Published in final edited form as:

Med Care. 2012 April ; 50(4): 311–319. doi:10.1097/MLR.0b013e3182422aec.

Accurately Predicting Bipolar Disorder Mood Outcomes— Implications for the Use of Electronic Databases

Alisa B. Busch, MD, MS^{1,2,3}, Brian Neelon, PhD⁴, Katya Zelevinsky, MA³, Yulei He, PhD³,
and Sharon-Lise T. Normand, Ph.D.³

¹McLean Hospital, Belmont, MA

²Department of Psychiatry, Harvard Medical School

³Department of Health Care Policy, Harvard Medical School

⁴Nicholas School of Environment, Duke University

Abstract

Background—Successful use of health information to monitor mental health treatment outcomes will require understanding which patient information is needed in electronic format and is feasible to obtain in routine care.

Objective—To examine whether bipolar disorder outcomes can be accurately predicted using information of limited clinical detail but feasible to collect electronically.

Research Design, Data Sources and Subjects—Longitudinal study of bipolar disorder patients treated 2000–2004 in the 19-site Systematic Treatment Enhancement Program for Bipolar Disorder observational study arm (N=3,168). Clinical data were obtained at baseline and quarterly over one year. We fit a “gold-standard” longitudinal random-effects regression model using detailed clinical information and estimated the area under the ROC curve (AUC) to predict accuracy. The model was then modified to include patient characteristics feasible in routinely collected electronic data (such as in administrative data). We compared the AUCs for the “limited-detail” and gold-standard models using the full sample and clinical subgroups.

Measure—Remission, defined as a Montgomery-Asberg Depression Rating Scale score<5 and a Young Mania Rating Scale score<4.

Results—The gold-standard models had baseline AUC of 0.91 ([95% CI=0.89–0.93]) and 0.89[0.86–0.92] at one-year follow-up. The limited-detail models had similar predictive accuracies (all AUC = 0.89). Both the gold-standard and limited-detail models performed well and similarly in predicting remission rates at one year.

Conclusions—Predicting bipolar disorder remission is feasible and accurate, even using data of limited clinical detail that are feasible to gather in routine care.

Keywords

observational data; behavioral health care; bipolar disorder; health services research; outcomes assessment

Introduction

The U.S. spends more for its health care than other developed nations, yet fares comparatively poorly in outcomes.¹ Increasingly, health policy is focused on using electronic health information to help reduce unnecessary costs and improve outcomes through improved clinical information sharing.^{2,3} Over 22 billion dollars will be spent as part of the American Recovery and Reinvestment Act of 2009 to speed adoption of electronic health information technology towards these aims.^{4,5}

To succeed in improving treatment quality and outcomes across a system of care, the electronic data gathered must be both clinically useful for predicting outcomes and financially feasible to obtain. Deciding which or how much information should be extracted electronically across systems of care is an important step in achieving routine quality assessment. For mental health treatment, no existing research provides guidance as to how much detail regarding patient characteristics or process of care measures is needed to accurately predict outcomes. Therefore, it is important to determine if clinically limited but easily obtained data can be used to discriminate outcomes.

We examine this question using observational data from the National Institute of Mental Health funded Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) study. Bipolar disorder is a prevalent, disabling, and deadly illness^{6–10} affecting an estimated 2.6% of the US adult population.¹¹ It is characterized by elevated mood states, which can impair function and judgment, and frequently treatment-resistant depressions. The goal of the STEP-BD observational arm was to provide an infrastructure for longitudinal research on bipolar disorder outcomes in a demographically and clinically heterogeneous bipolar population. Participating STEP-BD clinics were often located in academic medical centers but community provider clinics were included as well. In the STEP-BD, bipolar diagnoses were established via structured clinical interviews and additional clinical information by semi-structured interview. Additionally, while providers received additional education about treating bipolar disorder, decisions by the treating psychiatrists were not constrained.¹² Thus, STEP-BD patients and providers are a hybrid between what is seen in academic and usual care.

In this study, STEP-BD observational data are used to construct two sets of models. In one, we use detailed clinical information to predict remission rates. We compare these results to a second model in which the clinical detail is limited to information typically seen in administrative billing files. Our aims are twofold: 1) to determine the accuracy of clinical registry data to serve as a “gold standard” for predicting bipolar disorder outcomes, and 2) to determine if demographically and clinically limited data, such as that already exist in administrative billing data, can adequately predict bipolar disorder outcomes. If so, then this study can provide valuable guidance from a health policy perspective regarding patient

characteristics that can feasibly be captured electronically across a system of care for outcomes assessment.

Methods

Data Source and Population

The STEP-BD participants were recruited from 19 sites across 12 states. Study subjects gave informed consent and IRB approval was obtained at each site for participation in STEP-BD. Further IRB approval was obtained by the McLean Hospital and Harvard Medical School for this analysis.

The STEP-BD observational arm study protocol called for all participants to receive a baseline and quarterly assessments for the first year. The baseline assessment was completed by the treating psychiatrist and included the Affective Disorders Evaluation (ADE). The ADE contains a semi-structured format to assist in establishing a DSM-IV diagnosis and to obtain other clinical information useful for diagnosis and treatment planning. Mood disorders were diagnosed using a modified version of the Structured Clinical Inventory DSM-IV Axis I Disorders [SCID] mood modules. No other potentially comorbid psychiatric disorder had a structured diagnostic format. In addition to the ADE, the treating psychiatrist completed a Care Monitoring Form at baseline and scheduled appointments with the patient. Unscheduled visits to the treating psychiatrist were documented in a Brief Contact-Clinical Monitoring Form. These forms detailed changes in affective symptoms, medications, and adherence to medications.

At baseline, clinical specialists (non-treating psychiatrist, psychologist, social worker, or psychiatric nurse) independently diagnosed patients using the Mini-International Neuropsychiatric Interview (MINI).¹³ They also conducted baseline and quarterly assessments that included the Young Mania Rating Scale (YMRS)¹⁴ and the Montgomery-Asberg Depression Rating Scale (MADRS).¹⁵ These scales provide a measure of current manic and depressive mood symptoms and severity. Baseline and quarterly assessments also included a Care Utilization Form, in which study subjects detailed their mental health services use over the prior three months. Baseline demographic information was obtained by trained research assistants. We included all STEP-BD participants who entered the study between years 2000 and 2004 (N=3,168).

Primary Outcome

Our primary outcome of interest was euthymia/remission, defined as a YMRS score<4 and a MADRS score<5, which was measured at each quarterly follow-up. While these thresholds are lower than those typically seen in the literature, research has shown that they correspond to a global clinical measure of remission for patients with bipolar disorder.¹⁶ We also examined YMRS<7 and MADRS<10 as an alternative, less stringent cut-off that allows some minimal symptoms and has some support in the literature as well.^{17,18}

Explanatory Variables

We fit two separate models: a “gold-standard” model that contained richer clinical detail and a second, modified model containing variables of limited demographic and clinical detail, henceforth denoted the “limited-detail” model. Covariates thought to be correlated with symptom severity, treatment complexity, or remission were chosen based on clinical knowledge and prior literature (Table 1). Both models included self-report explanatory variables for baseline data describing study entry year, age (as a continuous variable), sex, race/ethnicity, insurance type, and the presence of a comorbid seizure disorder. Also included in the model was longitudinal information on treatment characteristics, such as medications prescribed and mean mental health visits (inpatient, outpatient or emergency room) during the 3 months prior to each visit, and patient self-report of adherence with antimanic medication.

Both models included explanatory variables for bipolar disorder type (bipolar-I/schizoaffective disorder, bipolar-II and other bipolar spectrum disorders) and specific comorbid psychiatric illnesses (or their pharmacologic treatment) that can worsen the course of bipolar disorder. These include substance use disorders [SUDs], anxiety disorders, and attention deficit hyperactivity disorder [ADHD]. However, comorbid diagnostic information for the gold-standard model was obtained via the structured MINI whereas for the limited-detail model, comorbid diagnostic information was collected from the baseline ADE (semi-structured for non-mood diagnoses). We used the ADE diagnoses for the limited-detail model because the semi-structured format of the ADE would better approximate usual care diagnoses than would structured research interviews.

Another difference between the MINI and ADE diagnoses related to comorbid SUD. We included only current SUD ascertained by the MINI in the gold standard model because a current SUD would be more reliably detected in routine clinical assessments than a past SUD, and is more clinically relevant to bipolar outcomes. However the ADE, used in the limited-detail model, did not distinguish between current versus past SUD.

The “gold-standard” model included some baseline non-time varying information not routinely found in administrative data and so was excluded from the limited-detail model. These characteristics were: clinical history variables (history of psychosis, legal problems or violence, suicide attempt), the Bipolarity Index (BPI),¹⁹ and a baseline Global Assessment of Functioning (GAF).²⁰ The BPI is composed of five categories that describe the characteristics of prior bipolar episodes, age of onset, illness course, treatment response and family history. The score ranges from 0–100, with higher scores indicating characteristics most consistent with bipolar disorder. The GAF is a commonly used measure of clinical symptoms and functioning and ranges from 0 to 100. Higher GAF scores indicate fewer symptoms/higher functioning.

Finally, time, measured as days since baseline, was included as a continuous explanatory variable in both models to capture longitudinal trends. We also included time in both models to permit non-linear trends in the log-odds of remission.

Statistical Analysis

Our analyses involved several steps. We first multiply imputed missing covariates so that our findings are based on combined multiply imputed data sets. We conducted bivariate tests to compare remission rates at baseline and the 12-month follow-up by sample characteristic. These included chi-square tests among categorical variables and Wilcoxon rank sum tests for continuous variable. Next, we estimated the gold-standard model using longitudinal random-effects regression models. Developmental and validation samples were created, model fit and accuracy indices were computed, and then a final model was estimated by recombining the developmental and validation samples. We next estimated a random-effects regression model using the variables containing limited information. We then compared the predictive accuracy of the gold standard and limited-detail models.

Missing Data

We examined the frequency of missing responses at baseline and at each quarterly follow-up. Observations measured at baseline and quarterly visits were used to predict remission based on the composite outcome of MADRS<5 and YMRS<4 at approximately one year post study entry. Because some explanatory variables had missing values, we used multiple imputation techniques to fill in missing values for baseline covariates. This was accomplished using the SAS (version 9.2) macro, “IVEWare,” that uses sequential imputation techniques.²¹ No values were imputed for missing outcomes (i.e., for those lost to follow-up). Five “completed” data sets were generated for analyses. All analyses were based on combining test statistics or regression estimates from the five completed data sets using SAS procedure MIanalyze.

Model Development and Analysis

Gold Standard Model—We randomly split one of the completed datasets 80%–20% into “developmental” and “validation” data sets respectfully. We used the developmental data set to build a predictive but parsimonious “gold-standard” model. We included both random intercepts and time (measured in days) effects for each subject to capture heterogeneity in patient baseline rates and in remission trajectories. We refined the model through elimination of predictors not statistically significant at the $\alpha=0.05$ level. Because we did not want to develop an overly complex model that would exaggerate its accuracy in future studies, we fit the gold-standard model to the validation sample to calculate overfit indices.²² Once we were satisfied with model fit, we estimated the random-effects regression models to the combined validation and developmental samples.

To estimate the prediction accuracy of the models, we categorized the predicted remission rates into deciles, and compared the observed and predicted remission rates within each decile. Observed and predicted rates that are close within a decile indicate that the model is highly predictive for that particular range. Finally, we calculated areas under the ROC curve (AUCs).

Limited-Detail Model—We modified the variables in the gold-standard model to approximate variables typically found in administrative data. We compared the gold-standard and limited-detail models by testing for differences in the AUCs for the combined

developmental and validation samples. This test accounted for the correlation induced by the same subjects appearing in the gold-standard and limited-detail models.

Additional Analyses

We repeated the above steps of calculating the AUCs, comparing observed and predicted outcomes, and estimating overfit indices to the alternative, less stringent clinical outcome that included mild symptoms (i.e., MDRS<10, YMRS<7). Lastly, because comorbid anxiety, ADHD and SUD were determined via clinical interview in the limited-detail data model but via the structured MINI in the gold-standard model, we also examined the degree of concordance between the two diagnostic methods.

Results

Among the 3,168 participants, representing 9,181 person-observations, missing item response rates ranged from 0% to 3.2%. There was significant attrition of the study population over the course of the year with only 17.2% of the subjects completing follow-up 4 (Table 1). Approximately 26% were in remission at the 12-month follow-up compared to 12.3% at baseline. Comorbid mental health diagnoses frequencies varied based on whether they were ascertained via the ADE or MINI. The prevalence of comorbid anxiety disorders was similar across the two diagnostic methods (e.g., 33.6% versus 37.9% at baseline). However, ADE interviews by the treating psychiatrists yielded half as many ADHD diagnoses (4.6% versus 9.0% at baseline) and nearly double the SUD diagnoses (35.2% versus 16.9% at baseline) as the MINI interview conducted by the clinical research specialist. Bivariate analyses of these two diagnostic methods found 80.1% concordance for anxiety disorders, 73.2% for SUD, and 93.5% for ADHD.

In bivariate analyses, variables suggestive of more complicated or severe illness (e.g., psychosis, legal problems/violence, history of suicide attempt, lower GAF score), or the presence of comorbid medical or mental health conditions, were associated with a lower probability of remission at baseline (Table 2). While this finding was not consistently observed at Follow-up 4 (i.e., the 12-month follow-up) there were considerably fewer people at the 12-month follow-up, making it difficult to draw firm conclusions. Also notable is that the prescribing of adjunctive medications (e.g., antidepressants, benzodiazepines) was associated with a lower probability of remission at both baseline and Follow-up 4, suggesting that these patients likely had a more complicated and treatment-resistant course.

In the random-effects logistic regression analyses, the gold-standard model indicated no over-fitting (estimated intercept = .062, estimated slope = .936). In the gold-standard model, a mood stabilizer prescription during at least 80% of the quarter was not associated with euthymia (estimate = .102[s.e. = .089], p. value = .25) but was in the “limited detail” model (.174[.086], p. = .044). Mood stabilizer adherence was highly significant in both models. (p. < .0001) (Appendix 1). Both the gold-standard and limited-detail models had excellent ability to discriminate between remission and non-remission at baseline (gold standard model AUC [95% CI] = .91[.89-.93]; limited-detail model AUC = .92[.91-.93]) and at Follow-up 4 (gold standard model AUC = .89[.86-.92]; limited-detail model AUC = .89[.86-.91]) (Figures 1a and 1b). The correlated ROC test indicated no statistical difference in accuracy between the

gold-standard and limited-detail models (Baseline $p=.34$; Follow-up 4 $p=.47$). The models also discriminated well in important clinical sub-populations (Appendix 2).

Additional Analyses

Our AUC results were similar when we repeated these analyses, predicting a less stringent remission definition that included mild symptoms (i.e., MADRS <10 and YMRS <7).

Finally, comparison of the observed and predicted values of remission (Figures 2a-b) indicated that at baseline both the gold-standard and limited-detail models exhibited poor agreement between observed and predicted values for remission, typically over-predicting remission when the actual observed remission rate was low, and under-predicting when the observed rate was high. However, both models did well (and similarly) in predicting remission after longitudinal follow-up of approximately 1 year (Follow-up 4).

Discussion

This is the first study to demonstrate that clinically limited information can be used to accurately predict bipolar outcomes as compared to a model containing detailed clinical data. From a health policy perspective, this is reassuring, given that data sets with limited clinical detail are more financially feasible for studying outcomes in systems of care than clinical registry data.

The limited-detail model included patient and treatment information typically available in administrative data (e.g., age, race, gender, comorbid psychiatric and medical conditions, visits and medications prescribed and adherence). Notably, in the limited-data model both prescribing a mood stabilizing/antimanic agent and patient adherence were significantly associated with euthymia. The corresponding administrative data variable would be medication possession ratios (MPRs). Our findings are consistent with other research using administrative data that found improved bipolar outcomes with MPR 80%.^{23,24} Thus, our results provide additional support for medication possession as a process quality measure associated with improved outcomes in observational study.

Model accuracy was similar for the gold-standard and limited-detail models. However, the sub-sample comorbidity populations differ between the two models (due to diagnostic discordance between them) and the AUC confidence intervals were large due to small sample sizes in the subpopulations, particularly by Follow-up 4. Both model types adequately discriminated outcomes, despite that the populations identified as having a comorbid condition were not always the same. This suggests that comorbid diagnoses based on a less structured interview method, which is more similar to how diagnoses are often ascertained in usual care, distinguished between patient subtypes with clinically significant comorbid symptoms that are also predictive of outcome.

There are several limitations to consider when interpreting these results. There was significant loss to follow-up of the study subjects in the four quarterly assessments. We assumed that this loss to follow-up was missing at random after conditioning on the demographic, clinical and treatment data although it is possible that this is not the case. If

not missing at random, our predicted outcomes are subject to a selection bias, although the direction of this bias is unknown. Despite the high attrition rate, the prediction accuracies of our models are high. Moreover, this attrition rate is not unlike that observed in usual care; therefore these predicted outcomes may reasonably mirror what one would see in administrative data of continued longitudinal follow-up.

Although the STEP-BD observational arm was designed to include a broader, more generalizable population than is usually seen in clinical trials, and treatment choices were not constrained, STEP-BD patients and treatment may differ from those in usual care populations in a variety of ways. For example, STEP-BD's use of a structured interview to determine the bipolar diagnosis may mean that STEP-BD bipolar diagnoses are more accurate than what would be seen in usual care settings. That said, bipolar disorder cohorts established from usual care administrative data have been found to be highly accurate when compared to chart review.²⁵

Additionally, STEP-BD may be more accurate in detecting comorbid conditions than what is typically seen in usual care. For example, we compared the prevalence of comorbid SUD as defined by the "limited detail" model in STEP-BD bipolar-I participants with those in prior administrative data-based studies.^{26–29} STEP-BD treating clinicians identified a higher prevalence of SUD (privately insured=34%, Medicaid insured=38%) than the administrative data studies (privately insured 3% and 8%; Medicaid 24%). Given prior literature on the prevalence of comorbid SUD in bipolar disorder outpatient populations,^{30,31} it appears that comorbid SUD was better detected in STEP-BD than the administrative data studies cited above. This observation could have implications for the accuracy of predicting bipolar outcomes using usual care data if important comorbid conditions that can affect bipolar outcomes (such as SUD) are not accurately ascertained or recorded in usual care.

These limitations highlight that while our study shows promise for the comparative accuracy of limited-detail clinical data versus more detailed clinical registry data for predicting bipolar outcomes, an important next step would be to apply similarly limited data to a truly usual care database that includes patient outcomes.

In conclusion, using a large multi-site contemporary cohort of bipolar patients, this is the first study demonstrating that it is feasible to discriminate well between bipolar patients who remit from those who do not using data that are limited in clinical detail. Further, predicting accurate remission rates for bipolar disorder patients occurring approximately one year of follow-up are possible. Additional study is needed to further refine the precision of predicting bipolar outcomes (particularly at baseline), and to test if the results are consistent with true usual care data. Nonetheless, these results are promising for the potential to predict population wide outcomes using routinely collected data that are limited in their clinical detail. This study informs the discussion on the feasibility and potential accuracy of collecting limited clinical detail in electronic data for routine quality assessment purposes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding :The authors gratefully acknowledge the following funding agencies: the National Institute of Mental Health K01MH071714 (Busch, Zelevinsky) and R01MH61434 (Neelon and Normand) and RC4MH092717-01(Busch, He). An earlier version of this analysis was presented at the 2010 AcademyHealth Annual Meeting in Boston, MA.

References

1. Health at a Glance 2009: OECD Indicators. Organization for Economic Cooperation and Development; 2009.
2. Conway PH, Clancy C. Transformation of health care at the front line. *JAMA*. 2009; 301(7):763–765. [PubMed: 19224753]
3. Girosi, F.; Meili, R.; Scoville, R. Extrapolating evidence of health information technology savings and costs. RAND Corporation; 2005.
4. American Recovery and Reinvestment Act of 2009, (2009).
5. Department of Health and Human Services. [Accessed March 2, 2011] <http://www.hhs.gov/recovery/>
6. American Psychiatric Association. Practice guideline for the treatment of patients with bipolar disorder. *American Journal of Psychiatry*. Dec.1994 151(12, Suppl):40. [PubMed: 8267133]
7. American Psychiatric Association. Practice guidelines for the treatment of patients with bipolar disorder (revision). *American Journal of Psychiatry*. 2002; 159 Supplement(4):1–50.
8. Coryell W, Scheftner W, Keller M, Endicott J, Maser J, Klerman GL. The enduring psychosocial consequences of mania and depression. *American Journal of Psychiatry*. 1993; 150(5):720–727. [PubMed: 8480816]
9. Goodwin, FK.; Jamison, KR., editors. Manic-Depressive Illness: Bipolar disorders and recurrent depression. 2. New York: Oxford University Press; 2007.
10. Judd LL, Akiskal HS, Schettler PJ, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Archives of General Psychiatry*. 2002; 59(6):530–537. [PubMed: 12044195]
11. Kessler RC, Chiu WT, Demler O, Walters E. Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry*. 2005; 62(6):617–627. [PubMed: 15939839]
12. 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*. 2003; 53(11):1028–1042. [PubMed: 12788248]
13. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*. 1998; 59(Suppl 20):22–57. [PubMed: 9881538]
14. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: Reliability, validity, and sensitivity. *Br J Psychiatry*. 1978; 133:429–435. [PubMed: 728692]
15. Montgomery A. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979; 134:382–389. [PubMed: 444788]
16. Berk M, Ng F, Wang WV, et al. The empirical redefinition of the psychometric criteria for remission in bipolar disorder. *J Affect Disord*. 2008; 106:153–158. [PubMed: 17655936]
17. Chengappa KNR, Baker RW, Shao L, et al. Rates of response, euthymia and remission in two placebo-controlled olanzapine trials for bipolar mania. *Bipolar Disord*. 2003; 5:1–5. [PubMed: 12656931]
18. Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery–Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *Journal of Psychiatric Research*. 2004; 38:577–582. [PubMed: 15458853]

19. Sachs GS. Strategies for improving treatment of bipolar disorder: integration of measurement and management. *Acta Psychiatrica Scandinavica*. 2004; 110(Suppl 422):7–17. [PubMed: 15330934]
20. Jones SH, Thornicroft G, Coffey M, Dunn G. A brief mental health outcome scale: reliability and validity. *British Journal of Psychology*. 1995; 166:654–659.
21. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*. 2001; 27(1):85–95.
22. Harrell, FE. Regression modeling strategies: with applications to linear models, logistic regression and survival analysis. New York: Springer-Verlag New York, Inc; 2001.
23. Svarstad BL, Shireman TI, Sweeney J. Using drug claims data to assess the relationship of medication adherence with hospitalization and costs. *Psychiatric Services*. Jun; 2001 52(6):805–811. [PubMed: 11376229]
24. Sajatovic M, Valenstein M, Blow F, Ganoczy D, Ignacio R. Treatment adherence with lithium and anticonvulsant medications among patients with bipolar disorder. *Psychiatr Serv*. 2007; 58(6):855–863. [PubMed: 17535948]
25. Unutzer J, Simon G, Pabiniak C, Bond K, Katon W. The use of administrative data to assess quality of care for bipolar disorder in a large staff model HMO. *General Hospital Psychiatry*. 2000; 22(1):1–10. [PubMed: 10715498]
26. American Medical Association. Current Procedural Terminology. AMA Press; 2005.
27. Busch AB, Frank RG, Sachs G. Bipolar-I depression outpatient treatment quality and costs in usual care practice. *Psychopharmacology Bulletin*. 2008; 41(2):24–39. [PubMed: 18668015]
28. Busch AB, Lehman AF, Goldman HH, Frank RG. Changes over time and disparities in schizophrenia treatment quality. *Medical Care*. 2008 in press.
29. Busch AB, Ling D, Frank RG, Greenfield SF. Changes in bipolar-I disorder quality of care during the 1990's. *Psychiatric Services*. 2007; 58(1):27–33. [PubMed: 17215409]
30. Chengappa KNR, Levine J, Gershon S, Kupfer DJ. Lifetime prevalence of substance or alcohol abuse and dependence among subjects with bipolar I and II disorders in a voluntary registry. *Bipolar Disorders*. 2000; 2(3):191–195. [PubMed: 11256686]
31. McElroy SL, Altshuler LL, Suppes TS, et al. Axis I Psychiatric Comorbidity and its relationship to historical illness variables in 288 patients with bipolar disorder. *Am J Psychiatry*. 2001; 158:420–426. [PubMed: 11229983]

1a: Gold standard models **1b: Limited-detail models** **Figure 1.**

Figures 1a & 1b: ROC curves for remission at baseline and follow-up 4 for gold-standard and administrative models.

2a: Gold standard model **2b: Limited-detail models**

Figure 2.
 Figures 2a & b: Observed versus predicted deciles of remission

Author Manuscript

Author Manuscript

Characteristics included in model		Baseline		Follow-up 1		Follow-up 2		Follow-up 3		Follow-up 4	
		N	%	N	%	N	%	N	%	N	%
Sample size		3168	100	2428	76.6	1824	57.6	1216	38.4	545	17.2
	x	625	19.7	439	18.1	319	17.5	214	17.6	98	18.0
Clinical characteristics											
	x	2062	65.1	1586	65.3	1203	66.0	787	64.7	362	66.4
	x	671	21.2	520	21.4	392	21.5	277	22.8	106	19.4
	x	435	13.7	322	13.3	229	12.6	152	12.5	77	14.1
	x	186	5.9	144	5.9	100	5.5	63	5.2	32	5.9
	x	1194	37.7	935	38.5	717	39.3	475	39.1	228	41.8
	x	1065	33.6	809	33.3	604	33.1	392	32.2	177	32.5
	x	1160	36.6	859	35.4	639	35.0	415	34.1	186	34.1
	x	72.2 [15.9]		72.1 [15.7]		72.5 [15.5]		72.5 [15.5]		72.1 [15.4]	
	x	60.0 [12.0]		64.5 [11.4]		65.8 [11.4]		66.9 [11.1]		67.3 [10.7]	
Treatment characteristics											
		2474	78.1	2182	89.9	1719	94.2	1162	95.6	528	96.9
	x			1673	68.9	1238	67.9	830	68.3	386	70.8
	x	2374	74.9	1790	73.7	1362	74.7	908	74.7	407	74.7
	x	1434	45.3	1138	46.9	1004	55.0	740	60.9	361	66.2
	x	871	27.5	687	28.3	632	34.6	467	38.4	222	40.7
	x	3.11 [5.3]		3.32 [5.8]		3.3 [5.2]		3.3 [6.3]		3.1 [4.9]	
Diagnoses Defined by MINI											
	x	1201	37.9	919	37.9	682	37.4	445	36.6	206	37.8
	x	284	9.0	211	8.7	148	8.1	91	7.5	52	9.5
	x	534	16.9	380	15.7	264	14.5	162	13.3	61	11.2
Diagnoses Defined by Clinical Interview in ADE											
	x	1064	33.6	824	34.0	630	34.5	438	36.0	199	36.5
	x	145	4.6	109	4.5	75	4.1	49	4.0	28	5.1

		Baseline		Follow-up 1		Follow-up 2		Follow-up 3		Follow-up 4	
Characteristics included in model		N	%	N	%	N	%	N	%	N	%
		Sample size									
	x	3168	100	2428	76.6	1824	57.6	1216	38.4	545	17.2
		1114	35.2	892	36.7	669	36.7	443	36.4	192	35.2

¹ LD= Limited-detail model

² Antimanic medication: lithium, valproate, carbamazepine, lamotrigine or antipsychotic.

³ Adherence defined from information in the Care Monitoring Form (CMF) as: percent of CMF visits where patient reports missing any antimanic medications. If at least 3 CMF visits observed, adherence = missing antimanic dosages 25% of CMF visits; if 1 – 3 CMF visits, adherence = missing antimanic dosages 33% of CMF visits.

⁵ GS = Gold-standard model;

Table 2

Remission rates by sample characteristics for study subjects, at baseline and follow-up 4 (approximately 12 months).

	Baseline			P	Follow-up 4				P
	Over all N	Full Remission ¹	%		Overall N	Full Remission ¹	N	%	
Characteristic (person level)									
Study quarter									
	3168	389	12.3	<.0001	545	144	26.4	<.0001	
Study entry year									
2000	540	99	18.3	<.0001	125	42	33.6	.08	
2001	773	109	14.1		147	38	25.9		
2002	739	69	9.3		154	33	21.4		
2003	692	68	9.8		106	25	23.6		
2004	424	44	10.4		13	6	46.2		
Demographic characteristics									
Age categories at study start									
Age 16–34	1201	138	11.5	.29	155	47	30.3	.19	
Age 35–64	1967	251	12.8		390	97	24.9		
Gender									
Male	1324	185	14.0	.01	230	59	25.7	.73	
Female	1844	204	11.1		315	85	27.0		
Race/ethnicity									
Black	168	18	10.7	.14	14	6	42.9	.30	
White	2724	342	12.6		494	127	25.7		
Hispanic	166	12	7.2		22	5	22.7		
Other	110	17	15.5		15	6	40.0		
Insurance									
Private	1970	274	13.9	.005	346	90	26.0	.11	
Medicare only	305	27	8.9		75	16	21.3		
Medicaid only	189	13	6.9		18	2	11.1		
Both	78	9	11.5		8	4	50.0		

	Baseline			Follow-up 4		
	Over all N	Full Remission ¹	P	Overall N	Full Remission ¹	P
None	625	66	10.6	98	32	32.7
Clinical characteristics						
Bipolar disorder category						
Bipolar-I disorder	2062	277	13.4	362	108	29.8
Bipolar-II disorder	671	89	13.3	106	22	20.8
Other bipolar spectrum disorders	435	23	5.3	77	14	18.2
Comorbid seizure disorder						
Yes	186	11	5.9	32	6	18.8
No	2881	369	12.8	503	136	27.0
History of psychosis						
Yes	1194	172	14.4	228	73	32.0
No	1880	205	10.9	314	71	22.6
History of legal problems/violence						
Yes	1065	98	9.2	177	43	24.3
No	2011	286	14.2	363	99	27.3
History of suicide attempt						
Yes	1160	95	8.2	186	42	22.6
No	1940	290	15.0	349	99	28.4
BPI						
0-25	5	1	20.0	0	0	0
26-50	352	22	6.3	59	10	17.0
51-75	1489	153	10.3	262	63	24.0
76-100	1260	207	16.4	220	70	31.8
Admission GAF						
0-70	2672	183	6.9	344	56	18.8
71-100	473	205	43.3	185	85	46.0
Treatment characteristics						
Antimanic medication prescribed ³						

	Baseline			Follow-up 4		
	Over all N	Full Remission ¹	P	Overall N	Full Remission ¹	P
Yes	2474	338	13.7	528	139	26.3
No	694	51	7.4	17	5	29.4
Antimanic medication prescribed for at least 80% of follow-up						
Yes				386	107	27.7
No				159	37	23.3
Patient adherent with antimanic medication ^{3,4}						
Yes	2374	318	13.4	407	118	29.0
No	794	71	8.9	138	26	18.8
Antidepressant prescribed ⁵						
Yes	1434	159	11.1	361	80	22.2
No	1734	230	13.3	184	64	34.8
Benzodiazepine/sedative prescribed ⁵						
Yes	871	70	8.0	222	48	21.6
No	2297	319	13.9	323	96	29.7
Mental health visits during the past quarter (mean/s.d.)						
Diagnostic Variables Defined by MINI						
Comorbid anxiety disorder						
Yes	1201	52	4.3	206	37	18.0
No	1954	335	17.1	337	106	31.5
Comorbid attention deficit hyperactivity disorder						
Yes	284	10	3.5	52	7	13.5
No	2875	377	13.1	491	136	27.7
Comorbid substance use disorder						
Yes	534	30	5.6	61	16	26.2
No	2627	358	13.6	483	128	26.5
Diagnostic Variables Defined by Clinical Interview in ADE						
Comorbid anxiety disorder						
Yes	1064	67	6.3	199	45	22.6
						.13

	Baseline			P	Follow-up 4			P
	Over all N	Full Remission ¹			Overall N	Full Remission ¹		
No	2104	322	15.3		346	99	28.6	
Comorbid attention deficit hyperactivity disorder								
Yes	145	11	7.6	.08	28	6	21.4	.54
No	3023	378	12.5		517	138	26.7	
Comorbid substance use disorder								
Yes	1114	101	9.1	<.0001	192	49	25.5	.72
No	2054	288	14.0		353	95	26.9	

³Antimanic medication: lithium, valproate, carbamazepine, lamotrigine or antipsychotic.

⁴In Care Monitoring Form (CMF), patient defined as non-adherent if psychiatrist recorded patient missed mgs of antimanic medication at least 25% (if CMF >3 visits) or 33% of time (if CMF = 1–3 visits). If patient prescribed >1 antimanic medication, both required to have mgs missed to count as “non-adherent”.

⁵At any time during baseline or study follow-up.

Imputed results not included in table but all p. values adjusted for imputation.