

# How should the samples be selected when evaluating the effectiveness of treatments for depression in routine clinical practice?

Mark Zimmerman, MD; Reina Kiefer, BA; Sophie Kerr, BA; Caroline Balling, BS; Theresa Morgan, PhD

**Introduction:** A better understanding of the effectiveness of psychiatric treatment in clinical practice will depend on systematically measuring outcome. In the present report from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we examined the impact of sample selection on evaluating the effectiveness of an intensive treatment program in treating depression. **Methods:** From a sample of 1,596 partial hospital patients, 656 (41.1%) received the Structured Clinical Interview for DSM-IV (SCID) (1) and the borderline personality section of the Structured Interview for DSM-IV Personality (SIDP-IV) (2) at intake. Most patients were not evaluated with these measures due to a lack of available interviewers. All patients completed a daily version of the Clinically Useful Depression Outcome Scale (CUDOS-D) (3) each day they attended the program. Analysis was limited to patients who attended the program for at least 3 days and completed the CUDOS-D at admission and discharge. We evaluated outcome in the following ways: (1) change in mean scores on the CUDOS-D from intake to discharge, (2) percent of patients who responded to treatment, with response defined as a reduction of the CUDOS-D scores from admission to discharge of at least 50%, (3) percent of patients who remitted, with remission defined at a discharge CUDOS-D score of 10 or less, and (4) effect size. **Data Analysis:** We examined the 4 outcome measures with 3 different approaches to sample selection: (1) in all patients, regardless of diagnosis, (2) in all patients, regardless of diagnosis, but requiring a minimum baseline CUDOS-D severity score, and (3) in 3 non-overlapping diagnostic groups of patients, also requiring a minimum baseline CUDOS-D severity score (a score of 20 indicates depression of at least mild severity and a score of 30 indicates that of at least moderate severity) (Table 1). Of the 1,596 partial hospital patients (65.4% female, 32.6% male, 2.0% transgender; mean age of 37.3 ± 13.9 years; 76.6% white), 56.8% were diagnosed with MDD (with an additional 3.3% in partial remission), 7.8% with bipolar depression (with an additional 0.2% in partial remission), 77.2% with any mood disorder diagnosis, and 22.8% had no mood disorder diagnosis. **Results:** Implementing a symptom severity requirement had the greatest impact on remission rates

and effect sizes: as the minimum severity threshold increased, the remission rate dropped and the effect size increased, regardless of diagnostic group. When examining the sample based on different diagnostic groups, we found that approximately half of the patients with MDD/bipolar depression responded to treatment and less than one-third remitted by time of discharge. There was no difference in the response or remission rate between patients with other mood disorders and no mood disorder. **Conclusions:** Given that the sample selection for effectiveness studies is intended to yield a sample that is more representative of patients treated in routine clinical practice, we do not plan to utilize a minimum severity threshold in future analyses of our outcome database. Further, we plan to examine diagnostic groups separately, to examine the change in mean scores from admission to the end of treatment, and to report response rates, remission rates, and effect sizes. We strongly recommend that a minimum set of reporting standards be adopted to enhance comparability across studies and to reduce the likelihood of bias due to selective reporting of results.

Table 1. Change in Scores on the CUDOS-D, Response Rates, Remission Rates, and Effect Size in Non-Overlapping Diagnostic Groups in 1,596 Partial Hospital Patients

	All patients	A MDD or Bipolar Depression <sup>a</sup>	B Other Mood Disorder <sup>b</sup>	C No Mood Disorder	3-group test <sup>d</sup>	Pairwise group differences
No min. baseline CUDOS-D score (n)	1,596	1,030	202	364		
Admission, M (SD)	36.0 (12.1)	38.9 (10.4)	31.9 (11.7)	30.1 (14.1)	62.1***	B, C<A***
Discharge, M (SD) <sup>c</sup>	18.6 (14.1)	20.5 (14.0)	14.8 (13.2)	15.4 (13.7)	377.5***	B, C<A***
% change in M scores	48.3	47.3	53.6	48.8		
Responder, % (n)	53.9 (860)	50.9 (524)	63.4 (128)	57.1 (208)	12.6**	A<B**; A<C*
Remission, % (n)	35.5 (583)	29.6 (305)	52.0 (105)	47.5 (173)	61.0***	A<B, C***
Effect size	1.3	1.5	1.4	1.1		
Min. baseline CUDOS-D ≥ 20 (n)	1,417	981	166	270		
Admission, M (SD)	38.9 (9.2)	40.1 (9.0)	35.8 (8.8)	36.6 (9.5)	17.9***	B, C<A***
Discharge, M (SD) <sup>c</sup>	19.9 (14.0)	20.9 (14.0)	16.3 (13.2)	18.5 (13.9)	270.2***	B<A***; C<A*
% change in M scores	48.8	47.9	54.5	49.5		
Responder, % (n)	53.1 (752)	50.6 (496)	63.3 (105)	55.9 (151)	10.3**	A<B**
Remission, % (n)	31.6 (448)	27.9 (274)	47.0 (78)	35.6 (96)	26.2***	A<B***; A<C*; C<B*
Effect size	1.6	1.6	1.7	1.5		
Min. baseline CUDOS-D ≥ 30 (n)	1,171	849	125	197		
Admission, M (SD)	41.8 (7.3)	42.4 (7.2)	39.3 (6.9)	40.9 (7.2)	7.8***	B<A***; C<A*
Discharge, M (SD) <sup>c</sup>	21.5 (14.2)	22.1 (14.1)	18.3 (13.5)	20.9 (14.6)	166.7***	B<A**
% change in M scores	48.6	47.9	53.4	48.8		
Responder, % (n)	52.0 (609)	49.9 (424)	60.8 (76)	55.3 (109)	6.2*	A<B*
Remission, % (n)	27.7 (324)	25.6 (217)	39.2 (49)	29.4 (58)	10.5**	A<B**
Effect size	1.8	1.8	2.0	1.7		

1. First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W. (1995). Structured Clinical Interview for DSM-IV Axis I Disorders – Patient edition (SCID-I/P, version 2.0). Biometrics Research Department, New York State Psychiatric Institute, New York.

2. Pfohl, B., Blum, N., Zimmerman, M. (1997). Structured Interview for DSM-IV Personality. American Psychiatric Press, Inc., Washington, D.C.

3. Zimmerman, M., Harris, L., Martin, J., McGonigal, P. (2018). Reliability and validity of a self-report scale for daily assessments of the severity of depressive symptoms. *Psychiatry Res.* 270, 581-586.