Additional file 1: TRIPOD+AI checklist

Checklist from Collins GS, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385:e078378. Page numbers refer to the submitted manuscript. D, development; E, evaluation; AUPRC, area under the precision–recall curve; AUROC, area under the receiver operating characteristic curve; BPB, below personal baseline; CI, confidence interval; EMA, ecological momentary assessment; MAE, mean absolute error; MHI, Mental Health Index; RMSE, root mean squared error; TOST, two one-sided tests. In this study models were developed and evaluated (temporal hold-out) in the same cohort.

Section/topic	Item	D/E	Checklist item	Reported (page)	How it is addressed
Title					
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Title (1)	Prediction model development and temporal validation study; forecasting a causal Mental Health Index from passive smartphone sensing in university students
Abstract					
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	Abstract (1–2)	Structured abstract: design, data, models, primary analysis, results with CIs, conclusions
Introduction					
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	Background (3–4)	Prognostic, next-day forecasting for mental health monitoring; rationale and prior personalized models
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users	Background (3); Discussion, Implications (23)	University students; monitoring between service contacts; intended users would be monitoring services. No deployment is proposed
	3c	D;E	Describe any known health inequalities between sociodemographic groups	Background (3); Methods, Data source (6)	Prevalence among students cited; demographic data were not part of the data used, stated explicitly
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	Background (4)	Three research questions; development and temporal evaluation
Methods					
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets, the rationale for using these data, and representativeness of the data	Data source and participants (5–6)	StudentLife cohort; rationale (multi-item EMA with sensing and PHQ-9); same source for development and hold-out
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	Data source (5); Data partition (8)	24 March–5 June 2013; hold-out from 19 May 2013
Participants	6a	D;E	Specify key elements of the study setting including the number and location of centres	Data source (5)	One university (Dartmouth College), spring term 2013
	6b	D;E	Describe the eligibility criteria for study participants	Data source (6)	$\geq$72 h history, $\geq$50% accelerometer coverage; $\geq$20 development and $\geq$3 hold-out days
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	–	Not applicable (observational cohort, no treatments recorded)
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	Data source (6); Predictors (7–8); Data partition (8–9)	Time-zone conversion, proportions, session spreading, causal imputation, availability flags, automated leakage tests; identical for all participants
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale, and consistency across groups	Outcome (6–7)	Causal MHI from four EMA items, behavioural day 04:00–04:00, forecast for the coming day; same definition for all participants
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	–	Not applicable (self-reported EMA, computed algorithmically)

Section/topic	Item	D/E	Checklist item	Reported (page)	How it is addressed
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	–	Not applicable (outcome computed from self-reports independently of predictors)
Predictors	9a	D	Describe the choice of initial predictors and any pre-selection of predictors before model building	Predictors (7)	All available hourly sensing modalities; no data-driven pre-selection
	9b	D;E	Clearly define all predictors, including how and when they were measured	Predictors (7–8); Fig. 2; Additional file 2, Table S1	29 channels defined; 72 h before 04:00 of the forecast day
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	–	Not applicable (passive sensors)
Sample size	10	D;E	Explain how the study size was arrived at, and justify that the study size was sufficient. Include details of any sample size calculation	Data source (6); Statistical analysis (11–13)	Fixed by the public dataset and eligibility; detectable difference (0.017 MAE, 80% power) and equivalence margin reported; small sample acknowledged
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	Predictors (7); Data source (6)	Causal imputation with availability flags; days with $<$50% coverage or $<$72 h history excluded
Analytical methods	12a	D	Describe how the data were used in the analysis, including whether the data were partitioned	Data partition (8–9); Figs 1–2	Calendar hold-out used once; four expanding folds per participant with 1-day gap; chronological early stopping
	12b	D	Describe how predictors were handled (functional form, rescaling, transformation, standardisation)	Predictors (7–8); Data partition (8–9); Identity strategies (9)	Log transforms, proportions, training-fold standardisation, causal personal normalisation (I1)
	12c	D	Specify the type of model, rationale, all model building steps, including any hyperparameter tuning, and method for internal validation	Identity strategies (9); Models (9–10); Baselines (10–11); Figs 3–4; Additional file 2, Figure S1 and Table S2	Four encoders, five identity strategies, nine baselines; Optuna tuning; 10 seeds; internal validation by chronological folds
	12d	D;E	Describe if and how any heterogeneity in estimates and model performance was handled and quantified across clusters	Statistical analysis (11)	Participants treated as clusters: cluster bootstrap CIs, per-participant MAE comparisons, within-person correlation
	12e	D;E	Specify all measures and plots used to evaluate model performance and, if relevant, to compare multiple models	Statistical analysis (11–13)	MAE (primary), RMSE, R^2, within-person r; Wilcoxon with Holm; TOST; AUROC, AUPRC, decision curves
	12f	E	Describe any model updating arising from the model evaluation	–	None (no recalibration or updating after hold-out evaluation)
	12g	E	For model evaluation, describe how the model predictions were calculated	Models (9–10); Availability (25)	Seed-ensembled predictions from trained models; code provided
Class imbalance	13	D;E	If class imbalance methods were used, state why and how	–	Not applicable (regression; no imbalance methods for BPB)
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	Data source (6)	None; demographic data not available (stated)
Model output	15	D	Specify the output of the prediction model. Provide details and rationale for any classification and how the thresholds were identified	Models (9); Statistical analysis (12)	Continuous MHI in $(-1, 1)$; BPB alert threshold set on validation predictions
Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in setting, eligibility, outcome and predictors	Data partition (8); Table 1 (15)	Same setting and definitions; later calendar period; lower BPB prevalence in hold-out reported
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe consent or waiver	Ethics approval (24–25)	Original study approved by the Dartmouth College IRB (no reference number reported) with signed informed consent; no additional approval required for secondary analysis of public de-identified data

Section/topic	Item	D/E	Checklist item	Reported (page)	How it is addressed
Open science					
Funding	18a	D;E	Give the source of funding and the role of the funders	Funding (25)	Ministry grant; funder had no role
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	Competing interests (25)	None declared
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	Study design (4)	No separate protocol; analysis plan fixed in code before hold-out examination
Registration	18d	D;E	Provide registration information or state that the study was not registered	Study design (4)	Not registered
Data sharing	18e	D;E	Provide details of the availability of the study data	Availability (25); Additional file 3	Original data public; derived analysis-ready dataset and hourly inputs provided as Additional file 3 (item-level PHQ-9 withheld)
Code sharing	18f	D;E	Provide details of the availability of the analytical code	Availability (25); Additional file 3	Complete analysis code with locked package versions provided as Additional file 3
Patient and public involvement					
PPI	19	D;E	Provide details of any patient and public involvement or state no involvement	Study design (4)	No involvement (stated)
Results					
Participants	20a	D;E	Describe the flow of participants through the study, including numbers with and without the outcome and follow-up time	Data source (6); Participants (14); Table 1 (15)	$48 \rightarrow 46 \rightarrow 28 \rightarrow 19$ participants; participant-days and BPB days per partition
	20b	D;E	Report the characteristics overall and for each data source or setting, including key dates, predictors, sample size, outcome events, follow-up and missing data	Table 1 (15); Participants (14)	Days, MHI, BPB events, coverage, PHQ-9; demographics not available
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors and outcome	Table 1 (15)	Development vs hold-out: days, MHI, BPB prevalence
Model development	21	D;E	Specify the number of participants and outcome events in each analysis	Data partition (8–9); Table 1 (15); Additional file 2	19 participants in all folds and hold-out; 682/164 days; 118/9 BPB days
Model specification	22	D	Provide details of the full prediction model to allow predictions in new individuals and third-party evaluation, including any restrictions	Models (9–10); Fig. 4; Additional file 2 Table S2; Availability (25)	Architectures and hyperparameters; code and trained configuration available
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for key subgroups. Consider plots	Results (14–20); Tables 2–4; Figs 5–8; Additional file 2, Tables S3–S14	Cluster-bootstrap CIs; subgroup by baseline PHQ-9
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters	Results (16–18); Additional file 2, Tables S5–S7	Per-participant comparisons and within-person correlations
Model updating	24	E	Report the results from any model updating	–	Not applicable (no updating)
Discussion					
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Discussion (21–23)	Principal findings, mechanisms, comparison with literature; fairness not assessable (stated)
Limitations	26	D;E	Discuss any limitations and their effects on biases, statistical uncertainty and generalisability	Strengths and limitations (23)	Small single-site sample, 2013 data, sparse EMA, few BPB events, post hoc additional analyses, no demographics

Section/topic	Item	D/E	Checklist item	Reported (page)	How it is addressed
Usability	27a	D	Describe how poor quality or unavailable input data should be assessed and handled when implementing the model	Predictors (7)	Availability flags and causal imputation; coverage threshold; no implementation proposed
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and the level of expertise required	Implications (23)	Not intended for deployment; models showed no clinical utility
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalisability	Implications (23); Conclusions (24)	Longer horizons, sustained change, larger and newer cohorts, leakage audits