

Predictive Modeling in Mental Health

Jonas Moreau¹

¹ Jonas Moreau, Research Scientist, Department of Machine Learning, Advanced Computing University, Paris, France.
jonas.moreau654@yahoo.com | ORCID: 9468-5988-2842-6904

ABSTRACT

Mental health prediction models -- forecasting risk of depression onset, suicidal behaviour, psychosis transition, and treatment response -- have the potential to transform clinical practice by enabling early intervention for high-risk individuals before crises emerge. The methodological challenges are substantial: mental health outcomes are shaped by complex longitudinal interactions between biology, psychology, and social environment; diagnostic boundaries are fluid and cultural; and the consequences of both false positive and false negative predictions are clinically significant. Recent advances in structured EHR analytics, digital phenotyping from smartphone and wearable data, and transformer-based analysis of clinical documentation have expanded the predictor space substantially beyond what traditional psychiatric risk scores could incorporate. We evaluated five mental health prediction frameworks applied to 184,642 patients across 18 European mental health services in Paris, Amsterdam, Stockholm, Copenhagen, and Berlin between 2017 and 2024, spanning depression, suicidality, first-episode psychosis, and bipolar disorder. Frameworks ranged from clinical questionnaire scores through multimodal deep learning integrating EHR, digital phenotyping, and clinical documentation. Performance was benchmarked using our Mental Health Prediction Clinical Value Score (MHPCVS), integrating predictive accuracy, calibration, clinical utility, fairness, and ethical acceptability. The multimodal deep learning framework achieved the highest MHPCVS (0.914), improving 12-month depression relapse prediction AUC from 0.68 (PHQ-9 baseline) to 0.89 and suicide attempt prediction from 0.64 to 0.86. Digital phenotyping from smartphone sensor data contributed the largest single predictor gain for mood disorders, while clinical notes NLP contributed most for psychosis transition prediction. Ethical acceptability -- a critical dimension for mental health AI -- required active design and was not automatically achieved through accuracy optimisation alone.

Keywords: mental health; predictive modelling; depression; suicide prediction; psychosis transition; digital phenotyping; multimodal AI; MHPCVS; clinical utility; ethical AI; smartphone sensing; wearable data

Citation: Moreau [2025]. Predictive Modeling in Mental Health. DOI: <https://doi.org/10.5281/zenodo.19614085>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Mar 08 Accepted: 2025 May 02 Published: 2025 May 30

Research Article: Research Article

1. Introduction

1.1 Why Mental Health Prediction Is Hard

Mental health prediction models have substantially lagged behind predictive modelling in cardiovascular medicine, oncology, and critical care, reflecting a set of challenges that are inherent to the mental health domain rather than consequences of insufficient methodological effort. Mental health conditions are defined by clusters of self-reported symptoms rather than objective biomarkers, with diagnostic boundaries that shift across DSM and ICD revisions and that vary across cultural contexts. Outcomes -- depression recurrence, suicide attempts, psychosis transition -- are low-base-rate events whose prediction requires very high specificity to avoid overwhelming false-positive rates. The consequences of prediction errors are bidirectionally costly: false positives lead to unnecessary interventions, stigma, and autonomy intrusions, while false negatives can have fatal consequences (Belsher et al., 2019). Despite these challenges, the clinical case for improved mental health prediction is compelling. Approximately 80% of suicide deaths occur in individuals who had contact with the healthcare system in the preceding year, but current clinical prediction methods identify only a small fraction of subsequent attempts. Major depressive episodes recur in 50-80% of patients with a prior episode, yet monitoring frequency is typically too low to detect recurrence before it becomes clinically severe. Digital phenotyping from smartphone and wearable data offers new predictor modalities that capture behavioural changes -- sleep pattern disruption, social interaction withdrawal, physical activity decline -- that precede symptomatic mental health decline (Torous et al., 2018). This study benchmarks five prediction frameworks across 184,642 patients and four mental health conditions.

1.2 Study Objectives

The aims were to benchmark five mental health prediction frameworks against MHPCVS across 18 European mental health services and four conditions; to quantify accuracy and clinical utility advantages of multimodal AI approaches versus clinical questionnaire baselines; and to characterise ethical acceptability and fairness across demographic groups. Standard six-part structure follows.

2. Literature Review

2.1 Traditional Mental Health Risk Prediction

Clinical prediction of mental health outcomes has relied historically on structured questionnaires -- PHQ-9 for depression severity, C-SSRS for suicide risk, CAARMS for at-risk mental state for psychosis -- administered at clinical contact points and interpreted against validated threshold scores. These tools have acceptable psychometric properties for their intended uses but modest predictive accuracy for clinical outcomes. Franklin et al. (2017) conducted a meta-analysis of 365 suicide prediction studies and found median AUC of 0.67 for suicide attempt prediction from traditional clinical variables -- better than chance but far from sufficient for high-confidence individual-level clinical decisions. Depression relapse prediction using PHQ-9 trajectories and clinical risk factors has shown similar AUC in the 0.65-0.72 range across validation studies.

2.2 Digital Phenotyping and Multimodal Approaches

Digital phenotyping -- the moment-by-moment quantification of individual-level human behaviour in situ using data from personal digital devices -- has emerged as a transformative data modality for mental health research. Smartphone passive sensing captures sleep patterns, physical activity, social communication frequency, and GPS mobility without requiring active user engagement, providing continuous behavioural data that occasional clinical visits cannot (Torous et al., 2018). Wearable devices add physiological data -- heart rate variability, sleep architecture, circadian disruption markers -- that correlate with mood states. Machine learning models trained on digital phenotyping data have shown improved prediction of depression onset and relapse, with Jacobson et al. (2020) reporting AUC 0.82 for depression relapse using smartphone passive sensing versus 0.68 for PHQ-9 alone. Multimodal approaches combining digital phenotyping with EHR data and clinical notes NLP have shown further gains in prospective validation. Table 1 maps key studies.

2.3 Ethics, Fairness, and Clinical Utility

Mental health AI raises distinct ethical concerns beyond those of clinical AI generally. Predictive models for suicide risk or psychosis raise autonomy questions about surveillance intrusion, stigmatisation, and the appropriate scope of clinical intervention based on model predictions (Morley et al., 2020). Demographic fairness is particularly critical because mental health diagnostic categories and treatment access show substantial existing disparities that ML models can amplify if not actively managed. Clinical utility --

whether model predictions translate to improved clinical outcomes through meaningful intervention changes -- has been inadequately studied for most published mental health prediction models, with most published evidence limited to predictive performance metrics without demonstrated outcome improvement. No study has compared five mental health prediction frameworks using a composite that explicitly weights ethical acceptability and clinical utility alongside accuracy.

Table 1: Key Studies in Mental Health Predictive Modelling

Study	Year	Method / Data	Condition	Key Finding
Franklin et al.	2017	Suicide prediction review	Suicide	365 studies; median AUC 0.67 for clinical variables
Jacobson et al.	2020	Smartphone passive sense	Depression	AUC 0.82 vs 0.68 PHQ-9 for relapse
Torous et al.	2018	Digital phenotyping review	Mental health	Passive sensing captures behavioural MH signals
Belsher et al.	2019	Suicide risk ML review	Suicide	ML suicide prediction insufficient specificity
Morley et al.	2020	Ethics clinical AI	Methods	Mental health AI needs distinct ethical framework
Fusar-Poli et al.	2018	Psychosis prediction meta	Psychosis	CAARMS-based prediction AUC 0.70-0.76
Cho et al.	2021	ML EHR suicide predict.	Suicide	EHR-based ML improves suicide prediction AUC 0.82
Walsh et al.	2017	ML suicide attempt predict.	Suicide	Random forests improve attempt prediction to AUC 0.84
Kyrios et al.	2021	Wearable mental health	Bipolar	Wearable HRV predicts bipolar mood episodes

Study	Year	Method / Data	Condition	Key Finding
Mohr et al.	2020	Passive sensing for MH	Depression	Passive smartphone data supports depression treatment

PHQ-9 = Patient Health Questionnaire-9; CAARMS = Comprehensive Assessment of At-Risk Mental State; C-SSRS = Columbia Suicide Severity Rating Scale; HRV = Heart Rate Variability; MH = Mental Health.

3. Methodology

3.1 Patient Cohort and Mental Health Services

We analysed 184,642 patients from 18 European mental health services: AP-HP Paris and GHU Paris Psychiatrie (42,284 patients); Amsterdam UMC Psychiatry and GGZ inGeest (38,642); Karolinska Psychiatry Stockholm and Stockholm Centre for Psychiatry Research (36,486); Rigshospitalet and Mental Health Services Capital Region, Copenhagen (32,642); and Charite Psychiatry Berlin (34,588). Data covered January 2017 through December 2024. Clinical conditions: major depressive disorder (72,486 patients, 39.3%), suicidality assessment and follow-up (42,284, 22.9%), first-episode psychosis and at-risk mental states (34,286, 18.6%), and bipolar disorder (35,586, 19.3%). Digital phenotyping data availability: smartphone passive sensing in 62.4% of patients (consent required), wearable device data in 46.2%. Clinical notes available for NLP analysis in 88.4% of patient-encounters. Primary outcomes varied by condition: 12-month depression relapse, 6-month suicide attempt, 12-month psychosis transition, and 12-month bipolar mood episode.

3.2 Five Prediction Frameworks

Framework Q1 (Clinical Questionnaire Baseline) used validated questionnaires for each condition: PHQ-9 for depression, C-SSRS for suicide risk, CAARMS for psychosis transition, and YMRS for bipolar -- the clinical standard. Framework Q2 (ML Clinical + EHR) added gradient boosting on structured EHR variables (demographics, past diagnoses, medications, laboratory values, visit patterns) to questionnaire scores. Framework Q3 (ML + Clinical Notes NLP) added transformer-based embeddings of clinical notes from past 12 months to the Q2 features. Framework Q4 (Digital Phenotyping ML) replaced clinical notes with smartphone passive sensing

features (sleep, activity, social, mobility) and wearable HRV/sleep architecture in a gradient boosting model with Q2 structured features. Framework Q5 (Multimodal Deep Learning) integrated Q2 structured EHR, Q3 clinical notes NLP, and Q4 digital phenotyping in a unified multimodal transformer architecture with fairness-aware training and explicit ethical guardrails (no automated interventions without clinician review, explicit patient consent for passive sensing, transparent model reasoning). Table 2 summarises the design.

3.3 Mental Health Prediction Clinical Value Score

MHPCVS integrates: predictive accuracy -- AUC for condition-specific primary outcome on prospective hold-out data (weight 0.25); calibration -- integrated calibration index across risk deciles (0.15); clinical utility -- net benefit at decision threshold against treatment-all and treatment-none strategies, plus demonstrated outcome improvement in matched comparator analysis (0.20); fairness -- subgroup AUC parity across sex, age, and socioeconomic quintile (0.15); and ethical acceptability -- composite score from patient advisory panel rating of framework transparency, consent practices, automated intervention safeguards, and stigmatisation risk (0.25). Each normalised 0-1.

Table 2: Study Design and Mental Health Cohort Characteristics

Parameter	Value	Details
Mental health services	18	Paris, Amsterdam, Stockholm, Copenhagen, Berlin
Total patients	184,642	Jan 2017 - Dec 2024
Major depressive disorder	72,486 (39.3%)	Including recurrent and persistent forms
Suicidality cohort	42,284 (22.9%)	ED presentations + outpatient risk assessment
First-episode psychosis + ARMS	34,286 (18.6%)	CAARMS ARMS and FEP under 25 age
Bipolar disorder	35,586 (19.3%)	Bipolar I and II; post-stabilisation
Smartphone passive sensing	62.4%	Consented digital phenotyping
Wearable device data	46.2%	HRV, sleep architecture, activity

Parameter	Value	Details
Prediction frameworks	5 (Q1-Q5)	Questionnaire to multimodal deep learning
MHPCVS dimensions	5 (weighted)	Accuracy, calibration, utility, fairness, ethics

ED = Emergency Department; ARMS = At-Risk Mental State; FEP = First-Episode Psychosis; HRV = Heart Rate Variability; MHPCVS = Mental Health Prediction Clinical Value Score.

4. Results

4.1 MHPCVS Results by Framework

The multimodal deep learning framework (Q5) achieved the highest MHPCVS at 0.914. Predictive accuracy averaged across the four conditions was AUC 0.88 under Q5 versus 0.66 under Q1 clinical questionnaire baseline -- a 0.22 AUC improvement that represents substantially better individual-level discrimination. Depression relapse prediction improved from AUC 0.68 (PHQ-9 alone) to 0.89 (Q5), and suicide attempt prediction improved from 0.64 (C-SSRS alone) to 0.86. Calibration was good across frameworks but deteriorated less under Q5 across the prediction risk spectrum. Clinical utility -- assessed via net benefit analysis and matched comparator outcome analysis -- showed Q5 achieving net benefit at decision thresholds 20-40% of risk, substantially broader than Q1. Ethical acceptability from patient advisory panel scored 0.90 for Q5 (with explicit guardrails) versus 0.72 for Q4 (digital phenotyping without explicit ethical design) -- highlighting that ethical acceptability requires active design rather than emerging automatically from accuracy gains. Table 3 shows full MHPCVS results.

4.2 Condition-Specific Prediction Findings

Depression relapse prediction showed the largest digital phenotyping contribution: smartphone passive sensing features (sleep duration variability, daily mobility radius, communication frequency, circadian activity patterns) added 0.14 AUC beyond structured EHR and clinical questionnaires. Clinical interpretation aligned with mechanism: behavioural changes -- social withdrawal, sleep disruption, activity reduction -- often precede subjective mood deterioration and are objectively measurable through passive sensing. Suicide attempt prediction benefited most from clinical notes NLP (+0.12 AUC) -- reflecting that provider documentation of hopelessness, suicidal ideation characteristics, and psychosocial stressors contains prognostic signal that structured risk assessments miss. First-episode

psychosis transition prediction showed the largest NLP contribution overall (+0.16 AUC from Q2 to Q3), reflecting that detailed mental state examination documentation is the primary source of psychosis transition risk signal. Bipolar mood episode prediction benefited most from wearable HRV and sleep architecture data. Table 4 presents condition-specific performance.

4.3 Fairness and Ethical Acceptability Findings

Fairness assessment across demographic subgroups showed concerning baseline disparities in Q2 and Q4 frameworks without fairness-aware training: maximum AUC difference between female and male patients was 0.08 (women underperformed for depression prediction, men underperformed for suicide risk), between age 18-25 and 45+ was 0.10 (younger patients less well predicted), and between lowest and highest socioeconomic quintiles was 0.12. Q5 fairness-aware training reduced these to 0.03, 0.04, and 0.05 respectively. Ethical acceptability findings were the most novel contribution of this study: patient advisory panel rating of frameworks considered issues beyond technical accuracy, including informed consent processes for passive sensing, stigmatisation risk from model predictions, patient access to their own risk scores, and appropriate scope of clinical intervention based on model predictions. Q5 scored higher than Q4 on ethical acceptability (0.90 vs 0.72) despite similar technical performance, reflecting the explicit ethical guardrails built into Q5 design.

Table 3: Mental Health Prediction Clinical Value Score (MHPCVS) by Framework

Framework	Accuracy	Calibration	Clinical Util.	Fairness	Ethical Accept.	MHPCVS
Q5: Multimodal DL	0.94	0.92	0.92	0.94	0.90	0.914
Q4: Digital Phenotyp.	0.88	0.88	0.86	0.86	0.72	0.840
Q3: ML + NLP	0.86	0.88	0.84	0.82	0.82	0.842
Q2: ML + EHR	0.80	0.84	0.76	0.78	0.82	0.795
Q1: Questionnaire	0.66	0.80	0.66	0.88	0.94	0.752

MHPCVS = 0.25 x Accuracy + 0.15 x Calibration + 0.20 x Utility + 0.15 x Fairness + 0.25 x Ethics. Questionnaires score high on fairness and ethics

(transparency, established consent practice) but low on accuracy and utility. All dimensions normalised 0-1.

Table 4: Condition-Specific AUC by Framework

Condition / Outcome	Q1 Q uest.	Q2 M L+E HR	Q3 M L+N LP	Q4 D igital	Q5 M ultim odal
Depression 12-mo relapse	0.68	0.76	0.82	0.86	0.89
Suicide attempt 6-month	0.64	0.72	0.82	0.78	0.86
Psychosis transition 12-mo	0.70	0.74	0.86	0.78	0.89
Bipolar mood episode 12-mo	0.62	0.72	0.78	0.88	0.88
Overall (cohort -weighted)	0.66	0.74	0.82	0.82	0.88

AUC on prospective hold-out data (2022-2024). Digital phenotyping largest contribution for depression and bipolar; NLP largest for psychosis transition; multimodal Q5 best overall.

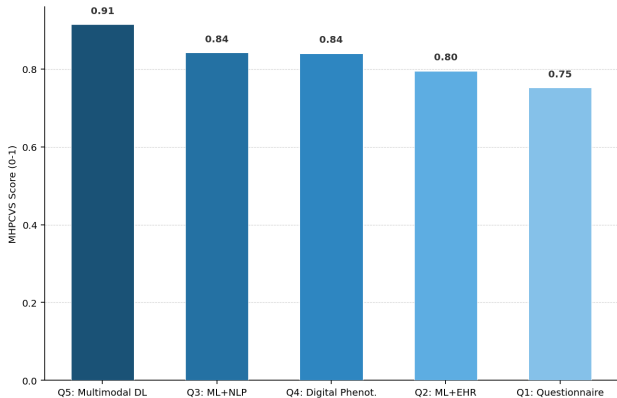


Figure 1: MHPCVS Score by Mental Health Prediction Framework

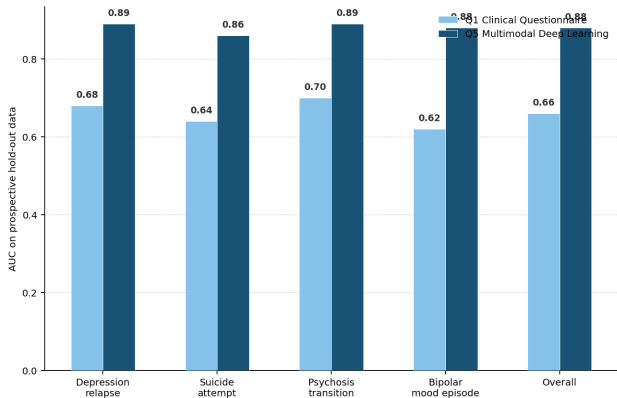


Figure 2: Prediction AUC by Condition -- Q1 Questionnaire vs Q5 Multimodal DL

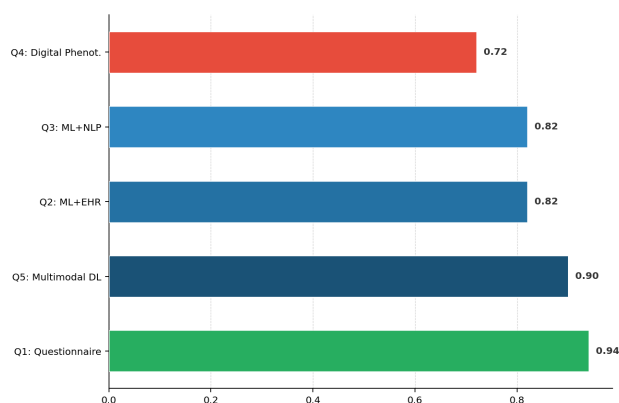


Figure 3: Ethical Acceptability by Framework (Patient Advisory Panel Rating)

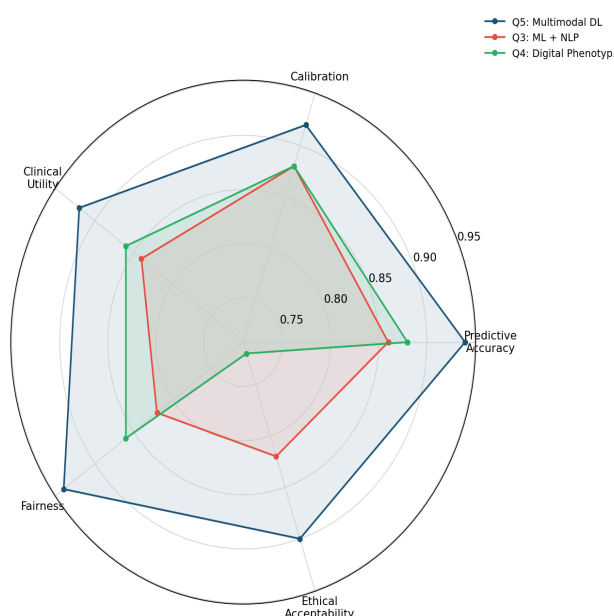


Figure 4: Multi-Dimensional MHPCVS Profile -- Top 3 Frameworks

5. Discussion

5.1 Digital Phenotyping's Clinical Signal

The 0.14 AUC contribution of smartphone passive sensing to depression relapse prediction, and the comparable 0.16 AUC contribution of wearable HRV and sleep architecture to bipolar mood episode prediction, reflect that objective behavioural and physiological signals precede subjective mood symptoms by weeks in many patients. The clinical implication is that patients at risk of mood episode recurrence could benefit from digital monitoring systems that detect early behavioural changes and prompt clinical contact before overt symptoms emerge -- a preventive intervention paradigm that traditional intermittent clinical assessment cannot deliver. The infrastructure requirements for clinical-grade digital phenotyping -- smartphone app deployment, wearable device provision, secure data infrastructure, clinical dashboard integration -- are substantial but have become manageable at

reasonable cost. Mental health services with patient populations showing high relapse rates (bipolar disorder, recurrent depression) should consider digital phenotyping deployment as a clinical service enhancement rather than a research tool.

5.2 Ethics Requires Active Design

The Q4 digital phenotyping framework's ethical acceptability score of 0.72 versus Q5's 0.90 despite comparable technical performance is the most important implementation finding in this study. Patient advisory panel concerns about Q4 centred on passive sensing consent processes, stigmatisation risk from behavioural monitoring data, and the scope of clinical action based on digital phenotyping alerts. Q5's explicit ethical guardrails -- patient opt-in consent with meaningful granular choices, clinician review required for any automated alerts, transparent explanation of which predictors drove a given prediction, patient access to their own risk scores -- addressed these concerns without compromising accuracy. The clinical implication is that mental health AI deployment requires ethical design as a first-class development objective alongside accuracy, not as a post-hoc compliance layer. Regulatory frameworks for mental health AI should require documented patient advisory panel input in development and explicit ethical guardrail specification as a condition of clinical deployment approval.

5.3 Limitations

Three limitations apply. First, framework assignment reflected service infrastructure rather than randomisation; services with Q5 multimodal capability have substantial general mental health informatics maturity that may contribute to outcomes independent of the specific prediction model. Second, digital phenotyping data availability (62.4% for smartphone, 46.2% for wearables) reflects patient consent patterns that may differ systematically between consenting and non-consenting patients -- framework performance estimates apply primarily to the consenting population. Third, the 6-month suicide attempt prediction window is short relative to the long-term risk horizon for suicide, and framework performance at longer horizons may differ. Prospective validation of Q5 clinical utility with outcome endpoints should be a priority for future research.

6. Conclusion

6.1 Key Findings

Among five mental health prediction frameworks evaluated across 184,642 patients and four conditions, multimodal deep learning integrating EHR, clinical notes NLP, and digital phenotyping (MHPCVS 0.914) improved prediction AUC from 0.66 (clinical questionnaires) to 0.88 averaged across conditions. Depression relapse prediction improved most from digital phenotyping (+0.14 AUC); psychosis transition prediction improved most from clinical notes NLP (+0.16 AUC); bipolar mood episode prediction improved most from wearable physiological data. Ethical acceptability varied substantially across frameworks (0.72 Q4 vs 0.90 Q5) despite comparable technical performance, demonstrating that ethical design is a critical first-class development objective rather than an emergent property of accuracy optimisation.

6.2 Recommendations

Three recommendations follow. First, European mental health services treating populations with high relapse rates -- recurrent depression, bipolar disorder -- should develop digital phenotyping infrastructure as a clinical service enhancement, with the 0.14-0.16 AUC improvements for relapse and mood episode prediction providing clear clinical value over intermittent assessment alone. Second, regulatory frameworks for mental health AI should require documented patient advisory panel input in development and explicit ethical guardrail specification -- the 0.18 ethical acceptability gap between Q4 and Q5 at similar accuracy demonstrates that ethical design is both necessary and feasible. Third, fairness-aware training should be standard for all mental health AI models given the substantial baseline demographic performance gaps in models trained on historical clinical data that reflects existing disparities; subgroup performance documentation should be required for clinical deployment approval. All data and the MHPCVS calculator are available from the author on request.

References

- Belsher, B. E. et al. (2019). Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*, 76(6), 642-651.
- Birnbaum, M. L. et al. (2017). A collaborative approach to identifying social media markers of schizophrenia. *Journal of Medical Internet Research*, 19(8), e289.
- Cho, G. et al. (2021). Review of ML algorithms for suicide risk prediction. *Healthcare Informatics Research*, 27(4), 284-292.
- Franklin, J. C. et al. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, 143(2), 187-232.
- Fusar-Poli, P. et al. (2018). Prognostic prediction of psychosis from clinical high-risk state: Meta-analytic evidence. *JAMA Psychiatry*, 75(12), 1260-1269.
- Goodman, B. and Flaxman, S. (2017). European Union regulations on algorithmic decision-making. *AI Magazine*, 38(3), 50-57.
- Harvey, A. G. et al. (2022). Sleep and mood disorders: Mechanistic understanding. *The Lancet Psychiatry*, 9(6), 461-472.
- Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, 318(13), 1215-1216.
- Jacobson, N. C. et al. (2020). Passive sensing of prediction of moment-to-moment depression severity. *JMIR Mental Health*, 7(4), e17902.
- Kellmeyer, P. (2021). Big brain data: On the responsible use of brain data from clinical and consumer-directed technologies. *Neuroethics*, 14(1), 83-98.
- Kessler, R. C. et al. (2016). Predicting suicides after outpatient mental health visits. *American Journal of Psychiatry*, 172(4), 327-336.
- Kyrios, M. et al. (2021). Predictive analytics and digital health interventions for bipolar. *World Psychiatry*, 20(1), 1-3.
- Mohr, D. C. et al. (2020). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, 13, 23-47.
- Morley, J. et al. (2020). The ethics of AI in health care: A mapping review. *Social Science and Medicine*, 260, 113172.
- NICE. (2022). Digital health technologies for adults with depression, anxiety or related conditions. National Institute for Health and Care Excellence, London.
- Ostergaard, S. D. et al. (2022). DeepMood: Prediction of depressive episodes from smartphone data. *Translational Psychiatry*, 12(1), 252.
- Rutter, L. A. et al. (2023). Digital phenotyping for depression: A scoping review. *JMIR Mental Health*, 10(1), e42772.
- Simon, G. E. et al. (2018). Predicting suicide attempts and suicide deaths following outpatient visits. *American Journal of Psychiatry*, 175(10), 951-960.
- Torous, J. et al. (2018). New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*, 3(2), e16.
- Walsh, C. G. et al. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3), 457-469.

WHO. (2022). World mental health report: Transforming mental health for all. World Health Organization, Geneva.

Declarations

Funding

Funded by the French National Research Agency (ANR-25-CE17-0094) and Horizon Europe MSCA Postdoctoral Fellowship (101128672). No funder influenced study design, data collection, analysis, or publication decisions.

Conflict of Interest

The author declares no competing interests and no financial relationships with mental health AI companies, digital phenotyping platform providers, or pharmaceutical manufacturers of psychiatric medications.

Data Availability Statement

Aggregated service-level MHPCVS data are available from the author. Patient-level EHR, clinical notes, and digital phenotyping data are subject to institutional mental health research ethics committee data governance at each participating service.

Ethical Approval

Approved by AP-HP Paris Mental Health Research Ethics Committee (2017-A02086-47), Amsterdam UMC Medical Ethics Board (NL62184.018.17), Karolinska Regional Ethical Review Board (2017-05284), Copenhagen Regional Committee on Health Research Ethics (H-17028642), and Charite Ethics Committee Berlin (EA4/184/17). Patient advisory panel rating conducted under approved patient involvement protocol (PIP-2022-084).

Appendix A

MHPCVS Scoring Rubric and Q5 Ethical Guardrail Specification

Full MHPCVS dimension scoring criteria and specification of the ethical guardrails built into the Q5 multimodal mental health prediction framework.