

## ORIGINAL ARTICLE

## Data-driven classification of bipolar I disorder from longitudinal course of mood

AL Cochran<sup>1</sup>, MG McInnis<sup>2</sup> and DB Forger<sup>1,3</sup>

The Diagnostic and Statistical Manual of Mental Disorder (DSM) classification of bipolar disorder defines categories to reflect common understanding of mood symptoms rather than scientific evidence. This work aimed to determine whether bipolar I can be objectively classified from longitudinal mood data and whether resulting classes have clinical associations. Bayesian nonparametric hierarchical models with latent classes and patient-specific models of mood are fit to data from Longitudinal Interval Follow-up Evaluations (LIFE) of bipolar I patients ( $N=209$ ). Classes are tested for clinical associations. No classes are justified using the time course of DSM-IV mood states. Three classes are justified using the course of subsyndromal mood symptoms. Classes differed in attempted suicides ( $P=0.017$ ), disability status ( $P=0.012$ ) and chronicity of affective symptoms ( $P=0.009$ ). Thus, bipolar I disorder can be objectively classified from mood course, and individuals in the resulting classes share clinical features. Data-driven classification from mood course could be used to enrich sample populations for pharmacological and etiological studies.

*Translational Psychiatry* (2016) 6, e912; doi:10.1038/tp.2016.166; published online 11 October 2016

## INTRODUCTION

Bipolar disorder (BP) is characterized by profound pathological mood swings from mania to depression, interspersed with periods of subsyndromal symptoms (that is, symptoms that are insufficient to constitute a major mood episode). The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV and now DSM-5) enables a common clinical language between practitioners, but is yet to anchor an understanding of disease etiology of BP.

Longitudinal course of mood is the defining feature of BPI, determining how patients are diagnosed and subtyped. We propose that BPI patients can be objectively classified based solely on their longitudinal course of mood states. A data-driven approach is likely to better represent the random processes behind transitions between mood states and, in doing so, help identify classes of clinical significance.<sup>1–7</sup>

Here, we introduce a methodology that uses mood data to objectively identify new classes within DSM categories such as BP. It consists of descriptions of (i) a single patient's mood as a random process and (ii) BPI patients as a Bayesian nonparametric hierarchical model with latent classes. We fit these models to self-reported longitudinal mood data from BPI patients ( $N=209$ ) tracked over one or more years to determine whether the data substantiate BPI classes.

## MATERIALS AND METHODS

## Data

Data from BPI patients ( $N=209$ ) were drawn from the Prechter Longitudinal Study of Bipolar Disorder at the University of Michigan.<sup>8</sup> Patients were evaluated using the Longitudinal Interval Follow-up Evaluation (LIFE)<sup>9</sup> administered by a clinician involving retrospective analysis of symptoms over the past 1–3 years. Patients were included provided they had a diagnosis of BPI and at least one LIFE. Patients were rated from 0 to 3 weekly for depression, mania and hypomania. A rating of zero was given

if they had no history of symptoms in the respective category, one if they had a history of symptoms but no symptoms for the given week, two if they had one or more symptoms for the respective category (mania, hypomania and depression) and three if they met full DSM-IV criteria. Serial assessments were combined, resulting in 1–8 years of observation for study patients (see Supplementary Figure 1). Of the 209 patients, 141 were female, 179 patients were white, 10 black or African-American, 1 American Indian or Alaskan Native, 1 Asian, 11 more than one race and 7 patients of unknown race; 197 patients were not-Hispanic, 5 Hispanic and 7 of unknown ethnicity. Patients were an average  $40.3 \pm 12.2$  ( $\pm$  s.d.) years of age and had an average  $15.6 \pm 3.1$  ( $\pm$  s.d.) years of education at initial interview for the Prechter study. Years of education was unknown for two patients. The University of Michigan's Biomedical Institutional Review Board approved all recruitment, assessment and research procedures for the main study HUM00000606. Patients provided written informed consent after receiving a complete description of the study.

## Patient model of mood

Mood was modeled as a discrete-time Markov Chain with finite states (DTMC). The finite states represented mood states (defined below). Central to DTMC are one-step transition probabilities, which define the probability that a patient observed in one-mood state is in a certain mood state at the next observation time. The underlying assumption of DTMC (the Markov assumption) is that a patient's state depends on its history only through their current mood state. Time between observations represented 1 week with respect to LIFE.

We defined mood states in three ways using (i) full DSM-IV criteria, (ii) subsyndromal symptoms and (iii) subsyndromal symptoms with mixed states (Table 1). First, a patient was considered manic (M) if they met full DSM-IV criteria for mania or depressive (D) if they met the full DSM-IV criteria for depression but not mania. A non-episode (nEp) state was defined as the absence of meeting criteria for either mania or depression. Second, a patient was considered in a subsyndromal state of mania (m) if manic/hypomanic symptoms (one or more) were present and, if not present, they were considered in a subsyndromal state of depression (d) if depressive symptoms were present. The euthymic state (Eu) was defined as the absence of m or d. Lastly, we included mixed states, where a patient

<sup>1</sup>Department of Mathematics, University of Michigan, Ann Arbor, MI, USA; <sup>2</sup>Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA and <sup>3</sup>Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. Correspondence: Professor DB Forger, Department of Mathematics, University of Michigan, 530 Church Street, Ann Arbor, MI 48105, USA.

E-mail: forger@umich.edu

Received 19 May 2016; accepted 1 June 2016

**Table 1.** Criteria from LIFE data for defining mood states in three models

Model	Mood states	Criteria based on LIFE
DSM-IV model	Mania (M)	3 On mania scale
	Depression (D)	0–2 On mania scale; 3 on depression scale
	Non-episodic (nEp)	0–2 On mania scale; 0–2 on depression scale
Subsyndromal model	Subsyndromal mania (m)	2–3 On mania scale or 2–3 on hypomania scale
	Subsyndromal depression (d)	0–1 On mania scale; 0–1 on hypomania scale; 2–3 on depression scale
	Euthymia (Eu)	0–1 On mania scale; 0–1 on hypomania scale; 0–1 on depression scale
Subsyndromal—mixed state model	Subsyndromal mania (m)	2–3 On mania or 2–3 on hypomania scale; 0–1 on depression scale
	Subsyndromal depression (d)	0–1 On mania scale; 0–1 on hypomania scales; 2–3 on depression scale
	Euthymia (Eu)	0–1 On mania scale; 0–1 on hypomania scale; 0–1 on depression scale
	Subsyndromal mixed state (md)	2–3 On mania or 2–3 on hypomania scale; 2–3 on depression scale

Abbreviations: 0, no history of symptoms; 1, no symptoms over past week; 2, one or more symptoms over past week; 3, symptoms meet full DSM-IV criteria over past week; DSM, Diagnostic and Statistical Manual of Mental Disorder; LIFE, Longitudinal Interval Follow-Up Evaluations.

was considered in a mixed state (md) if they reported subsyndromal symptoms for both depression and mania/hypomania.

### Bayesian nonparametric hierarchical model

Patient models of mood were embedded into a Bayesian nonparametric hierarchical model.<sup>10</sup> A hierarchical model uses other patients' data to improve inferences on an individual. A nonparametric hierarchical model allows for a variable number of classes (with high probability) and are used when it is important to determine the number of classes.

The following hierarchy was evaluated:

BipolarI → Class → Patient → Mood,

which involves patient-specific mood dynamics. This hierarchy was found to be a better model, as measured using information criteria, than a hierarchy that has bipolar I, class, and mood levels, but no patient level (see Supplementary Methods).

Patient data were collected in a matrix  $Y$  counting week-to-week transitions between mood states (see Supplementary Methods). A patient's mood is governed by a transition probability matrix  $X$ , which along with the DTMC formulation specifies the probability of observing data  $Y$  given  $X$ . The patient's transition matrix  $X$  was modeled as a Bayesian nonparametric hierarchical model.<sup>10</sup> The model assumes that a patient belongs to one of the classes. Transition probability matrices for patients in a class are identically distributed, leading to similar mood dynamics for a class. Probabilities of belonging to a class and class-specific parameters are drawn from an approximate Dirichlet process.<sup>10</sup>

### Model fitting

Markov Chain Monte Carlo estimated the posterior distribution of parameters. Markov Chain Monte Carlo is an iterative algorithm generating a sequence of samples that approximate a target probability distribution. Following a particular blocked Gibbs algorithm,<sup>11</sup> each iteration updated sequentially (i) patient transition matrices, (ii) class parameters, (iii) patient assignments, (iv) probabilities of belonging to a class and (v) additional parameters. Gibbs's sampling was used when possible (i, iii, iv); otherwise slice sampling was used (ii,v; Supplementary Methods). To determine whether the results generalize to a larger population, threefold cross-validation was performed in which the population was randomly divided into three equal-sized groups and the analysis repeated three times with one of the groups omitted in each case.

### Post analysis

Classes were evaluated for over/under-representation of categorical variables and for differences in the means of certain covariate variables (complete list of variables is in Supplementary Tables 1–3). All variables were collected on year 5 of the Prechter study, resulting in the exclusion of 13 of the original 209 patients from post analysis for joining the study after year 5. In SPSS, significant associations were evaluated using cross-tabulation with a Person's  $\chi^2$ -test for categorical variables and one-way analysis of variance for covariates. Homogeneity of variances was tested

using a Levene test and, when violated, analysis of variance was replaced by a Welch test. *Post hoc* analysis was performed using standardized Pearson residuals for categorical variables; a Tukey's test for covariates when homogeneity of variances was not violated; and a Games–Howell test for covariates when homogeneity of variances was violated. Significance was considered an alpha level of 0.05. Significance was not adjusted for multiple testing.

### Code availability

Analysis was performed in Matlab (Mathworks, Natick, MA, USA). Code has been made available at <https://sites.google.com/site/amylouissecochran/code>.

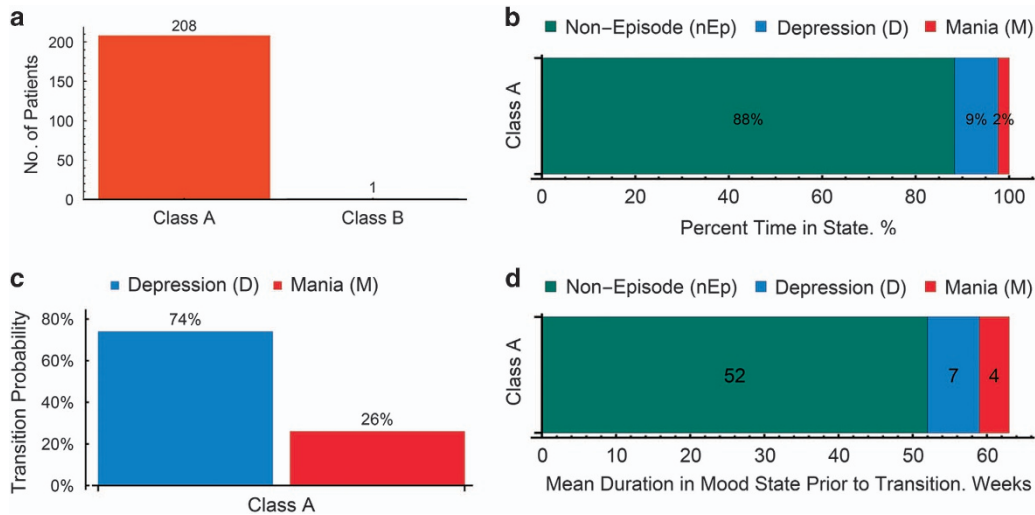
## RESULTS

### No classes identified using DSM-IV categories

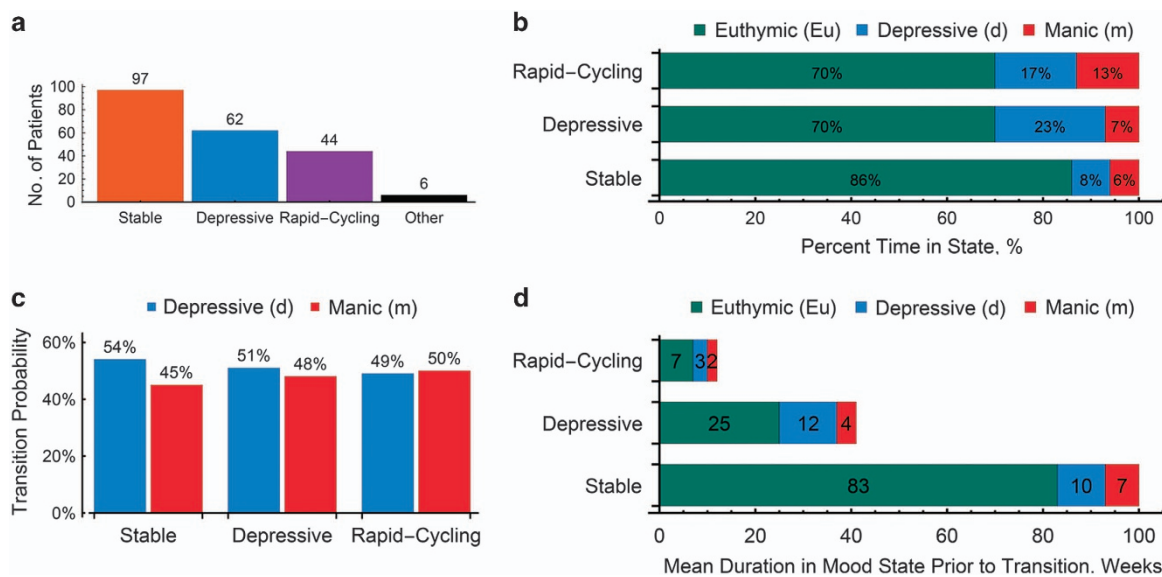
When mood states were defined using DSM-IV criteria, 208 of the 209 patients were assigned the same class, with assignments based on which class he/she was most likely to belong (Figure 1a). The lone patient was atypical for having over 15 episodes of each of mania (M) and depression (D) in 2 years. The large class provides expectations for any patient's illness course. We define a 'typical' patient as one whose parameters take estimated averages for the class, where averages are taken with respect to the estimated posterior distribution. A typical patient spends ~88% of their time in nEp, followed by 9% of their time in the depression state D and 2% in the mania state M (Figure 1b; see Supplementary Table 1 for average transition probability matrix of the class). The propensity toward the depression state D over the mania state M is attributed to both a greater chance of transitioning from the non-episodic state nEp to the depression state D (74% versus 26%) and longer episodes of the depression state D (7 versus 4 weeks). For comparison, the non-episodic state nEp lasts ~52 weeks. Threefold cross-validation also revealed only one large class for each run made up of 139 of 139 patients in the run, 138 of 139 in the second run and 138 of 140 in the third run.

### Distinct classes in the subsyndrome

Using partial criteria (that is, subsyndromal states m,d and Eu), three classes are identified, labeled as 'stable', 'depressive' and 'rapid cycling' in a decreasing order of size (Figure 2a). The stable class consists of an estimated  $42 \pm 7\%$  ( $\pm$ s.d.) of the BPI population; 97 patients in our study are classified as the stable class. The depressive class has  $28 \pm 7\%$  ( $\pm$ s.d.) of the BPI population and 62 study patients, whereas the rapid-cycling class has  $25 \pm 5\%$  ( $\pm$ s.d.) of the BPI population and 44 study patients. Six patients were not assigned to one of the three classes. Threefold



**Figure 1.** No classes in Diagnostic and Statistical Manual of Mental Disorder (DSM) syndrome. In **a**, every patient except one is assigned to the same class when using the DSM model of mood states. The average parameters for the main class can be used to establish expectations for **(b)** the percent time in a particular state, **(c)** the relative transition probability into mania, M, or depression, D, from a non-episodic state (nEp); and **(d)** the mean duration of mood state before transition to another state. The outlying patient is atypical for short episodes.



**Figure 2.** Three classes in subsyndrome. In **a**, three classes emerge when using the subsyndromal model of mood states. **(a)** A typical patient for each class differs in **(b)** the percent time in a particular state, **(c)** the relative transition probability into a manic state, m, or depressive state, d, from euthymia, Eu; and **(d)** the mean duration of mood state before transition to another state. The three classes were labeled ‘stable’, ‘depressive’ and ‘rapid cycling’ to reflect the percent time in a state and the mean duration of mood states. Subsyndromal mood states, m and d, define mood symptoms that meet partial or *full* Diagnostic and Statistical Manual of Mental Disorder (DSM) criteria.

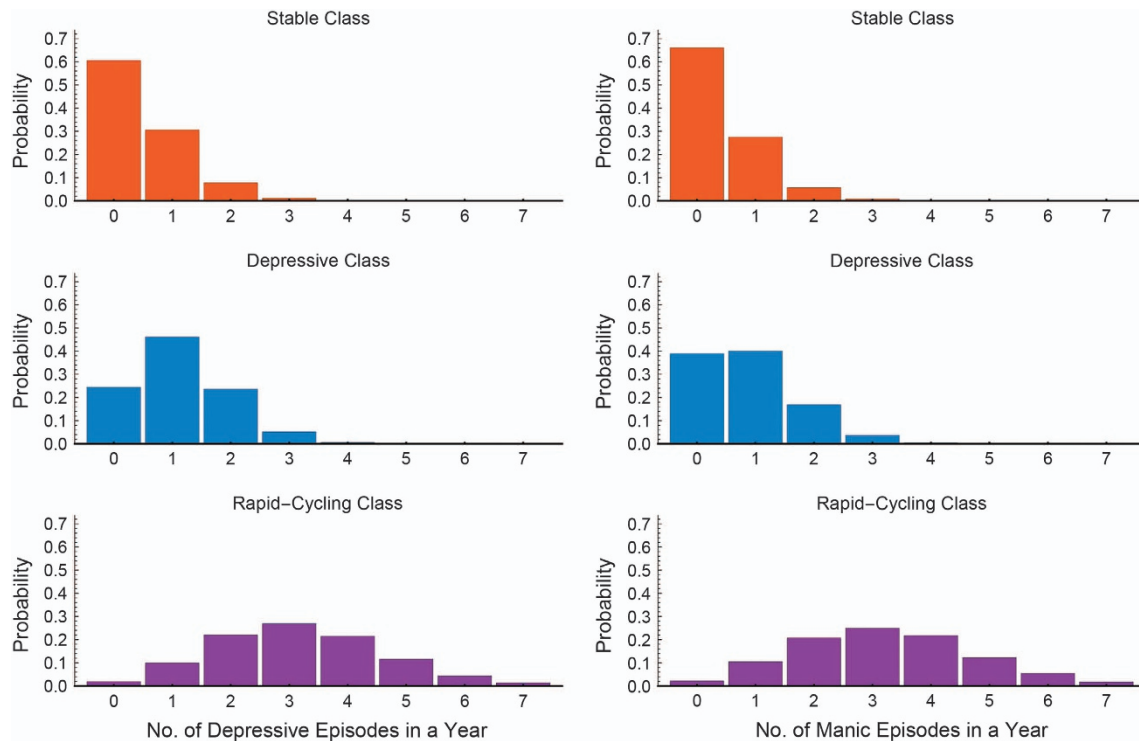
cross-validation revealed three classes in two of the three runs, and two classes in the remaining run. When allowing for a mixed state, we still identified three classes with 93% of patients separated into the same classes (Supplementary Figure 2). We focus on the classes from subsyndromal model without a mixed state.

**Description of subsyndromal classes**

The stable class is labeled ‘stable’ for spending the most time in the euthymic state Eu. A typical stable patient spends 86% of their time in the euthymic state Eu and only 8% of their time in the depressive state d and 6% in the manic state m (Figure 2b; see

Supplementary Table 1 for average transition probability matrices of each class). They are more likely to transition to the depressive state d over the mania state m (54% versus 46%) after the Eu (Figure 2c) and spend an average of 83 weeks in the Eu, 10 weeks in the depressive state d and 7 weeks in the manic state m before transitioning to another state (Figure 2d).

Compared with the stable class, the depressive class is labeled ‘depressive’ because a typical patient spends 23% of their time in the depressive state d and only 70% in the Eu (leaving 7% of their time in the manic state m). They last an average of 12 weeks in the depressive state d, 25 weeks in the Eu and 4 weeks in the manic state m before a transition to another state. Similar to the stable



**Figure 3.** Number of subsyndromal mood episodes in a year. A typical patient in the rapid-cycling class is more likely to have more subsyndromal episodes in a year than typical patients in the stable and depressive classes. However, if classes were determined based on number of episodes, patients will always be misclassified. For example, a typical patient from each class has a significant probability of having exactly one subsyndromal episode in a year.

class, they also are more likely to transition to the depressive state *d* over the manic state *m* (54% versus 46%) after being in the euthymic state *Eu*.

The final class is ‘rapid-cycling’ because a typical patient has frequent transitions between states. A typical patient lasts an average of only 7 weeks in the euthymic state *Eu* compared with 3 weeks in the depressive state *d* and 2 weeks in the manic state *m* before a transition to another state. An estimated 13% of their time is spent in the manic state *m*, the most of the three classes. However, they spend a similar fraction of time in the euthymic state *Eu* (70%) to the depressive class, leaving 17% of their time in the depressive state *d*. They were least likely to transition to the depressive state *d* over the manic state *m* (49% versus 51%) after in the euthymic state *Eu*.

Each class is defined based on transition probabilities rather than the number of episodes in a year. Hence, there is always a chance that a particular patient in the rapid-cycling class has less subsyndromal mood episodes in a year than a patient in another classes, even though on average the rapid-cycling class is characterized by more subsyndromal mood episodes in a year (see Figure 3). In other words, the number of episodes in a year cannot alone determine classes. For example, all three classes have a significant probability of having one subsyndromal mood episode.

Classes revealed from threefold cross-validation could also be distinguished based on whether patients were relatively more stable or not, and whether patients were relatively more depressive or not (Supplementary Figure 3).

#### Comorbid clinical phenotypes of subsyndromal classes

Depressive and rapid-cycling classes have a worse clinical outcome than the stable class (Figure 4). Suicide attempts were significantly different between classes ( $P=0.017$ ), with suicide attempts more frequent in the depressive class and less in the

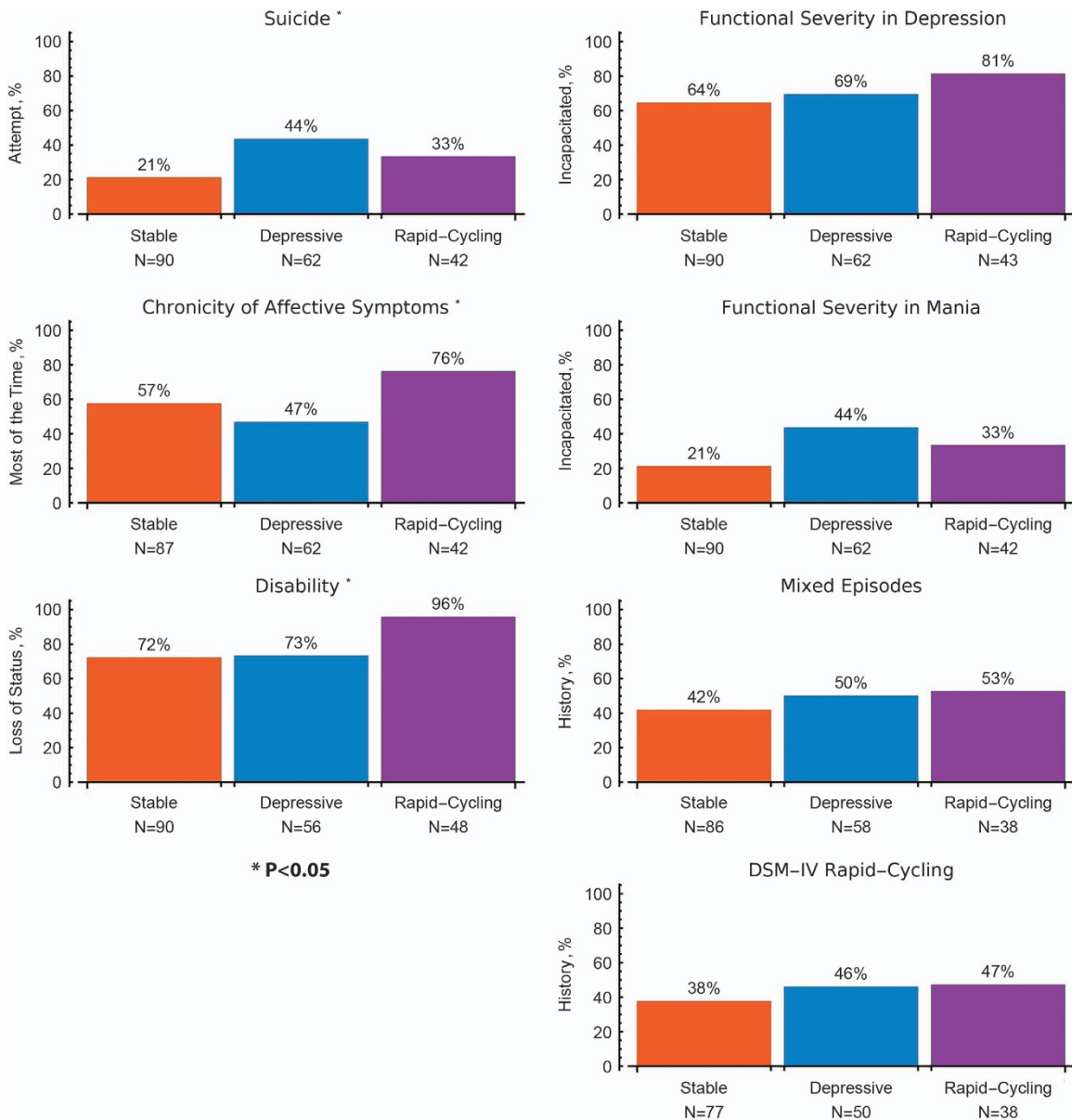
stable class. Classes differed in overall chronicity (mood symptoms most of the time or otherwise;  $P=0.009$ ) and general impairment (no loss of status or otherwise;  $P=0.012$ ), with increased chronicity and general impairment in the rapid-cycling class and decreased chronicity in the depressive class. Classes did not differ with respect to rapid cycling (history or no history), functionality during depression or mania (incapacitated or otherwise) and mixed episodes (history or no history; Supplementary Table 2).

On the Hamilton Rating Scale for Depression, classes differed significantly in the mean scores ( $P=0.014$  from Welch test; homogeneity of variance violated). The rapid-cycling class had a significantly higher mean score than the depressive class. On the Young Mania Rating Score, classes differed significantly in the mean scores ( $P=0.006$  from Welch test; homogeneity of variance violated) and baseline score ( $P=0.002$  from Welch test; homogeneity of variance violated). The rapid-cycling class had a higher mean score on the Young Mania Rating Score than the depressive class; the depressive class had a lower baseline score than both the depressive and rapid-cycling classes. Classes did not differ in age of onset and number of depressive, manic and hypomanic episodes (Supplementary Tables 3 and 4).

Classes were not significantly different with respect to sex ( $P=0.242$ ) or age at baseline ( $P=0.052$ ); the depressive class was slightly younger at baseline. Type of baseline medication (antidepressant, anti-epileptic, antipsychotic, benzodiazepine, hypnotic, lithium and stimulant) did not differ between the classes.

#### DISCUSSION

We have demonstrated the utility of objectively classifying bipolar I patients from course of mood. Our method consists of statistically fitting mood data to a Bayesian nonparametric hierarchy with latent states and patient-specific mood dynamics.



**Figure 4.** Clinical relevance of subsyndromal classes. Over and under-representation of specific patient classes in the subsyndromal classes. Suicide attempts are more common in the depressive class, whereas chronicity of affective symptoms and general impairment is more common in the rapid-cycling class.

Our methodology was applied to 209 bipolar I patients, identifying three classes of similar size when mood states were defined using partial DSM-IV criteria. No classes were identified when relying on strict DSM-IV criteria. Each class shared clinical features, for example, the ‘depressive’ class had higher rates of suicide attempts.

#### Implication

Current classification often relies on categories whose distinctions are not supported scientifically.<sup>12,13</sup> There was debate surrounding the development of DSM-5 about how to weigh scientific evidence against historical practice in establishing definitions.<sup>12</sup> Arguments against historical definitions are that disease causes remain largely unknown and boundaries between categories are more fluid than is recognized. The NIMH even proposed a system to research mental illness, the Research Domain Criteria, which

emphasizes measurements from genomics to self-reports over DSM categories.<sup>14</sup>

The present study presents a method to objectively classify BPI patients using longitudinal course of mood. Our results show that data-driven classification from mood course is not only viable, but can divide patients into groups with shared clinical features and can agree with clinical intuition, for example, certain patients are often considered to be a more ‘depressive’ subtype. The presented method can be applied to other recurrent mood disorders and other data sets of longitudinal course of mood.

Because data-driven classes emerge based solely on data, rather than expert consensus, they may better reflect etiological differences in what causes mood transitions. For example, genetic studies have looked to better subtype bipolar patients, as patient groups that are defined using imprecise or broad criteria can weaken the statistical inferences that can be obtained.<sup>15,16</sup> Moreover, when compared with other criteria for data-driven

classification,<sup>17,18</sup> longitudinal course of mood states has the unique advantage that it defines BP patients.

Data-driven classes may also better reflect etiological differences between patients because they emerge only when substantiated by data and cannot be determined by other means, for example, using age, sex or baseline medication. Note classes were not revealed using full DSM criteria and our 'rapid-cycling' class could not be determined from the number of subsyndromal mood episodes or a patient's history of rapid cycling. Certain DSM boundaries may be unsupported by data. For example, rapid cycling based on DSM-5 criteria is defined as at least four episodes in 1 year. Kupka *et al.*<sup>19</sup> find factors (for example, gender) that are increasingly associated with the number of episodes. They and others argue that there is no empirical basis for separating patients on number of episodes.<sup>20,21</sup>

#### Comparison with longitudinal studies

Identifying only subsyndromal classes supports monitoring symptoms weaker in severity, number and/or duration.<sup>22,23</sup> Stronger criteria can focus clinical efforts on critical patients, but weaker criteria provide more information over time. For example, major episodes are rare and longer observation may be needed to classify patients from major episodes alone. In other words, subsyndromal symptoms offer extensive clinical information.

Judd *et al.*<sup>24</sup> and Paykel *et al.*<sup>25</sup> also find that weaker symptoms predominate severe symptoms, depressive symptoms predominate hypomanic/manic symptoms and moods change frequently in severity and polarity. All three results are consistent with current findings. In addition, they find that BPI patients spend significant time with mood symptoms, for example, 47.3% of the time in Judd *et al.*<sup>24</sup> and 53% of the time in Paykel *et al.*<sup>25</sup> Depending on the class, a typical patient in our study spends ~12–32% of their time without these symptoms. Differences in chronicity may reflect discrepancies in how mood states are defined between the studies, sample populations or treatment.

#### Methodological comparison

Our classification method has advantages over structural equation modeling (for example, latent class or latent growth analysis), the current approach to subtype and classify BP and other recurrent mood disorders.<sup>17,18</sup> First, we used a random model of course of symptoms to better represent the unpredictable (and nonlinear) nature of mood states. Second, we used a *nonparametric* hierarchical model to objectively determine the number of patient classes. Structural equation models are typically *parametric* in that the number of classes are specified for each model.

Random models of mood had previously been fitted to data for major depression,<sup>1,2</sup> BP<sup>3–5</sup> and youth BP.<sup>5,7</sup> For example, van der Werf *et al.*<sup>1</sup> fit a random model to describe time-to-recovery from depressive episodes for a cohort of depressive patient. They present a patient-specific model, but do not fit it to data. Bonsall *et al.*<sup>4</sup> fit modifications of autoregressive models and extended in Moore *et al.*<sup>3,5</sup> The theses of Lopez<sup>6</sup> and Fan<sup>7</sup> are more similar to the present study in their use of DTMC and in inferring patient clusters using Bayesian techniques. They find clusters in youth BP, but do not allow mood dynamics to be patient-specific. Patient-specific models outperformed class-specific models in this study.

#### Limitations

The data used in this study were weekly retrospective data based on the LIFE. Retrospective data can be less accurate than prospective data because memory can be unreliable and dependent on mood. Despite its limitations, retrospective data have one significant advantage: enough data could be collected in one visit to classify a patient. Future work will need to test data-

driven classification using prospective data, with consideration for length of observation needed to classify. Weekly data have the disadvantage that they do not capture all the transitions within the week. Mood in BP will often shift dramatically in a matter of days. It may be that the inclusion of daily information could better define classes of clinical relevance.

Another limitation of our study is the use of relatively small ( $N=209$ ) and homogeneous (mostly female Caucasian) population. Therefore, whereas the utility of data-driven classification can be established, the exact classes identified may not generalize to other populations. To explore this generalizability, our analysis was repeated on subsamples of our data set using a threefold cross-validation. This additional analysis also found no classes when using DSM-IV states and two or three classes when using subsyndromal states with classes separated based on whether individuals were more or less stable and more or less depressive.

We also did not identify a mechanism to explain why individuals have different longitudinal patterns of mood, although results suggest that these differences are probably not explained by baseline medication, gender, age or rapid cycling. Identifying a mechanism may require exploring connections between classes and complete medical history that includes hospitalizations, medication response and comorbidities. This exploration, in turn, may also elucidate whether data-driven classes could be used to improve treatment.

Lastly, our model makes assumptions to simplify the analysis that may affect the utility of the model. For example, mood is divided into three or four states in the model, whereas mood may be better described with additional states or as continuous. In addition, the model does not capture factors that could cause changes in longitudinal patterns, such as life events and changes in medication, among others. This assumption and others (such as the Markov assumption) will need to be considered in the future.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGMENTS

This research was supported by Heinz C Prechter Bipolar Research Fund at the University of Michigan Depression Center; the Richard Tam Foundation; and Human Frontiers of Science Program Grant RPG 24/2012.

#### REFERENCES

- 1 van der Werf SY, Kaptein KI, de Jonge P, Spijker J, de Graaf R, Korf J. Major depressive episodes and random mood. *Arch Gen Psychiatry* 2006; **63**: 509–518.
- 2 Kaptein KI, De Jonge P, Korf J, Spijker J, De Graaf R, Van Der Werf SY. Random-mood interpretation of determinants for major depression. *Psychol Med* 2007; **37**: 1261–1271.
- 3 Moore PJ, Little MA, McSharry PE, Geddes JR, Goodwin GM. Forecasting depression in bipolar disorder. *IEEE Trans Bio-med Eng* 2012; **59**: 2801–2807.
- 4 Bonsall MB, Wallace-Hadrill SM, Geddes JR, Goodwin GM, Holmes EA. Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. *R Soc Proc Biol Sci* 2012; **279**: 916–924.
- 5 Moore PJ, Little MA, McSharry PE, Goodwin GM, Geddes JR. Mood dynamics in bipolar disorder. *Int J Bipolar Disord* 2014; **2**: 11.
- 6 Lopez A. Markov models for longitudinal course of youth bipolar disorder, Doctoral Dissertation, University of Pittsburgh, 2008.
- 7 Fan J. On Markov and hidden Markov models with applications to trajectories, Doctoral Dissertation, University of Pittsburgh 2014.
- 8 Langenecker SA, Ryan KA, Marshall DM, Kamali M, McClinnis MG. Longitudinal stability of intermediate cognitive phenotypes in bipolar disorder. *Biol Psychiatry* 2015; **77**: 319s–320ss.
- 9 Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, Mcdonaldscott P *et al.* The longitudinal interval follow-up evaluation - a comprehensive method for assessing outcome in prospective longitudinal-studies. *Arch Gen Psychiatry* 1987; **44**: 540–548.

- 10 Ishwaran H, Zarepour M. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* 2000; **87**: 371–390.
- 11 Ishwaran H, James LF. Gibbs sampling methods for stick-breaking priors. *J Am Stat Assoc* 2001; **96**: 161–173.
- 12 Kendler KS. A history of the DSM-5 Scientific Review Committee. *Psychol Med* 2013; **43**: 1793–1800.
- 13 Kendler KS. The transformation of American psychiatric nosology at the dawn of the twentieth century. *Mol Psychiatry* 2016; **21**: 152–158.
- 14 Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K *et al*. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010; **167**: 748–751.
- 15 MacQueen GM, Hajek T, Alda M. The phenotypes of bipolar disorder: relevance for genetic investigations. *Mol Psychiatry* 2005; **10**: 811–826.
- 16 Ivleva EI, Morris DW, Moates AF, Suppes T, Thaker GK, Tamminga CA. Genetics and intermediate phenotypes of the schizophrenia–bipolar disorder boundary. *Neurosci Biobehav Rev* 2010; **34**: 897–921.
- 17 Musliner KL, Munk-Olsen T, Laursen TM, Eaton WW, Zandi PP, Mortensen PB. Heterogeneity in 10-year course trajectories of moderate to severe major depressive disorder: a danish national register-based study. *JAMA Psychiatry* 2016; **73**: 346–353.
- 18 Monahan PO, Stump T, Coryell WH, Harezlak J, Marcoulides GA, Liu H *et al*. Confirmatory test of two factors and four subtypes of bipolar disorder based on lifetime psychiatric co-morbidity. *Psychol Med* 2015; **45**: 2181–2196.
- 19 Kupka RW, Luckenbaugh DA, Post RM, Suppes T, Altshuler LL, Keck PE Jr. *et al*. Comparison of rapid-cycling and non-rapid-cycling bipolar disorder based on prospective mood ratings in 539 outpatients. *Am J Psychiatry* 2005; **162**: 1273–1280.
- 20 Schneck CD, Miklowitz DJ, Miyahara S, Araga M, Wisniewski S, Gyulai L *et al*. The prospective course of rapid-cycling bipolar disorder: findings from the STEP-BD. *Am J Psychiatry* 2008; **165**: 370–377.
- 21 Bauer M, Beaulieu S, Dunner DL, Lafer B, Kupka R. Rapid cycling bipolar disorder—diagnostic concepts. *Bipolar Disord* 2008; **10**(1 Pt 2): 153–162.
- 22 Marangell LB. The importance of subsyndromal symptoms in bipolar disorder. *J Clin Psychiatry* 2004; **65**(Suppl 10): 24–27.
- 23 Bauer M, Glenn T, Grof P, Schmid R, Pfennig A, Whybrow PC. Subsyndromal mood symptoms: a useful concept for maintenance studies of bipolar disorder? *Psychopathology* 2010; **43**: 1–7.
- 24 Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA *et al*. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry* 2002; **59**: 530–537.
- 25 Paykel ES, Abbott R, Morriss R, Hayhurst H, Scott J. Sub-syndromal and syndromal symptoms in the longitudinal course of bipolar disorder. *Br J Psychiatry* 2006; **189**: 118–123.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

© The Author(s) 2016

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)