BIPOLAR MIXED STATES SYNTHETIC DATASET

Version 1.0 | Executive Summary

Metric	Value
Total Patients	800
Total Observations	5,550
Clinical Variables	35
Privacy (k-anonymity)	k = 12
Differential Privacy	ε < 0.8

MentalData.io *— Research-Grade Synthetic Mental Health Data*

1. Dataset Overview

The Bipolar Mixed States Synthetic Dataset is a high-fidelity, privacy-preserving longitudinal dataset modeling bipolar disorder with mixed features (ICD-10: F31.6x). This dataset enables researchers to develop and validate ML/AI models, clinical decision support systems, and educational tools without privacy concerns.

1.1 Clinical Scope

The dataset captures the full spectrum of bipolar mixed states across seven ICD-10 subcategories, from mild episodes (F31.61) to severe with psychotic features (F31.64). Each patient trajectory includes seven longitudinal visits, enabling treatment response modeling and outcome prediction research.

ICD-10 Code	Description	Patients
F31.60	Mixed Episode, Unspecified	250 (31.5%)
F31.61	Mixed Episode, Mild	100 (12.3%)
F31.62	Mixed Episode, Moderate	150 (18.9%)
F31.63	Mixed Episode, Severe	100 (12.3%)
F31.64	Severe w/ Psychotic Features	100 (12.6%)
F31.75	In Partial/Full Remission	50 (6.1%)
F31.89	Other Specified Bipolar	50 (6.2%)

Table 1: Diagnosis Distribution (N = 800 patients)

2. Key Clinical Features

2.1 Clinical Rating Scales

Scale	Mean ± SD	Range	Clinical Interpretation
YMRS	29.6 ± 12.3	0–47	Moderate manic symptoms
HAM-D	18.1 ± 11.7	0–39	Mild-moderate depression
MADRS	27.2 ± 18.7	0–50	Moderate depression
GAF	44.6 ± 21.4	17–89	Moderate impairment
PHQ-9	15.0 ± 4.9	0–26	Moderately severe

Table 2: Clinical Scale Summary Statistics

2.2 Unique Psychological Phenomena

This dataset uniquely captures underrepresented psychological patterns in bipolar mixed states that are clinically significant but rarely available in existing datasets:

Feature	Prevalence	Clinical Significance
Post-Diagnosis Identity Crisis	38.8%	Patients questioning past decisions
Sleep Aversion in Mania	65.1%	"Sleep is wasting time" cognition
Stimulant Misuse Pattern	39.2%	Self-medication with stimulants
Polypharmacy Conflict	27.9%	Simultaneous stimulant + sedative use

Table 3: Unique Psychological Features

3. Privacy & Quality Guarantees

The dataset employs multiple layers of privacy protection to ensure no individual could ever be re-identified while maintaining high statistical fidelity for research applications:

Metric	Value	Target	Status
k-Anonymity	k = 12	k ≥ 12	✓ PASS
Differential Privacy	ε = 0.78	ε < 1.0	✓ PASS
MIA Resistance	AUC = 0.52	AUC < 0.55	✓ PASS
Statistical Fidelity	87%	> 80%	✓ PASS
Clinical Validity	90/100	> 80	✓ PASS

Table 4: Privacy & Quality Metrics

4. Recommended Use Cases

Use Case	Suitability	Notes
ML/AI Development	✓ Recommended	Train mood prediction, treatment response models
Clinical Decision Support	✓ Recommended	Prototype CDS systems for bipolar management
Educational Purposes	✓ Recommended	Teaching psychiatric assessment methods
Research Publications	✓ w/ Caveats	Document synthetic nature, cite properly
Clinical Practice	✗ Not Recommended	Not validated for real patient decisions

Table 5: Use Case Recommendations

5. Licensing & Contact

License: CC BY-NC 4.0 (Attribution-NonCommercial)

Website: https://mentaldata.io

Email: contact@mentaldata.io

For commercial licensing inquiries or custom dataset requests, please contact us directly.