of DFD based on the PHQ-9

The procedure for estimating DFD from PHQ-9 results followed the same method as described above for the HSCL-20, with substitution of appropriate cutoff scores. Using cutoffs established by Kroenke and colleagues (19), we characterized as fully depression free the days within assessment periods in which the average PHQ-9 scores were below 5 (classified as no depression) and hence assigned a score of 1. Days within an assessment period in which the average score was above 14 (classified as moderate to severe depression) were characterized as fully depressed days and were assigned a score of 0. Again, linear interpolation was used to convert average scores between our upper and lower cutoff scores into proportions between 0 and 1.

## Results

Depression outcomes for participants who received collaborative care in the IMPACT study are reported in Table 1.

As discussed, one of the advantages of using the DFD measure is the ability to incorporate multiple assessment points. The increase in reported symptom change on the HSCL-20 went from 153 DFDs with two assessment points, to 197 with three, to 204 with four assessment points—resulting in an increase of 33%. Similarly for the PHQ-9, two assessment points yielded 200 DFDs, whereas four assessment points yielded 265, an increase again of 33%. Using all available PHQ-9 assessments (mean of 16, range 8–38) yielded 273 DFDs, an incremental change from four assessment points of 3%.

## Discussion

We computed DFDs by using two assessment instruments and two methods for establishing cutoff scores for determining a DFD. In both approaches, we found that incorporating multiple assessment points changed the estimated effect size of treatment by 33%. The use of the DFD measure affords researchers advantages, namely the inherent ability to take advantage of multiple assessment points to increase accuracy in representing the course of symptom response and the ease with which cost analyses can be conducted.

In this study, we did not compare DFDs between the intervention and comparison groups from the IMPACT trial. The missing step of comparing the relative difference in DFDs between groups is a straightforward analytic process, and the results of this comparison have been reported elsewhere (24,25). Instead, we examined results from the intervention arm because they afford the ability to investigate a potential ceiling effect for measurement frequency on outcome. The most accurate method of calculating DFDs would likely use daily experience sampling (with a depression diary, for example). However, daily measurement is expensive, and follow-through by patients is a major barrier. Our results indicate that such methods may not provide substantial additional benefit in estimating DFDs. We found that four assessment points gave nearly as much information as using a combined sample with a mean of 16 assessment points, demonstrating that we do not need to measure depression severity more than four times over the course of a year to determine an accurate measure of DFD.

Our data do not allow us to determine an optimal number of assessment points for modeling DFD. It is possible that four is good but that five or six points would provide an adequate increase in accuracy to justify the added costs of assessment. We illustrated two approaches to selecting a cutoff for a DFD or a fully depressed day. For the HSCL-20 we illustrated selecting a cutoff based on a combination of cutoffs recommended in previous studies (23) and adjustment to the mean level of depression in our sample. For the PHQ-9 we used more general values associated with the instrument when used in population studies. Although the selection of cutoff values is not likely to affect between-group differences in any given analysis (because the selection is applied to both groups), the cutoffs have an impact on the magnitude of the DFD and hence its clinical relevance. Using either method has advantages in representing different groups, and the selection of cutoffs should be clearly articulated. Future studies could focus on determining optimal parameters for cutoff values (such as by using a daily-diary reporting method as a gold standard and comparing variable periodic assessment points with variable cutoffs).

## Conclusions

This study could have an impact on the future of clinical research with regard to depression treatment outcomes. Researchers have been debating the best methods for determining clinical significance, and very few solutions have been proposed that are useful. Metrics such as numbers needed to treat may be helpful in determining the overall effects of treatment but leave little information about the degree to which interventions have had an impact on individual lives. Jacobson and Truax (26) proposed a definition of clinical significance that is widely used in the psychological literature, but it assumes that the only clinically meaningful outcome for a psychiatric intervention is one that results in absolute eradication of symptoms; this is not a realistic expectation for real-world intervention. The DFD measure provides for more meaningful outcomes, and it has excellent face validity and direct clinical relevance to consumers of depression treatments.

## Acknowledgments

Original data collection was supported by grants from the John A. Hartford Foundation, the California Health Care Foundation, the Hogg Foundation, and the Robert Wood Johnson Foundation. Data analysis and preparation of this manuscript were supported by grants 5K24MH074717 and KL2RR025015 from the National Institute of Mental Health. The design, conduct, data collection, analysis, and interpretation of the results of this study were performed independently of the funders. The funding agencies also played no role in review or approval of the manuscript.

## References

1. Mallick R, Chen J, Entsuah AR, et al. Depression-free days as a summary measure of the temporal pattern of response and remission in the treatment of major depression: a comparison of venlafaxine, selective serotonin reuptake inhibitors, and placebo. *Journal of Clinical Psychiatry* 2003;64:321–330. [PubMed: 12716275]
2. Cohen J. A power primer. *Psychological Bulletin* 1992;112:155–159. [PubMed: 19565683]
3. Riso LP, Thase ME, Howland RH, et al. A prospective test of criteria for response, remission, relapse, recovery, and recurrence in depressed patients treated with cognitive behavior therapy. *Journal of Affective Disorders* 1997;43:131–142. [PubMed: 9165382]
4. Lave JR, Frank RG, Schulberg HC, et al. Cost-effectiveness of treatments for major depression in primary care practice. *Archives of General Psychiatry* 1998;55:645–651. [PubMed: 9672056]
5. Araya R, Flynn T, Rojas G, et al. Cost-effectiveness of a primary care treatment program for depression in low-income women in Santiago, Chile. *American Journal of Psychiatry* 2006;163:1379–1387. [PubMed: 16877650]

6. Ciechanowski PS, Russo JE, Katon WJ, et al. The association of patient relationship style and outcomes in collaborative care treatment for depression in patients with diabetes. *Medical Care* 2006;44:283–291. [PubMed: 16501401]
7. Liu CF, Hedrick SC, Chaney EF, et al. Cost-effectiveness of collaborative care for depression in a primary care veteran population. *Psychiatric Services* 2003;54:698–704. [PubMed: 12719501]
8. Montgomery SA, Andersen HF. Escitalopram versus venlafaxine XR in the treatment of depression. *International Clinical Psychopharmacology* 2006;21:297–309. [PubMed: 16877901]
9. Pyne JM, Tripathi S, Williams DK, et al. Depression-free day to utility-weighted score: is it valid? *Medical Care* 2007;45:357–362. [PubMed: 17496720]
10. Revicki DA, Siddique J, Frank L, et al. Cost-effectiveness of evidence-based pharmacotherapy or cognitive behavior therapy compared with community referral for major depression in predominantly low-income minority women. *Archives of General Psychiatry* 2005;62:868–875. [PubMed: 16061764]
11. Simon GE, Barber C, Birnbaum HG, et al. Depression and work productivity: the comparative costs of treatment versus non-treatment. *Journal of Occupational and Environmental Medicine* 2001;43:2–9. [PubMed: 11201765]
12. Simon RI. Suicide risk assessment: what is the standard of care? *Journal of the American Academy of Psychiatry and the Law* 2002;30:340–344. [PubMed: 12380411]
13. Trivedi MH, Wan GJ, Mallick R, et al. Cost and effectiveness of venlafaxine extended-release and selective serotonin reuptake inhibitors in the acute phase of outpatient treatment for major depressive disorder. *Journal of Clinical Psychopharmacology* 2004;24:497–506. [PubMed: 15349005]
14. Simon GE, Manning WG, Katzelnick DJ, et al. Cost-effectiveness of systematic depression treatment for high utilizers of general medical care. *Archives of General Psychiatry* 2001;58:181–187. [PubMed: 11177120]
15. Katon WJ, Unützer J, Fan MY, et al. Cost-effectiveness and net benefit of enhanced treatment of depression for older adults with diabetes and depression. *Diabetes Care* 2006;29:265–270. [PubMed: 16443871]
16. Katon W, Lin EH, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. *General Hospital Psychiatry* 2007;29:147–155. [PubMed: 17336664]
17. Rost K, Pyne JM, Dickinson LM, et al. Cost-effectiveness of enhancing primary care depression management on an ongoing basis. *Annals of Family Medicine* 2005;3:7–14. [PubMed: 15671185]
18. Spitzer RL, Williams JB, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *JAMA* 1994;272:1749–1756. [PubMed: 7966923]
19. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 2001;16:606–613. [PubMed: 11556941]
20. Derogatis LR, Lipman RS, Rickels K, et al. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behavioral Science* 1974;19:1–15. [PubMed: 4808738]
21. Unützer J, Katon WJ, Callahan CM, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 2002;288:2836–2845. [PubMed: 12472325]
22. Lowe B, Unützer J, Callahan, et al. Monitoring depression treatment outcomes with the Patient Health Questionnaire–9. *Medical Care* 2004;42:1194–1201. [PubMed: 15550799]
23. Simon GE. Evidence review: efficacy and effectiveness of antidepressant treatment in primary care. *General Hospital Psychiatry* 2002;24:213–224. [PubMed: 12100832]
24. Katon WJ, Schoenbaum M, Fan MY, et al. Cost-effectiveness of improving primary care treatment of late-life depression. *Archives of General Psychiatry* 2005;62:1313–1320. [PubMed: 16330719]
25. Simon GE, Katon WJ, Lin EH, et al. Cost-effectiveness of systematic depression treatment among people with diabetes mellitus. *Archives of General Psychiatry* 2007;64:65–72. [PubMed: 17199056]



26. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 1991;59:12–19. [PubMed: 2002127]

**Table 1**  
**Outcomes from the Improving Mood—Promoting Access to Collaborative Treatment trial based on various metrics**

Instrument	N	Baseline		12 month		Cohen's d	Proportion with 50% improvement	Estimated DFs <sup>d</sup>
		M	SD	M	SD			
Hopkins Symptom Checklist-20 <sup>b</sup>	868	1.67	.02	.92	.03	-1.23	47%	153
Patient Health Questionnaire-9 <sup>c</sup>	587	13.45	.03	4.70	.02	-1.78	69%	200

<sup>a</sup>Depression-free days

<sup>b</sup>Possible scores range from 0 to 4, with higher scores indicating more severe depression.

<sup>c</sup>Possible scores range from 0 to 27, with higher scores indicating more severe depression.