

Predicting Participation in Psychiatric Randomized Controlled Trials: Insights From the STEP-BD

Alisa B. Busch, M.D., M.S., Yulei He, Ph.D., Katya Zelevinsky, M.A., Alistair J. O'Malley, Ph.D.

Objective: Differences between patients who do and do not participate in randomized controlled trials (RCTs) could diminish the generalizability of results. This study examined whether RCT participants differ from non-RCT participants who are recruited from the same patient and provider population.

Methods: The Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) was an observational study in which participants also could enroll in an RCT during exacerbations of acute depression. The odds that a patient was enrolled in the STEP-BD acute depression RCTs (pharmacotherapy or psychotherapy) were estimated by fitting logistic regression models to STEP-BD participants with acute bipolar depression (total N=2,222; RCT, N=413; observational arm, N=1,809). Predictor variables included demographic characteristics, clinical information (including severity scales and comorbidities), and study site. The extent to which site determined RCT participation was estimated by using the area under the receiver operating characteristic curve (AUC).

Results: RCT participation was associated with having no insurance (odds ratio [OR]=1.58, 95% confidence interval [CI]=1.16–2.15), a Clinical Global Impression score indicating greater severity (severe versus mild: OR=1.52, CI=1.08–2.15), and site (predicted probability range 8%–31%). Site was the most significant predictor of RCT enrollment (model excluding site, AUC=.61, CI=.58–.64; full model, AUC=.70, CI=.67–.73).

Conclusions: STEP-BD RCT participants differed from those in the observational arm in few clinical or demographic characteristics. Site was the strongest predictor of RCT participation. Future study is needed to understand site characteristics associated with RCT participation and whether these characteristics are associated with patient outcomes and to test these findings in usual-care settings.

Psychiatric Services 2015; 66:817–823; doi: 10.1176/appi.ps.201300557

A criticism of randomized controlled trials (RCTs) is related to concerns about the generalizability of their results for clinical decision making in the broader patient population (1–7). These concerns stem from several factors. One is that RCTs often exclude individuals with clinical characteristics common to many patients seen in community settings, such as co-occurring substance use disorders, chronic general medical conditions, or suicidality (1,4,5,8,9). Another is that patients who participate in RCTs may differ from those seen in community settings on the basis of socioeconomic characteristics, educational attainment, or race-ethnicity (10). Patient characteristics that may influence an individual's participation in an RCT (for example, altruism and treatment adherence) (11) may also influence outcomes, and provider biases may influence which eligible patients are invited to participate. Another concern is that RCT results are not generalizable to usual-care settings, where care often is not delivered in highly protocol-guided, algorithmic ways and where structured outcomes are not routinely measured during treatment. These threats to the generalizability of RCT results to community populations and practice can have significant implications for our ability to translate knowledge gained from RCTs into an

understanding of which treatments will be effective for which patients and under what circumstances.

Broadening our understanding of patient characteristics associated with participation in an RCT could provide a useful context for interpreting RCT findings. We can examine some of the typical threats to RCT generalizability by using data from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) study (12), in which RCTs were embedded within a larger, multisite observational study population. Specifically, we can learn whether there are clinical or demographic differences between RCT and non-RCT participants who are drawn from a common pool of patients and sites or clinicians, when the patient pool is clinically diverse and described with considerable clinical detail, and when the RCTs have few clinical exclusion criteria.

METHODS

The STEP-BD Study and Population

The goal of the STEP-BD was to conduct clinical trials and other naturalistic studies that required a well-described, clinically diverse population of persons with bipolar disorder

(13). Twenty-one sites in 12 states participated. Several sites partnered with local clinics (six partnerships in five cities) to further increase participation by community clinics delivering mainstream care (14). Per STEP-BD policy, these local clinics did not contribute patients to the RCTs (personal communication, Sachs GS, 2011). STEP-BD study participants gave informed consent to participate in the observational arm and additional consent for RCT participation. Approval was obtained from an institutional review board (IRB) at each site. For the analysis reported here, further IRB approval was obtained from McLean Hospital and Harvard Medical School.

STEP-BD began in November 1999 and was conducted through September 2005. Recruitment advertising for the STEP-BD consisted of public service announcements by the National Institute of Mental Health (NIMH) released in several cities that contained STEP-BD sites. Sites were quickly inundated with prospective participants, obviating the need for further active recruitment (15).

STEP-BD participation was offered to new or existing patients at STEP-BD sites who met study criteria for bipolar disorder. Patients were informed about STEP-BD by their program psychiatrist. Participation meant, at a minimum, entering the observational study arm, which served as an overall structure for assessment and treatment of bipolar disorder. Providers in the observational arm received additional training in bipolar disorder treatment, but their treatment choices were not constrained. Participants in the observational arm who met criteria for one of the RCTs were offered an opportunity to participate in the RCT. RCT participants underwent additional assessments, as well as randomized assignment to the RCT treatment protocols. RCT enrollment could begin at any point of a person's STEP-BD participation (that is, at registration or thereafter) (13).

We compared two STEP-BD populations. Persons enrolled in at least one of two STEP-BD RCTs for the treatment of acute bipolar depression (the adjunctive antidepressant RCT or the psychosocial treatment RCT) (16,17) were compared with those in the observational arm who did not participate in either of these RCTs. In the adjunctive antidepressant RCT, participants were randomly assigned to receive either an adjunctive antidepressant medication or a placebo. In the psychosocial treatment RCT, participants were randomly assigned to receive one of three intensive psychosocial treatments or to a control arm of three educational sessions.

In brief, to be eligible, participants in both acute depression RCTs had to be adults (age 18 or older) who met *DSM-IV* criteria (18) for bipolar I or bipolar II disorder. Diagnoses were determined by a modified Structured Clinical Interview for DSM Disorders (19) and confirmed by the Mini-International Neuropsychiatric Interview (MINI) (20). Participants in the acute depression RCTs also met *DSM-IV* criteria for a major depressive episode and consented to take mood stabilizer and antipsychotic medication concomitantly (16). Few RCT exclusion criteria were employed. Both RCTs excluded persons who required short-term treatment for an active substance use disorder and those who were pregnant or planning to

become pregnant in the coming year. Additional exclusion criteria for the adjunctive antidepressant RCT were history of nonresponse to the study antidepressants (bupropion and paroxetine) and either introduction of an antipsychotic or a change in dosage of an antipsychotic that had been prescribed for a long time. In the psychosocial treatment RCT, individuals unwilling to discontinue their current (nonstudy) psychotherapy or taper the sessions to one or two per month were excluded. Individuals could choose to participate in the adjunctive antidepressant RCT and not the psychosocial treatment RCT. However, the psychosocial treatment RCT was initially limited to participants in the adjunctive antidepressant RCT. Study investigators later modified this to allow participation of persons ineligible for the adjunctive antidepressant RCT because of a history of nonresponse to the study antidepressants.

Time-varying clinical characteristics (mood state and symptom severity) were noted on the clinical monitoring form (CMF), a template progress note for the STEP-BD observational and RCT arms. We defined "index acute bipolar-depressed visits" for participants in each study arm. For RCT participants, we defined the index visit as the CMF completed closest to the date of RCT randomization. Preliminary analyses indicated that this occurred from seven days before to seven days after randomization for 92% (N=380) of the RCT sample. For participants in the observational arm who were acutely depressed and who had never been enrolled in either of the acute depression RCTs (adjunctive antidepressant or psychosocial treatment), we took the first major depression clinical status noted in a CMF as the index acute bipolar-depressed CMF visit. We excluded from our sample the participants without a CMF and RCT participants for whom we were not able to identify the RCT randomization date.

Primary Outcome

Our primary outcome was a dichotomous variable designating whether or not a STEP-BD participant had been enrolled in either of the RCTs.

Explanatory Variables

We compared demographic characteristics and clinical characteristics of the two groups (that is, those enrolled in an acute depression RCT or not). We also included in the model a categorical variable for site. The demographic characteristics were age (centered), gender, race-ethnicity, education, income (divided by the median income of \$40,000), and insurance type. Clinical characteristics included those related to bipolar disorder symptoms (domains from the Bipolarity Index [BPI] [21]) and baseline severity scores on the Clinical Global Impression (CGI) scales (22). The BPI is a categorical measure describing a patient's bipolar disorder symptom history, illness course, age at onset, family history, and prior treatment response. We included the two domains that we felt would be most pertinent for this study: symptom history and prior treatment response. We categorized the CGI as 1–3, no or mild symptoms; 4, moderate symptoms; and 5–7, severe symptoms.

We also included variables for specific co-occurring psychiatric and general medical conditions and characterized the comorbidity “burden” as categorical variables (for example, zero, one, two, or three or more co-occurring conditions). The co-occurring psychiatric and general medical conditions included were those that could complicate or otherwise influence bipolar disorder pharmacotherapy prescribing. For psychiatric conditions, these were anxiety disorders, attention-deficit hyperactivity disorder, and eating disorders. For general medical conditions, these were pregnancy or hepatic, renal, pancreatic, seizure, thyroid, or inflammatory disorders. Variables were also included for conditions or patient characteristics that often lead to exclusion from or influence selection into clinical trials (for example, substance use disorders).

The BPI and CGI are clinician-rated scales. All other explanatory variables were based on patient self-report, with the exception of comorbid mental and substance use disorders, which were determined by the MINI (20).

Statistical Analyses

This study was conducted with preexisting data from the STEP-BD, which we obtained from the STEP-BD Publications Committee; we subsequently obtained approval from NIMH for use of the data. We computed descriptive statistics (means and standard deviations) for the sample. Because some covariates were missing for some participants, we multiply imputed missing values. Our findings are based on combined multiply imputed data sets. We then fitted a mixed-effects logistic regression model in which site was a random effect (referred to as the full model). The mixed-effects model also enabled us to test the significance of site as an independent predictor of RCT participation allowing for the total number of patients the site contributed to STEP-BD. To separately quantify the impact of each characteristic significantly associated with RCT enrollment in the mixed-effects model, we fitted separate mixed-effects logistic regression models, each excluding a variable found to be significant in the full model.

To interpret the impact of significant variables, we estimated the receiver operating characteristic (ROC) curves of the fitted probabilities under both models (that is, the full model and the model missing the significant variable) and evaluated the difference in the area under the curves (AUCs). The AUC represents the probability that the model correctly classifies whether or not a randomly selected participant enrolls in an RCT. This can be thought of as a scale-free standardized effect size of a given explanatory variable on RCT participation. We treated the site dummies as random effects so that inferences pertain to an entire population of sites providing psychiatric care as opposed to their representing the sites that participated in STEP-BD. We then calculated the mean predicted probability of a particular site enrolling a patient in the RCT by using the results from the mixed-effects logistic regression model (23). This yielded mean site-specific probabilities adjusted for patient characteristics, thereby ensuring valid comparisons.

In discussions with the STEP-BD principal investigator, we learned that one site expressed reluctance to enroll participants with health insurance in the adjunctive antidepressant RCT (personal communication, Sachs GS, 2011). The reason given for this reluctance was that additional health care costs resulting from potential adverse outcomes related to study participation (for example, if participation in one study arm was harmful or was less efficacious than participation in another) would need to be shouldered by the health insurance plans. Therefore, we conducted a post hoc analysis in which we excluded that site to determine whether it altered the results for the association between insurance status and RCT participation.

We conducted other post hoc sensitivity analyses. Given that patient or site characteristics associated with enrollment in an adjunctive antidepressant RCT may differ from those associated with enrollment in a psychosocial treatment RCT, we fitted a separate model excluding participants who entered only the psychosocial RCT. However, the limited number of participants who enrolled only in the psychosocial RCT precluded us from also fitting a separate model to them alone. Although our fixed-effects modeling enabled us to examine site contribution to the RCT independent of site size, as an additional sensitivity analysis, we added a variable to the model that controlled for site volume in order to reduce the component of site variation in RCT enrollment that was explained by site volume.

RESULTS

We excluded 12 RCT participants either because they lacked CMFs or because we were unable to match them to the enrollment file where the date of their consent to the RCT was located. Bivariate analyses found no difference between participants in the RCT or in the observational arm in rates of missing data for any of the characteristics. Our total sample size was 2,222 (RCT, $N=413$; observational arm, $N=1,809$). As expected on the basis of the linked inclusion criteria of the two RCTs, the RCT participant populations overlapped considerably: among the 413 individuals who participated in at least one of the two STEP-BD acute depression RCTs, 56% ($N=233$) participated in both. The 233 individuals who participated in both RCTs represented 65% of the adjunctive antidepressant RCT participants ($N=359$), and 81% of the psychosocial treatment RCT participants ($N=287$).

Participants in the observational arm and in the RCTs were largely white ($>85\%$) (Table 1). About half had at least a college degree, and most were privately insured ($>55\%$). CGI scores were predominantly in the moderate to severe range; about half had scores in the moderate range, and 22%–26% had scores in the severe range. About half had a co-occurring substance use disorder, over two-thirds had one or more comorbid mental health conditions that were not substance use disorders, and nearly a third had one or more of the general medical conditions that can influence pharmacotherapy choices for bipolar disorder.

TABLE 1. Characteristics of 2,222 acutely depressed patients with bipolar disorder in STEP-BD by whether they enrolled in either of the acute bipolar depression RCTs or not^a

Characteristic	Observational arm (N=1,809)			RCT (N=413)		
	N	%	Missing data (%)	N	%	Missing data (%)
Demographic						
Age (M±SD)	40.2±12.5		0	40.5±11.5		<1
Male	711	39	<1	173	42	1
Race-ethnicity			<1			<1
White	1,574	87		366	89	
Black	80	4		22	5	
Hispanic	91	5		15	4	
Other	62	3		8	2	
Education			5			5
Not a high school graduate	50	3		9	2	
High school graduate	267	15		70	17	
Some college	460	25		117	28	
College graduate	936	52		198	48	
Income ≥\$40,000	786	44		160	39	
Insurance status			4			3
Medicare	221	12		40	10	
Medicaid	96	5		25	6	
Both	47	3		10	2	
Private	1,083	60		221	54	
None	298	17		104	25	
Clinical						
Bipolarity Index: course of illness			3			3
Symptoms with no evidence of bipolar disorder	21	1		3	<1	
Symptoms with possible relationship to or suggestive of bipolar disorder	245	14		64	16	
Known associated feature or convincing or most convincing characteristic of bipolar disorder	1,489	82		335	81	
Bipolarity Index: past bipolar treatment response			3			3
Symptoms with no relationship to bipolar disorder	116	6		37	9	
Symptoms with possible relationship to or suggestive of bipolar disorder	125	7		30	7	
Known associated feature or convincing or most convincing characteristic of bipolar disorder	1,513	84		334	81	
Clinical Global Impression scale			1			<1
1–3 (no or mild symptoms)	436	24		72	17	
4 (moderate symptoms)	937	52		230	56	
5–7 (severe symptoms)	415	23		108	26	
Comorbid mental or substance use disorder ^b						
Substance use disorder	867	48	8	223	54	6
Anxiety disorder ^c	1,041	58	8	256	62	6
Eating disorder	183	10	8	37	10	6
ADHD	220	12	8	52	13	6
Any of the above comorbid disorders except substance use disorder	1,128	62	8	272	66	6
None	545	30		114	28	
1	849	47		206	50	
2	242	13		61	15	
≥3	37	2		6	2	

continued

In the full mixed-effects model (Table 2), being uninsured, compared with having private insurance, was a significant predictor of RCT participation (odds ratio [OR]=1.58), as was a baseline CGI score in the severe range (OR=1.52), compared with a score indicating mild symptoms. Four sites had significantly higher odds of contributing patients to the RCT, whereas three had significantly lower odds: site C, OR=2.23; site K, OR=1.67; site L, OR=2.10; site P, OR=2.21; site B, OR=.59; site F, OR=.51; and site Q, OR=.40. The mean predicted probabilities of individual sites contributing patients to the RCT ranged from 8% to 31% (data not shown). Despite the STEP-BD policy that patients from community clinics would not be enrolled in the RCTs, among the five STEP-BD sites that partnered with the six community clinics, only one site had lower odds of contributing patients to an RCT. For the others, no greater or lesser likelihood was noted.

When the full mixed-effects model was compared with the model that did not include site, the difference in AUC suggested that site increased the accuracy of the model by 9 percentage points (model excluding site, AUC=.61; full model, AUC=.70) (Table 3). Insurance status did not significantly change the accuracy of the ROC for the model that excluded insurance, nor did excluding the baseline CGI score (AUC=.70).

None of our sensitivity analyses (adding site volume to the mixed-effects model, dropping persons who participated only in the psychosocial RCT, or dropping the

study site where site investigators expressed reluctance to enroll insured participants in an RCT) yielded different results.

DISCUSSION

Site was the strongest determinant of RCT participation. That is, not all STEP-BD sites contributed similar proportions of patients from their observational arm to the acute depression RCTs. Our findings are independent of the actual number of overall STEP-BD participants from a given site. This finding is notable because clinics were selected for STEP-BD participation on the basis of criteria that would favor a capacity for completing RCT research tasks.

Site contributions to RCT participation are of interest because the outcomes of clinical trials often vary by site (that is, a site $\times$ treatment interaction effect) (3,24–26). This effect can be attributable to variation in protocol adherence and in participant characteristics, although one purpose of multisite studies is to increase diversity in key participant characteristics (for example, race-ethnicity and socioeconomic characteristics) to improve generalizability (27). Our findings of differential recruitment to RCTs by site raises important questions that require further study about whether site-specific enrollment to a multisite RCT may be related to a site's resources to deliver care in general. For example, “high-enrolling” clinics may have different staffing composition or clinician-to-patient ratios that have an impact on care delivery or treatment quality. Important future work includes examining site characteristics in multisite clinical trials to understand why sites may differ in clinical trial outcomes and how we can interpret and extend results from RCTs to usual-care settings and populations.

Few clinical or demographic differences were found between participants in the STEP-BD acute depression RCTs and the observational arm. In contrast, previous studies have found that exclusion criteria typically used in RCTs often exclude patients with more complex presentations or heterogeneous characteristics (1,5,9). Our finding of few clinical or demographic differences between participants in the RCTs and in the observational arm is perhaps not surprising given STEP-BD's goal of a broader representation of patients with bipolar disorder than typically seen in clinical trials. Nevertheless, it is noteworthy that the STEP-BD achieved this aim for the RCTs.

TABLE 1, continued

Characteristic	Observational arm (N=1,809)			RCT (N=413)		
	N	%	Missing data (%)	N	%	Missing data (%)
Comorbid general medical condition ^d						
Seizure disorder	119	7	4	25	6	5
Thyroid disorder	302	17	2	62	15	3
Hepatic disorder	106	6	4	24	6	5
Renal disorder	4	<1	0	3	<1	0
Pancreatic disorder	1	<1	0	0	—	0
Inflammatory disease	67	4	0	14	3	0
Pregnancy	3	<1	0	1	<1	0
Any of the above comorbid general medical conditions			4			5
None	1,221	68		285	69	
1	70	4		17	4	
2	355	20		75	18	
≥ 3	87	5		17	4	

^a STEP-BD, Systematic Treatment Enhancement Program for Bipolar Disorder; RCT, randomized controlled trials

^b Current or past disorder (or both) confirmed by the Mini-International Neuropsychiatric Interview. The disorders listed are those that can affect the prognosis or pharmacotherapy of bipolar disorder or adherence to treatment.

^c Includes generalized anxiety disorder, PTSD, phobias, obsessive-compulsive disorder, agoraphobia, and panic disorder

^d Self-reported by study participant. The conditions listed are those that could complicate or influence options for bipolar disorder pharmacotherapy.

Also in contrast to prior research on RCTs that involved patients with general medical conditions (10,28), our study found that a lack of insurance was associated with RCT participation. This finding could be mediated by investigator biases or patient preferences and choices. Excluding from our analysis the STEP-BD site where investigators expressed reluctance to recruit insured individuals did not change the results; other investigators or sites may have had a similar reluctance but did not express it. If our finding is attributable to investigator biases, it raises ethical concerns that investigators may view patients with and without insurance differently in regard to RCT participation. Alternatively, uninsured patients may be more likely than those who are insured to choose RCT participation.

In STEP-BD, the RCT paid for the study treatments (for example, study antidepressants and psychosocial visits) but not for RCT participation, non-RCT-related study visits, or any other psychotropic medications. Thus study participation offered some financial benefit to uninsured individuals. If financial constraints due to lack of insurance encouraged some to participate in the RCTs, then this raises concerns about the disproportionate burden that persons without insurance may bear for clinical trials as a means to procure health care. Federal legislation to mandate parity and reduce the number of uninsured persons—that is, the Mental Health Parity and Equity Act and the Affordable Care Act—could reduce this ethical concern by providing more patients the opportunity to receive mental health care outside a clinical trial.

A limitation of our study is that it could not address all the potential threats to RCT generalizability. STEP-BD participants likely differed from patients seen in some community settings, such as in demographic characteristics (for example,

TABLE 2. Analysis of predictors of RCT enrollment among 2,222 acutely depressed patients with bipolar disorder in STEP-BD^a

Variable	OR	95% CI
Demographic		
Age (centered)	1.01	.99–1.02
Male (reference: female)	1.17	.91–1.49
Race-ethnicity (reference: white)		
Black	1.04	.61–1.77
Hispanic	1.06	.62–1.81
Other	.61	.29–1.30
Education (reference: not a high school graduate)		
High school graduate	1.32	.68–2.53
Some college	1.36	.68–2.53
College graduate	1.14	.61–2.15
Household income $\geq$ \$40,000 (reference: $<$ \$40,000)	.83	.62–1.11
Insurance status (reference: private)		
Medicare only	.85	.56–1.29
Medicaid only	1.17	.68–2.01
Both Medicare and Medicaid	.83	.37–1.86
None	1.58	1.16–2.15
Clinical		
Bipolarity Index: course of illness (reference: symptoms with no evidence of bipolar disorder)		
Symptoms with possible relationship to or suggestive of bipolar disorder	1.51	.74–3.10
Known associated feature or convincing or most convincing characteristic of bipolar disorder	1.32	.67–2.61
Bipolarity Index: past bipolar treatment response (reference: symptoms with no relationship to bipolar disorder)		
Symptoms with possible relationship to or suggestive of bipolar disorder	.68	.38–1.21
Known associated feature or convincing or most convincing characteristic of bipolar disorder	.70	.46–1.05
Clinical Global Impression scale (reference: no or mild symptoms)		
Moderate symptoms	1.36	1.00–1.83
Severe symptoms	1.52	1.08–2.15
Substance use disorder (current or past) (reference: none)	1.12	.87–1.44
N of comorbid mental or substance use disorders (reference: none) ^b		
1	1.03	.79–1.36
2	1.01	.69–1.48
≥ 3	.87	.30–2.47
N of comorbid general medical conditions (reference: none) ^c		
1	.88	.49–1.56
2	.89	.66–1.22
≥ 3	.86	.50–1.48
Site		
A	.95	.28–3.15
B	.59	.37–.95
C	2.23	1.45–3.44
D	.98	.39–2.49
E	.82	.26–2.57

*continued***TABLE 2, continued**

Variable	OR	95% CI
F	.51	.32–.83
G	.71	.47–1.09
H	.85	.26–2.72
I	1.24	.69–2.23
J	1.69	.96–2.97
K	1.67	1.01–2.76
L	2.10	1.01–4.39
M	.83	.46–1.48
N	1.73	.99–3.05
O	.61	.25–1.50
P	2.21	1.38–3.54
Q	.40	.20–.80
R	1.34	.67–2.70
S	.60	.24–1.48
T	1.00	.56–1.77
U	.70	.27–1.82

^a Mixed-effects regression analysis with site as a random effect. RCT, randomized controlled trial; STEP-BD, Systematic Treatment Enhancement Program for Bipolar Disorder

^b Disorders assessed were those that can affect the prognosis or pharmacotherapy of bipolar disorder or adherence to treatment.

^c Conditions assessed were those that could complicate or influence options for bipolar disorder pharmacotherapy.

race-ethnicity, educational attainment, and income) and in openness to participate in observational research. However, STEP-BD seemed to attract a patient population with a clinical complexity similar to that seen in usual-care settings; for example, compared with the community mental health center population in the Texas Medication Algorithm Project, STEP-BD participants had similar or higher proportions of chronic general medical conditions and co-occurring substance use disorders (29).

CONCLUSIONS

Our findings highlight the importance of future research focused on understanding not only how RCT patient populations may differ from patients seen in usual care but also the heterogeneity among clinics that participate in RCTs (that is, within an RCT) and whether this heterogeneity is predictive of patient outcomes as well.

TABLE 3. Comparison of effect sizes of models predicting RCT enrollment among 2,222 acutely depressed patients with bipolar disorder in STEP-BD^a

Model	AUC ^b	95% CI
Mixed-effects logistic model, site as random effect	.70	.67–.73
Fixed-effects logistic model, excluding site	.61	.58–.64
Mixed-effects logistic model, excluding past week CGI ^c	.70	.67–.72
Mixed-effects logistic model, excluding insurance	.69	.67–.72

^a RCT, randomized controlled trial; STEP-BD, Systematic Treatment Enhancement Program for Bipolar Disorder

^b Area under the curve

^c CGI, Clinical Global Impression scale

AUTHOR AND ARTICLE INFORMATION

Dr. Busch and Ms. Zelevinsky are with the Department of Health Care Policy, Harvard Medical School, Boston (e-mail: abusch@hcp.med.harvard.edu), where Dr. He was affiliated when this work was done. Dr. Busch is also with McLean Hospital, Belmont, Massachusetts. Dr. He is now with the Office of Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland. Dr. O'Malley is with the Dartmouth Institute of Health Policy and Clinical Practice, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire.

The authors gratefully acknowledge support from the National Institute of Mental Health (grants RC4 MH092717 and K01MH071714 to Dr. Busch) and from the Health Services Research Division, Partners Psychiatry and Mental Health. They also thank Andrew Nierenberg, M.D., and Gary Sachs, M.D., for sharing their knowledge and expertise about the STEP-BD study. The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics.

The authors report no financial relationships with commercial interests.

Received December 13, 2013; revision received October 27, 2014; accepted December 22, 2014; published online April 1, 2015.

REFERENCES

- Blanco C, Olfson M, Okuda M, et al: Generalizability of clinical trials for alcohol dependence to community samples. *Drug and Alcohol Dependence* 98:123–128, 2008
- Westen D, Morrison K: A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology* 69:875–899, 2001
- Peduzzi P, Henderson W, Hartigan P, et al: Analysis of randomized controlled trials. *Epidemiologic Reviews* 24:26–38, 2002
- Okuda M, Hasin DS, Olfson M, et al: Generalizability of clinical trials for cannabis dependence to community samples. *Drug and Alcohol Dependence* 111:177–181, 2010
- Le Strat Y, Rehm J, Le Foll B: How generalisable to community samples are clinical trial results for treatment of nicotine dependence? A comparison of common eligibility criteria with respondents of a large representative general population survey. *Tobacco Control* 20: 338–343, 2011
- Gross CP, Mallory R, Heiat A, et al: Reporting the recruitment process in clinical trials: who are these patients and how did they get there? *Annals of Internal Medicine* 137:10–16, 2002
- Wright JR, Bouma S, Dayes I, et al: The importance of reporting patient recruitment details in phase III trials. *Journal of Clinical Oncology* 24:843–845, 2006
- Rabinowitz J, Bromet EJ, Davidson M: Are patients enrolled in first episode psychosis drug trials representative of patients treated in routine clinical practice? *Schizophrenia Research* 61:149–155, 2003
- Zimmerman M, Chelminski I, Posternak MA: Generalizability of antidepressant efficacy trials: differences between depressed psychiatric outpatients who would or would not qualify for an efficacy trial. *American Journal of Psychiatry* 162:1370–1372, 2005
- Baquet CR, Commiskey P, Daniel Mullins C, et al: Recruitment and participation in clinical trials: socio-demographic, rural/urban, and health care access predictors. *Cancer Detection and Prevention* 30: 24–33, 2006
- Halpern SD, Karlawish JHT, Casarett D, et al: Hypertensive patients' willingness to participate in placebo-controlled trials: implications for recruitment efficiency. *American Heart Journal* 146:985–992, 2003
- Treatment for Bipolar Disorder. Rockville, Md, National Institute of Mental Health, 1998. Available at grants.nih.gov/grants/guide/notice-files/not98-051.html
- Sachs GS, Thase ME, Otto MW, et al: Rationale, design, and methods of the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Biological Psychiatry* 53:1028–1042, 2003
- Kogan JN, Bauer MS, Dennehy EB, et al: Increasing minority research participation through collaboration with community outpatient clinics: the STEP-BD Community Partners Experience. *Clinical Trials* 6:344–354, 2009
- Recruitment in STEP-BD. Cary, NC, 3-C Institute for Social Development, 4researchers.org, 2007
- Sachs GS, Nierenberg AA, Calabrese JR, et al: Effectiveness of adjunctive antidepressant treatment for bipolar depression. *New England Journal of Medicine* 356:1711–1722, 2007
- Miklowitz DJ, Otto MW, Frank E, et al: Psychosocial treatments for bipolar depression: a 1-year randomized trial from the Systematic Treatment Enhancement Program. *Archives of General Psychiatry* 64:419–426, 2007
- Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Washington, DC, American Psychiatric Association, 1994
- Spitzer RL, Williams JB, Gibbon M, et al: The Structured Clinical Interview for DSM-III-R (SCID): I. history, rationale, and description. *Archives of General Psychiatry* 49:624–629, 1992
- Sheehan DV, Lecrubier Y, Sheehan KH, et al: The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry* 59(suppl 20):22–33, quiz 34–57, 1998
- Sachs GS: Strategies for improving treatment of bipolar disorder: integration of measurement and management. *Acta Psychiatrica Scandinavica* 110(422):7–17, 2004
- Guy W: ECDEU Assessment Manual for Psychopharmacology, Revised. Rockville, Md, US Department of Health and Human Services, 1976
- Little RJA: Direct standardization: a tool for teaching linear models for unbalanced data. *American Statistician* 36:38–43, 1982
- Kaplan EL, Oakes JM, Johnson DR: Unexpected individual clinical site variation in eradication rates of group A streptococci by penicillin in multisite clinical trials. *Pediatric Infectious Disease Journal* 26:1110–1116, 2007
- McCarty D, Buti A, Kunkel LE, et al: Community-based clinical trials: site variation and adoption of innovation. Presented at the annual meeting of the American Public Health Association, Washington, DC, Oct 28–Nov 2, 2011
- Spirito A, Abebe KZ, Iyengar S, et al: Sources of site differences in the efficacy of a multi-site clinical trial: the Treatment of SSRI-Resistant Depression in Adolescents. *Journal of Consulting and Clinical Psychology* 77:439–450, 2009
- Kraemer HC: Pitfalls of multisite randomized clinical trials of efficacy and effectiveness. *Schizophrenia Bulletin* 26:533–541, 2000
- Blanch DC, Rudd RE, Wright E, et al: Predictors of refusal during a multi-step recruitment process for a randomized controlled trial of arthritis education. *Patient Education and Counseling* 73: 280–285, 2008
- Suppes T, Rush AJ, Dennehy EB, et al: Texas Medication Algorithm Project, phase 3 (TMAP-3): clinical results for patients with a history of mania. *Journal of Clinical Psychiatry* 64:370–382, 2003