Figure S1: ICCBD Diagnostician Review Criteria for Diagnosis & Confidence: Instructions to Reviewers

Step 1: Determine Diagnosis[1]:
- B = Bipolar Disorder
- NB = Not Bipolar Disorder
- NI = Not Enough Information/Not Classifiable
  - ❒ Internally inconsistent information
  - ❒ No actual diagnosis mentioned (i.e. "rule out…")
  - ❒ No strong evidence for any diagnosis

Step 2: Assign Confidence:
A. Bipolar Disorder
- 0 = Low Confidence
  - ❒ Diagnostic label of Bipolar Disorder with no evidence inconsistent with the diagnosis
- 1 = Intermediate Confidence
  - ❒ Diagnostic label of Bipolar Disorder

    AND

  - ❒ Documentation of a distinct period of abnormally and persistently elevated expansive or irritable mood

    AND

  - ❒ During the period of mood disturbance, ≤ 2 of the following symptoms (≤ 3 if the mood is irritable) and have been documented
    - ▪ Inflated self-esteem or grandiosity
    - ▪ Decreased need for sleep
    - ▪ More talkative than usual or pressure to keep talking
    - ▪ Flight of ideas or subjective experience that thoughts are racing
    - ▪ Distractibility
    - ▪ Increase in goal-directed activity or psychomotor agitation
    - ▪ Excessive involvement in pleasurable activities that have a high potential for painful consequences
- 2 = High Confidence
  - ❒ Diagnostic label of Bipolar Disorder

    AND

  - ❒ Documentation of prior or current distinct period of abnormally and persistently elevated expansive or irritable mood which includes documentation of ≥ 3 of the following symptoms (≥ 4 if the mood is irritable)
    - ▪ Inflated self-esteem or grandiosity
    - ▪ Decreased need for sleep
    - ▪ More talkative than usual or pressure to keep talking
    - ▪ Flight of ideas or subjective experience that thoughts are racing
    - ▪ Distractibility

---

[1] Diagnosis is based on a consensus clinical impression among at least 3 clinicians based on all available information, including family history, medication, and course.

- - Increase in goal-directed activity or psychomotor agitation
  - Excessive involvement in pleasurable activities that have a high potential for painful consequences
  - *AND*
  - ❒ Not better accounted for by substance use and/or general medical condition

General Note
When assigning a confidence rating of 1 or 2 for a case diagnosed with Bipolar Disorder, consider the implications of including this patient as a case in a genetic study of Bipolar Disorder.

Format for Diagnosis & Confidence:
- Diagnosis Code-Confidence Code
- Examples:
  - ➢ B-1 (indicating Bipolar Diagnosis with Intermediate Confidence)
  - ➢ NB-2 (indicating Not Bipolar with High Confidence)
  - ➢ NI (Indicating that there is not enough information to make a diagnosis)

Data S1: Diagnosis, Drug, and Comorbidity Code Definitions

CODE DEFINITIONS FOR DIAGNOSES

Bipolar disorder (BP) diagnosis

Code	Description
ICD9 296.0* ICD9 296.1*	Billing code for manic disorder
ICD9 296.4* ICD9 296.5* ICD9 296.6*	Billing code for bipolar affective disorder
ICD9 296.7*	Billing code for bipolar I disorder
ICD9 296.8*	Billing code for bipolar NOS
LPA44 LPB1598 ODA:YJHP1-1	Bipolar in EMR problem list

Major Depression (MDD) diagnosis

Code	Description
ICD9 296.2* ICD9 296.3*	Billing code for major depressive disorder
LPB101 LPB1593 ODA:YJSN1	Major depression in EMR problem list

Schizophrenia (Scz) diagnosis

Code	Description
ICD9 295.*	Billing code for schizophrenia
ODA:YJCM1 ODA:YJBG1 LPA487	Schizophrenia in EMR problem list

Organic Affective Disorder (OAS) diagnosis

Code	Description
ICD9 295.83	Billing code for Organic Affective Disorder

CODE DEFINITIONS FOR DRUGS

Lithium (Li) exposure

Code	Description
MDId 00061	Lithium (Eskalith, Lithobid) Dispensed by inpatient pharmacy or prescribed in EMR
LOINC 14334-7	Therapeutic blood lithium level

Valproic Acid (VPA) exposure

Code	Description
MDId 00083	Valproic Acid (Depakote, Depacon, Depakene) Dispensed by inpatient pharmacy or prescribed in EMR
MDId 03383	Divalproex sodium Dispensed by inpatient pharmacy or prescribed in EMR
LOINC 4086-5	Therapeutic blood valproic acid level

Secondary bipolar medication exposure

Code	Description
MDId 03180	Risperidone (Risperdal) Dispensed by inpatient pharmacy or prescribed in EMR
MDId 04050	Olanzapine (Zyprexa) Dispensed by inpatient pharmacy or prescribed in EMR
MDId 04220	Quetiapine (Seqouel) Dispensed by inpatient pharmacy or prescribed in EMR
MDId 04747	Ziprasidone (Geodon) Dispensed by inpatient pharmacy or prescribed in EMR
MDId 04825	Aripiprazole (Abilify) Dispensed by inpatient pharmacy or prescribed in EMR
MDId 00058	Carbamazepine (Tegretol) Dispensed by inpatient pharmacy or prescribed in EMR
MDId 03809	Lamotrigine (Lamictal, Lamotrizine) Dispensed by inpatient pharmacy or prescribed in EMR

CODE DEFINITIONS FOR COMORBIDITIES

Substance Abuse diagnosis

Code	Description
ICD9 291.*	Alcoholic Psychoses
ICD9 292.*	Drug Psychoses
ICD9 303.0* ICD9 303.9*	Alcohol dependence syndrome
ICD9 304.*	Drug dependence syndrome
ICD9 305.0* ICD9 305.2* ICD9 305.3* ICD9 305.4* ICD9 305.5* ICD9 305.6* ICD9 305.7* ICD9 305.8* ICD9 305.9*	Nondependent abuse of drugs

+ codes indicating patient is in remission are excluded
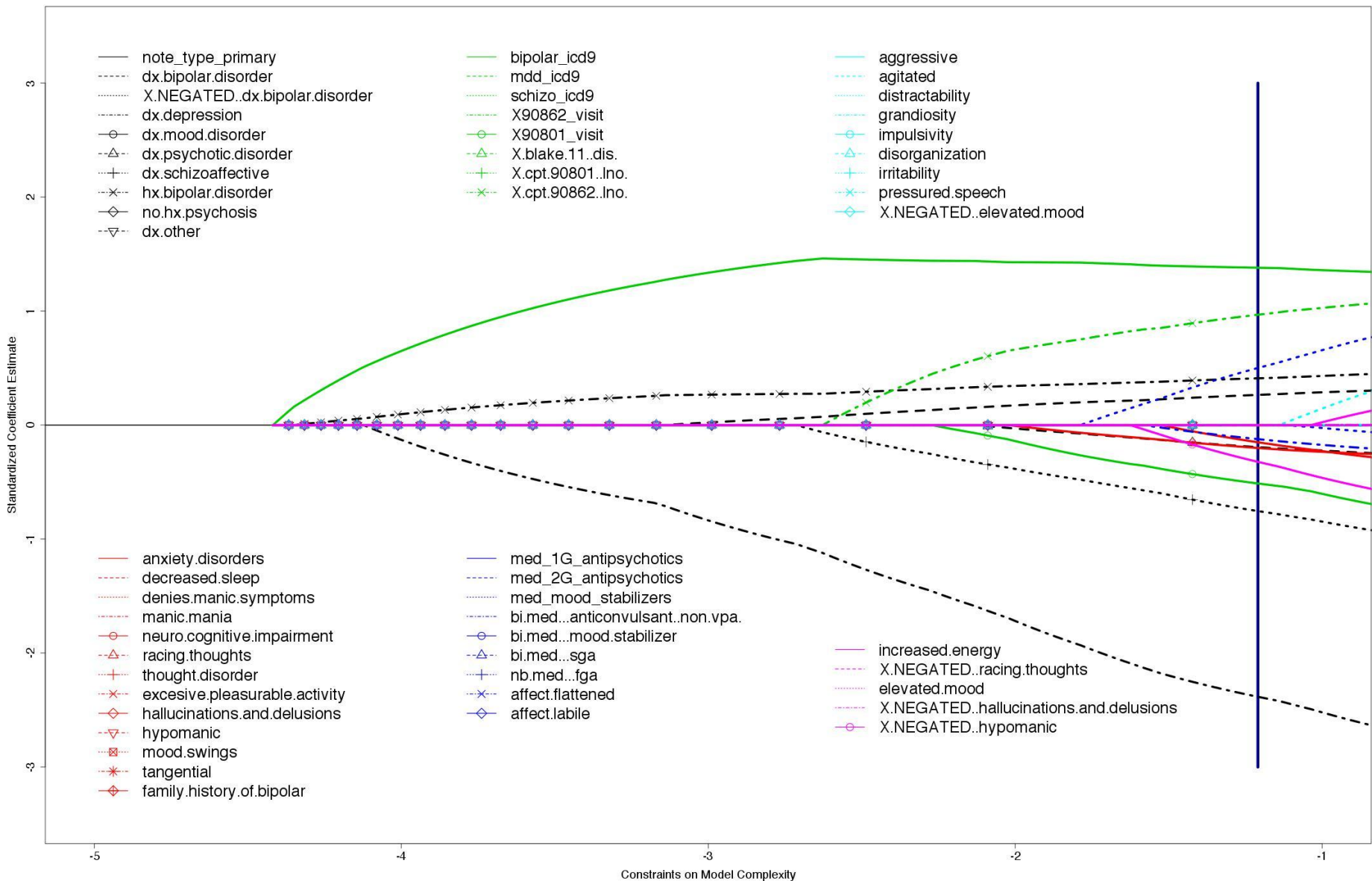
Data S2: NLP Feature Regular Expressions

See Table S1 in the online data supplement.

File description:

Field	Description
Id	identifier of regular expression
Type	feature is consistent with BD (CW_Bipolar) or inconsistent with BD diagnosis (IW_Bipolar)
Category	NLP feature category (corresponds to variables used to train 95-NLP Model)
Negatable	Feature is negatable (1= Yes, 0 = No)
Regex	Regular expression used to match this feature

Figure S2: Solution path for adaptive LASSO procedure.

Each curve represents the coefficient estimate of a feature for a given value of penalty parameter, where model complexity is constrained more for lower values of the x axis.

The vertical line at -1.2 represents the optimal penalty parameter corresponding to the final model selected based on the Bayesian Information Criteria.