Schizophrenia Bulletin vol. 47 no. 2 pp. 284–297, 2021
doi:10.1093/schbul/sbaa120
Advance Access publication 11 September 2020

MAJOR REVIEW

Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice

Gonzalo Salazar de Pablo[1–3], Erich Studerus[4], Julio Vaquerizo-Serrano[1–3], Jessica Irving[1], Ana Catalan[1,6–8], Dominic Oliver[1,9], Helen Baldwin[1], Andrea Danese[3,9,10], Seena Fazel[11], Ewout W. Steyerberg[12,13], Daniel Stahl[5], and Paolo Fusar-Poli*[,1,14–16]

[1]Early Psychosis: Interventions and Clinical-detection Lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 5th Floor, PO63, 16 De Crespigny Park, SE5 8AF London, UK; [2]Institute of Psychiatry and Mental Health, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón School of Medicine, Universidad Complutense, Instituto de Investigación Sanitaria Gregorio Marañón, CIBERSAM, Madrid, Spain; [3]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [4]Division of Personality and Developmental Psychology, Department of Psychology, University of Basel, Basel, Switzerland; [5]Biostatistics Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [6]Department of Psychiatry, Basurto University Hospital, Bilbao, Spain; [7]Mental Health Group, BioCruces Health Research Institute, Bizkaia, Spain; [8]Neuroscience Department, University of the Basque Country UPV/EHU, Leioa, Spain; [9]Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK; [10]National and Specialist CAMHS Clinic for Trauma, Anxiety, and Depression, South London and Maudsley NHS Foundation Trust, London, UK; [11]Department of Psychiatry, University of Oxford, Oxford, UK; [12]Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, the Netherlands; [13]Department of Public Health, Erasmus MC, Rotterdam, the Netherlands; [14]OASIS Service, South London and Maudsley NHS Foundation Trust, London, UK; [15]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; [16]National Institute for Health Research, Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK

*To whom correspondence should be addressed; tel: +44-0-20-7848-0900, fax:+44-0-20-7848-0976, e-mail: paolo.fusar-poli@kcl.ac.uk

Background: **The impact of precision psychiatry for clinical practice has not been systematically appraised. This study aims to provide a comprehensive review of validated prediction models to estimate the individual risk of being affected with a condition (diagnostic), developing outcomes (prognostic), or responding to treatments (predictive) in mental disorders.** ***Methods:*** **PRISMA/RIGHT/CHARMS-compliant systematic review of the Web of Science, Cochrane Central Register of Reviews, and Ovid/PsycINFO databases from inception until July 21, 2019 (PROSPERO CRD42019155713) to identify diagnostic/prognostic/predictive prediction studies that reported individualized estimates in psychiatry and that were internally or externally validated or implemented. Random effect meta-regression analyses addressed the impact of several factors on the accuracy of prediction models.** ***Findings:*** **Literature search identified 584 prediction modeling studies, of which 89 were included. 10.4% of the total studies included prediction models internally validated (n = 61), 4.6% models externally validated (n = 27), and 0.2% (n = 1) models considered for implementation. Across validated prediction modeling studies (n = 88), 18.2% were diagnostic, 68.2% prognostic, and 13.6% predictive. The most frequently investigated condition was psychosis (36.4%), and the most frequently employed predictors clinical (69.5%). Unimodal compared to multimodal models (β = .29, P = .03) and diagnostic compared to prognostic (β = .84, p < .0001) and predictive (β = .87, P = .002) models were associated with increased accuracy.** ***Interpretation:*** **To date, several validated prediction models are available to support the diagnosis and prognosis of psychiatric conditions, in particular, psychosis, or to predict treatment response. Advancements of knowledge are limited by the lack of implementation research in real-world clinical practice. A new generation of implementation research is required to address this translational gap.**

Key words: risk/prognosis/prediction/individualized/prevention/evidence/implementation/validation

Introduction

Precision medicine is an emerging approach for disease prevention, diagnosis, and treatment that considers individual variability in patient and disease characteristics, genes, environment, and lifestyle of each person.[1,2] The

concept of precision medicine is not new; clinicians have been working to personalize care tailored to people's individual health needs throughout the history of medicine (eg, matching human blood groups across donors and recipients during blood transfusion).[3] Yet, modern advancements of knowledge in the field of individualized prediction modeling have allowed the consolidation of an evidence-based science of precision medicine.[4] Prediction modeling can be used to forecast the probability of a certain condition being present (diagnostic models), outcomes (prognostic models), or the response to interventions (predictive models) at the individual subject level. From a methodological perspective, individualized prediction modeling research includes studies that investigate the development, internal or external validation of prediction models, and prediction model impact studies, which investigate the real-world effect of using prediction models in clinical practice.[5] External validity is the extent to which the predictions can be generalized to the data from plausibly related settings, while internal validity is the extent to which the predictions fit the derivation data after controlling for overfitting and optimism, with the latter representing the difference in a model's performance in the derivation data and unseen individuals (for further details see[4]).

More recently, individualized prediction models have been developed in psychiatry,[4] and a new field of precision psychiatry has emerged.[6–8] The area where individualized prediction models have been more extensively investigated in psychiatry relates to psychotic disorders. The high personal, clinical, and societal burden associated with psychosis, coupled with the limited pathophysiological understanding, has stimulated research into diagnostic prediction models. Incorporation of a clinical staging model for psychosis,[9] together with the emergence of the clinical high-risk state for psychosis (CHR-P),[10,11] has prompted research into prognostic prediction models, as well as several ongoing international collaborations.[12] The associated need to stratify or personalize early intervention or preventive treatment for psychosis[13,14] has stimulated research of predictive prediction models. Furthermore, emerging research has indicated that prediction modeling can benefit from transdiagnostic approaches that allow methodological cross-fertilization across other nonpsychotic disorders.[15–17]

Despite the increasing number of records published in this area over recent years, the impact of precision psychiatry for psychosis, and more broadly for clinical practice, is unclear. No study to our knowledge has comprehensively reviewed the advancements and challenges of prediction modeling in clinical psychiatry to date. Our primary aim was to systematically appraise diagnostic, prognostic, or predictive individualized prediction models that can be considered for clinical use in psychiatry, with a specific focus on psychosis; the secondary aim was to test potential moderating factors. The evidence reviewed was then used to formulate pragmatic recommendations to advance knowledge in this area. To address the potential impact of precision psychiatry, we focused on diagnostic, prognostic, and predictive prediction model studies with at least internal or external validation and implementation studies.

Methods

This study (study protocol: PROSPERO CRD42019155713) was conducted in accordance with the RIGHT[18] and PRISMA[19] statements (supplementary table 1).

Search Strategy and Selection Criteria

A multistep independent researcher systematic literature search strategy was used to identify the relevant articles. First, the Web of Science, Cochrane Central Register of Reviews, and Ovid/ PsycINFO database were searched, from inception until July 21, 2019 in English (specific search terms are reported in supplementary methods 1). Second, the references of the articles identified in previous reviews in the field and the references from the included studies were manually searched to identify additional relevant records. Abstracts identified through the previous step were then screened and, after the exclusion of those not relevant to the current study, their full texts were assessed against the inclusion and exclusion criteria. In a fourth step, a researcher with expertise in risk estimation models in psychiatry (E.S.) further checked the articles against the core biostatistical inclusion criteria (ie, presence of appropriate internal or external validation).

The inclusion criteria were (1) original studies or study protocols published in the databases searched or gray literature; (2) studies reporting the diagnostic (principally predicts the presence of a certain condition), prognostic (principally predicts the clinical outcomes in the absence of therapy[20]), predictive (principally predicts the response to a particular intervention[20]), or implementation of risk estimation models; (3) providing estimates at the individual subject level or in subgroups; (4) studies investigating individuals affected by mental disorders or mental conditions or individuals at risk of mental disorders, defined according to established psychometric criteria, and (5) diagnostic, prognostic, or predictive studies that performed at least a proper internal or external validation (see below). The exclusion criteria were: (1) abstracts, conference proceedings, reviews, or meta-analyses; (2) diagnostic, prognostic, or predictive models that did not provide individualized or subgroup risk estimates; (3) diagnostic, prognostic, or predictive studies that did not perform any proper internal or external validation (see supplementary methods 2); or (4) predictors-finding studies that did not report prediction models.

Descriptive Measures and Data Extraction

The variables extracted in the current review included items listed in the "Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies" (CHARMS[21]). Additional variables were included[22] as detailed in the supplementary methods 3. When more than one outcome per study was found in the same category, we extracted the information for the primary outcome, as defined in each article, unless the study reported multiple primary co-outcomes.

Quality Assessment

Risk of bias was assessed for each of the included studies adapting "The Prediction Model Risk of Bias Assessment Tool" (PROBAST v5/05/2019[5,23]). PROBAST includes 4 steps and assesses the risk of bias and applicability of 4 core domains (participants, predictors, outcome, and analysis) to obtain an overall judgment of the risk of bias.[5] An outcome is considered to be at high risk of bias when at least one of the questions answered is not appropriate (no or probably no). The overall risk of bias is considered high when one or more domains are considered to be at high risk[24] (details can be found in supplementary methods 4).

Data Analysis

All the included studies were systematically summarized in tables stratified by the model type (diagnostic, prognostic, and predictive)—those implemented were then discussed in a separate section—and reporting core descriptive variables (supplementary methods 5). The top 10% of the most widely employed predictors and all the studied conditions were summarized in graphs, and the specific methodological characteristics of the studies were summarized in a separate table. These descriptive analyses were complemented by the Pearson correlation between apparent vs external accuracy within the models that reported both.[16, 25–52] We further conducted meta-analytical regressions to estimate the association between accuracy and (1) the type of validation (internal vs external); (2) the type of accuracy measure (area under the curve [AUC] vs C-statistics vs accuracy, with the latter category including accuracy measures other than AUC or C-statistics as defined by each study); (3) the type of model (diagnostic vs prognostic vs predictive model); (4) the number of specific predictors; (5) the type of predictors (clinical or service use or sociodemographic vs any biomarker—neuroimaging or electroencephalography or magnetoencephalography or proteomic or genetic or cognitive—vs a combination of modalities); (6) the modality of predictors (unimodal, using only 1 type of predictor, eg, clinical only, vs multimodal, using more than 1 type of predictor, eg, clinical and biomarker); (7) type of analysis (machine learning vs statistical modeling, as defined in supplementary methods 6). For analyses 4–7, we also included the interaction between accuracy and meta-regressors. For analyses 2–7, we used accuracy values prioritizing external validation over internal validation, in line with the previous meta-analyses of prediction models.[53] In the case of multiple studies on the same prediction model in which the previous order of priority could not be applied, the study with the largest data set was employed. We performed a meta-regression of the difference between logit transformed accuracy (because of the bounded nature of AUC[53]) using a random effect meta-analysis model, taking 1–7 clustering of comparisons into account.[53] The analyses were performed with Comprehensive Meta-Analysis Version 3.[54]

Results

Database

The literature search yielded 50 698 records and, after the exclusion of nonrelevant abstracts, 1033 full-text articles were screened to identify a total of 584 prediction studies reporting on prediction models developed. These models were then screened for eligibility against the inclusion and exclusion criteria to identify 89 studies with individualized prediction models, which were validated or implemented and represented the final sample (PRISMA; figure 1): 61 were internally validated (10.4% of the total models developed), 27 were externally validated (4.6% of the total models developed), and 1 (0.2% of the total models developed) described a protocol for the implementation of a prediction model (figure 2). Thirty point three percent (27/89) of the prediction models included were externally validated. 8.2% studies reported on diagnostic prediction models, 68.2% on prognostic models, and 13.6% on predictive models; 55.6% of studies employed sociodemographic predictors, 69.5% employed clinical predictors, 10.2% employed cognitive predictors, 13.6% employed service use predictors, 25.0% employed physical health predictors, 17.0% employed neuroimaging predictors, 0.4% employed magnetoencephalography or electroencephalography predictors, 0.1% employed proteomic data, and 2.3% employed genetic predictors. The most frequently reported predictors were age (n = 38, 45.8%), sex (n = 27, 32.5%), education (n = 21, 25.3%), and depressive symptoms (n = 18, 21.7%; figure 3). The most frequently reported condition was psychosis (36.4%; figure 3). The total sample size was 3 889 457 individuals, ranging from 29[55] to 2 960 929[56] individuals. The average age ranged from 1.8[57] to 64.7[38] years. The source of data encompassed cohorts (46 studies, 52.3%), case-control studies (13 studies, 14.8%), clinical trials (16 studies, 18.2%), and registry data (13 studies, 14.8%).

The most frequent type of external validation was geographical, examining the model performance in other centers or regions 24/27 (88.9%). Internal validation was more

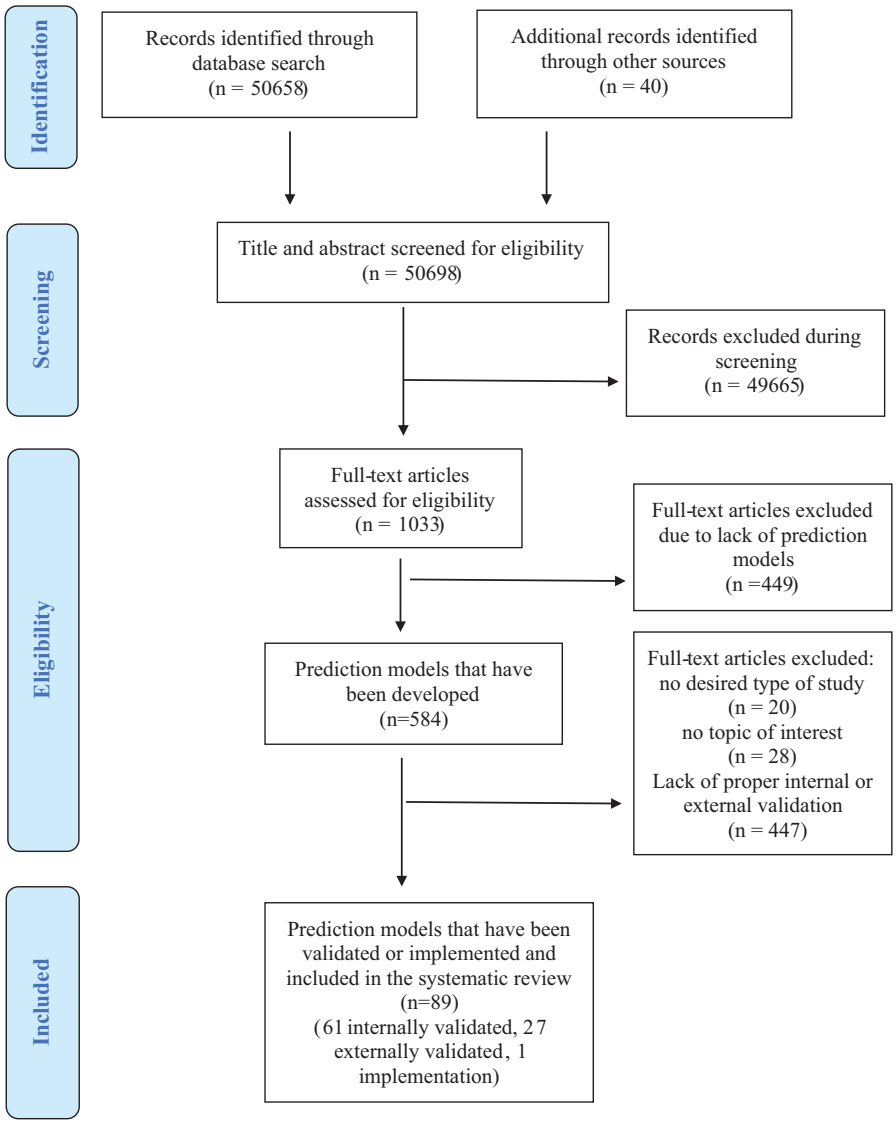

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart outlining study selection process.

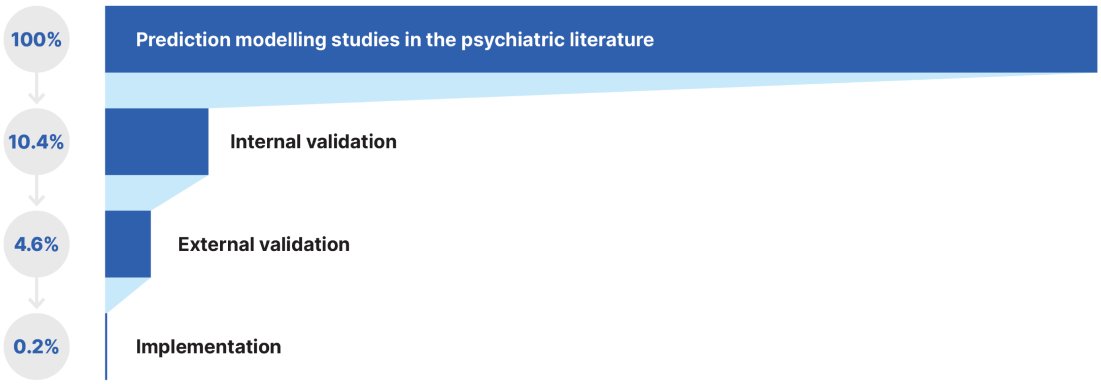

Fig. 2. Proportion of prediction models studies developed, internally validated, externally validated, and implemented in the psychiatric literature.

frequently done by cross-validation in 34/61 (55.7%). The most frequent modeling method was machine learning in 35/88 (39.8%) (supplementary table 2). In half of the studies (51.1%), there was no explicit handling of missing data; imputation (27.3%) was the most common method for data missingness (supplementary table 2). AUC was the

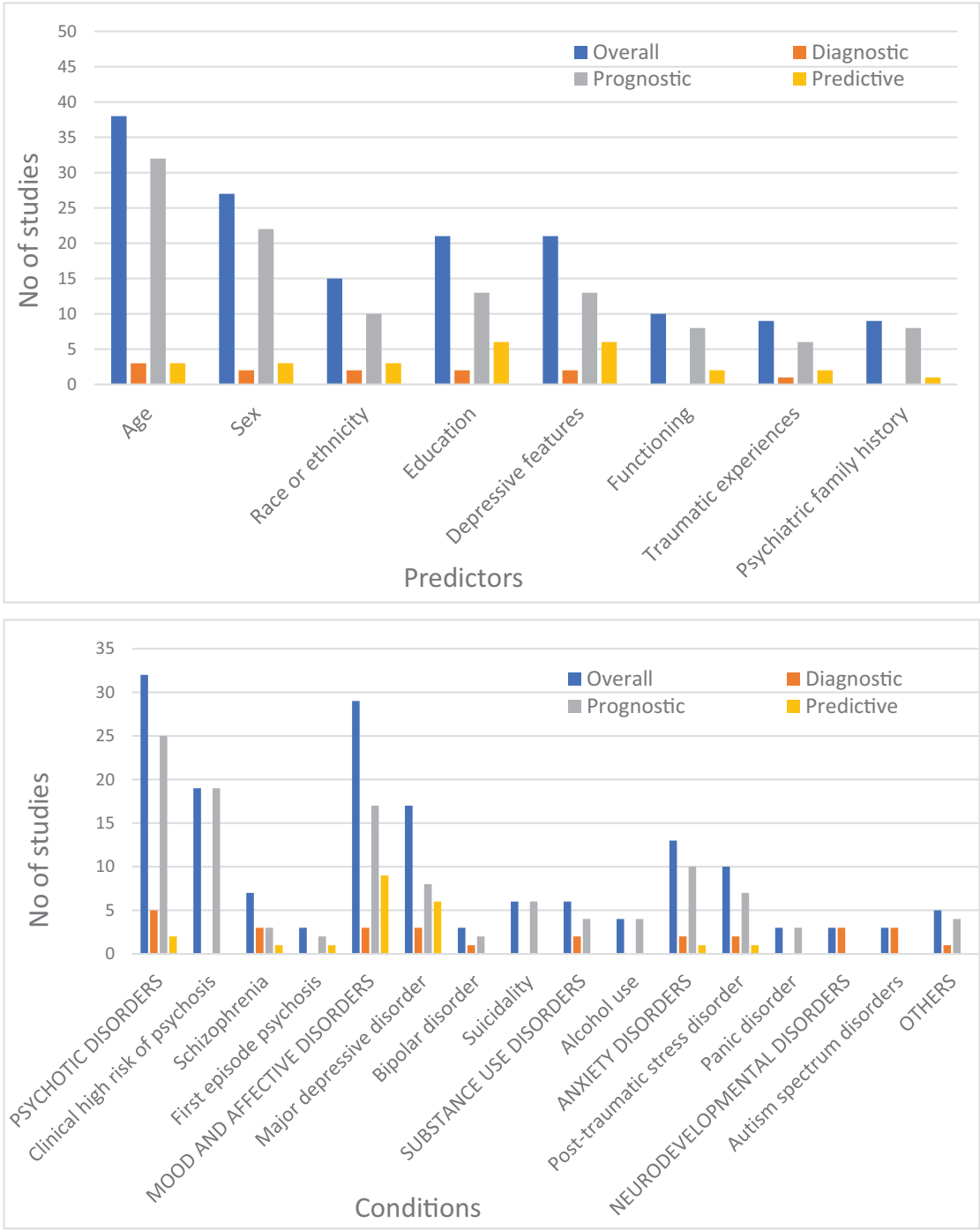

Fig. 3. Most frequently reported predictors (above, top 10%) and conditions (below, all) in the included studies.

most commonly reported measure of model performance (78.4%; supplementary table 2). Only 10.2% of the studies presented their model in full: almost half of them did not present any details of their model (46.6%) or the calibration results clearly (47.7%; supplementary table 2).

Diagnostic Risk Estimation Models

Four studies employed neuroimaging methods[58–60] and proteomic data[61] to classify individuals with schizophrenia compared to healthy controls (HC)[58,59,61] or to differentiate schizophrenia spectrum disorder and HC with or without impaired social functioning[60] (supplementary table 3). One study employed clinical predictors to discriminate between affective and schizophrenia spectrum psychoses.[62]

Two studies employed neuroimaging to differentiate unipolar vs bipolar depression[25] or major depression vs dysthymia in individuals with panic disorder and agoraphobia.[63] Another study used clinical predictors to

distinguish melancholic vs non-melancholic features in individuals with major depression.[64]

A neuroimaging study discriminated smokers and non-smoking HC.[65] Another study using sociodemographic, clinical, and cognitive data discriminated individuals with cocaine dependence from HC.[66] Problematic internet use was discriminated from HC using clinical and sociodemographic predictors.[26] Two studies classified posttraumatic stress disorder in veterans using sociodemographic and clinical predictors[67] or magnetoencephalography.[68] Three studies focused on autism spectrum disorders to discriminate them from attention deficit hyperactive disorder[69,70] or from HC[57] using clinical predictors[69,70] or genomic biomarkers[57] (supplementary table 3).

Prognostic Models

A considerable proportion of the prognostic risk estimation studies[16, 27–32, 71–82] (31.7%) investigated the CHR-P[83] (supplementary table 4). These studies focused on the prediction of psychosis onset in CHR-P individuals (n = 13),[27–29,71–80] functional outcomes and disability in CHR-P individuals (n = 2),[81,82] psychosis onset in individuals undergoing a CHR-P assessment (pretest risk n = 1),[30] and the transdiagnostic onset of psychosis in secondary mental health care (n = 3).[16,31,32] Six of these studies employed sociodemographic or clinical predictors only,[16,28,31,32,73,74] 1 employed sociodemographic and service use data,[30] 2 included cognitive measures beyond sociodemographic and clinical data,[27,72] 3 included cognitive measures alone,[29,71,77] 1 employed electroencephalography predictors,[75] 3 neuroimaging alone,[76,78,80] and 2 neuroimaging in association with clinical measures[81,82] or in association with sociodemographic, clinical, and cognitive measures (n = 1).[79] Four other studies focused on established psychosis using different combinations of sociodemographic, clinical, service use, cognitive, and physical health predictors to forecast psychotic relapses,[84] hospital admission,[33] employment, education or training status,[34] and mortality.[85] Nine studies focused on depression.[35–40,82,86,87] A combination of sociodemographic, clinical, and physical health factors was used by 3 studies to predict the onset of major depression in the general population[35–37] and by 5 other studies to predict persistence[38,86,87] or recurrence[39,40] of major depression. A further study predicted disability in recent-onset depression using clinical and neuroimaging data.[82] One study focused on the onset of bipolar spectrum disorders in youth at family risk using sociodemographic and clinical factors,[88] while another one predicted cognitive impairment in bipolar disorder using sociodemographic and cognitive factors.[89] Six studies used a combination of sociodemographic, clinical, physical health, and service use to predict suicidality, focusing on suicide ideation in the general population,[41,90] suicide attempts after outpatient visits,[56] suicide attempts in adolescents,[91] suicidal behavior,[92] or deaths by suicide after hospitalization in soldiers.[93] Seven studies focused on posttraumatic stress disorder (PTSD).[94–100] Three studies employed a combination of sociodemographic, clinical, physical health, and service use factors to predict the onset of PTSD[94–96] or the remission of PTSD (n = 3 studies),[97–99] and a further study used clinical predictors alone to forecast PTSD features in soldiers.[100] Sociodemographic, clinical, and physical health data were used by 2 studies[42,43] to predict the onset of generalized anxiety disorders and panic disorder in the general population and by another study to predict the recurrence of panic disorder.[44]

Two studies predicted alcohol use in young people using sociodemographic and clinical[45,46] predictors in combination with cognitive[46] predictors, while another 2 studies predicted abstinence from heavy drinking using sociodemographic and/or clinical[47,101] data. A prediction model forecasted offending behavior in schizophrenia and delusional disorder using forensic information.[102] Compulsory admission into psychiatric wards was predicted by a combination of sociodemographic, clinical, and service use factors,[103] and medication-induced altered mental status in hospitalized patients was predicted by sociodemographic, clinical, service use, and physical health data.[104] Other models predicted the onset of common mental disorder in a working population using sociodemographic, clinical, and physical health[105] variables, mental health hospital readmission using sociodemographic, clinical, and service use[106] data, and violent offending in severe mental disorders using sociodemographic, clinical, and service use[48] data.

Predictive Models

Two studies employed a combination of clinical, sociodemographic, or physical health features to predict remission[49,50] or response to antidepressants[107,108] in major depression. Three studies predicted the onset of treatment-resistant depression using clinical and sociodemographic variables,[51,52,109] service use data,[52,109] and physical health data.[109] A study employed clinical and sociodemographic data to predict the level of functioning at 4 and 52 weeks after antipsychotic treatment in patients with first-episode psychosis.[110] Two studies predicted the clinical response to transcranial magnetic stimulation combining neuroimaging and electroencephalography factors.[55,111] A further study employed clinical and physical health data to predict treatment dropout from psychotherapy in anxiety disorders[112] (supplementary table 5)

Implementation of Prediction Models

Among externally validated models, the transdiagnostic model predicting psychosis onset in secondary mental

Table 1: Replicated prediction models (all prognostic)

Studies	Data source, Outcome	Development sample size (mean age, Location)	Validation sample size (mean age, Location)	Predictors (n)	Performance (measure)[a]	Predictor categories			
						SOC	CLIN	COG	PHY
Cannon et al 2016[74]	Psychosis onset in CHR-P	596 (18.5, NAPLS-2, USA)		Age, family history, unusual thoughts and suspiciousness, lower verbal learning and memory performance, slower speed of processing, decline in social functioning (6)	0.71 (C-index)				
Carrión et al. 2016[28]			176 (16.6, EDIPPP, USA)[a]		0.79 (AUC, Ext)	X	X	X	
Fusar-Poli et al. 2017[32]	Transdiagnostic psychosis onset in secondary mental healthcare patients	33820 (34.4, SLAM[b], UK)	54716 (32.0, SLAM[c], UK)	Age, gender, ethnicity, age by gender, and ICD-10 index diagnosis (5)	0.8 (C-index, Dev); 0.79 (C-index, Ext)				
Fusar-Poli et al 2019[34]		33820 (34.4, SLAM[b], UK)	13702 (40.9, C&I NHS Trust)		0.73 (C-index, Ext)	X	X		
King et al. 2008[38]	MD onset in primary care	5216 (48.9, Europe[d])	1732 (47.0, Chile)	Age, sex, country educational status, difficulties in work, history of depression in first-degree relatives, experience of discrimination, lifetime major depression episode, mental quality of life, physical quality of life (10)	0.79 (C-index, Dev); 0.71 (C-index, Ext)				
Nigatu et al. 2016[39]		5216 (48.9, Europe[d])	29621 (43.8 NESARC, USA)		0.71 (AUC, Ext)	X	X		X
King et al. 2011[44]	GAD and PD onset in primary care	4905 (n.a., UK, Spain, Slovenia, Portugal)	5140 (n.a. NetherlandsEstonia, Chile)	Age, sex, country, difficulties in paid and unpaid work, history of depression in first-degree relatives, follow-up period, lifetime major depression episode, mental quality of life, physical quality of life (9)	0.75 (C-index, Dev); 0.71-0.81 (C- index, Ext)				
Nigatu et al. 2019[45]		4905 (n.a., UK, Spain, Slovenia, Portugal)	24626 (n.a. NESARC, USA)		0.62 (AUC, Ext)	X	X		X

[a]Trauma and life events predictors not included in the external validation; [b]Lambeth and Southwark; [c]Any other borough; [d]United Kingdom, Spain, Slovenia, Portugal, The Netherlands

AUC: area under the curve; C&I: Camden and Islington; CLIN: clinical; COG: cognitive; Dev: development; EDIPPP: Early Detection, Intervention, and Prevention of Psychosis Program; Ext: external validation; GAD: generalised anxiety disorder; ICD: International Classification of Diseases; MD: major depression; n.a.: not available; NAPLS-2: North American Prodrome Longitudinal Study- 2; NESARC: US National Epidemiologic Survey on Alcohol and Related Conditions; PD: panic disorder PHY: physical health; SLAM: South London and Maudsley; SOC: sociodemographic.

health care,[16,31,32] the model predicting psychosis onset in CHR-P,[27,72] the model predicting the onset of generalized anxiety disorders and panic disorder in the general population,[42,43] and the model predicting the onset of major depression in the general population[36,37] were all replicated twice (table 1). None of the models included in the current systematic review were fully implemented in clinical practice. However, 1 study[113] described the protocol for the implementation of the transdiagnostic risk calculator to detect individuals at risk of psychosis in secondary mental health care.[16,31,32] The core aim of this study was to integrate the prediction model in the local electronic health register and evaluate the clinician's adherence to the recommendations made by the risk calculator.[113]

Accuracy of Prediction Models and Meta-Regressions

Accuracy of prediction models was highly variable, ranging from 0.56[40] to 1.0[71] (0.69[25]–0.96[69] for diagnostic models, 0.56[40] to 1.0[71] for prognostic models, and 0.66[108] to 0.92[114] for predictive models) (supplementary tables 3–5). Within the nonoverlapping prediction model studies that reported apparent and external accuracy (n = 18), the 2 measures were strongly correlated (r = .78, 95% CI: 0.39–0.95, P < .001; figure 4). Meta-regressions revealed that accuracy was higher in unimodal (n = 25) vs multimodal (n = 71) prediction models (β = .29, P = .03), diagnostic (n = 14) vs prognostic (n = 51; β = .84, P < .001) models, and diagnostic (n = 14) vs predictive (n = 11; β = .87, P = .002) models, but no other significant meta-regressions or interactions were detected (supplementary results and supplementary table 6).

Quality of Prediction Models

Applying PROBAST, 94.3% of the included studies were found to be at high risk of bias. The results from the different domains were heterogeneous: 1.1% were at high risk of bias in the participants domain, 65.9% in the predictors domain, 90.9% in the outcomes domain, and 81.8% in the analysis domain (supplementary table 7; supplementary figure 1).

Discussion

This is the first large-scale systematic review to summarize the transdiagnostic and life span-inclusive evidence regarding diagnostic, prognostic, or predictive prediction models that have been internally and externally validated and, thus, can be considered for clinical implementation in psychiatry. Currently, only 10.4% of the total models developed are internally validated, 4.6% are externally validated, and 0.2% are considered for implementation. Most of the models validated were prognostic, followed by diagnostic and more infrequently predictive models. Most research in this area focused on psychosis and was life span inclusive. Most prediction models employed clinical predictors. Many studies were at high risk of bias and accuracy was mediated by several factors.

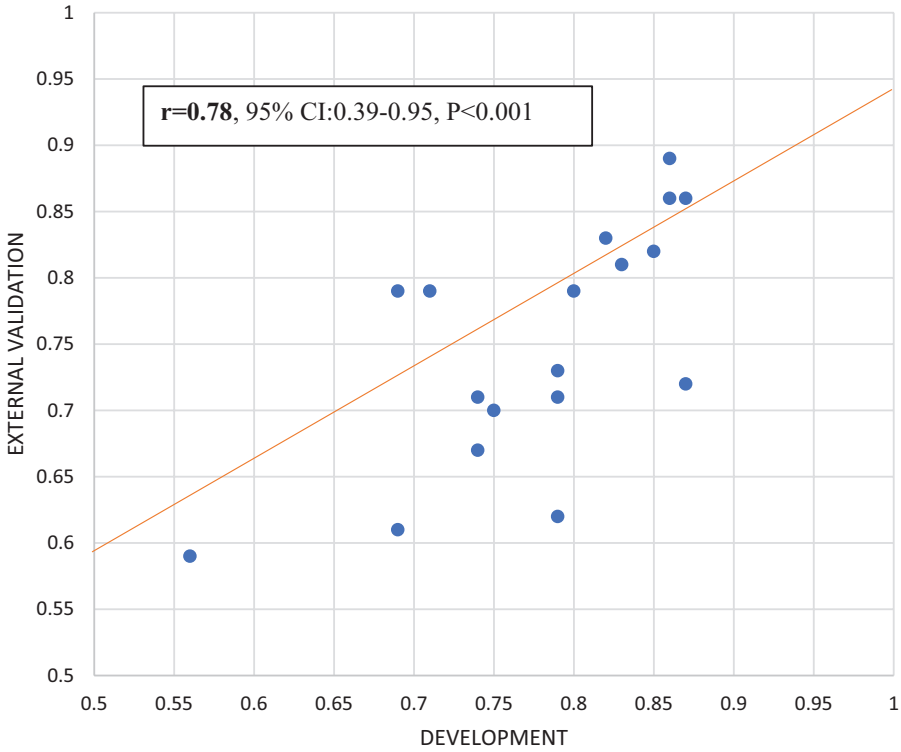

Fig. 4. Correlation between apparent and external accuracy (n = 18).

The main finding of this study is that precision psychiatry has developed into a consolidated area of clinical research, with a substantial number of individualized prediction models developed and validated on data from 3 889 457 participants aged from 1.8 to 64 years. These substantial advancements in the field of precision psychiatry reflect a life span-inclusive approach. Several validated individualized prediction models are nowadays available, transdiagnostically targeting many psychiatric conditions encompassing psychotic disorders, affective disorders, substance use disorders, anxiety disorders, neurodevelopmental disorders, and several clinically relevant outcomes as well. However, to date, psychosis research has mostly led (36.4%) precision psychiatry. Notably, the majority (68.2%) of the current psychiatric prediction models were prognostic, with CHR-P studies representing a leading field (31.7%) in this domain (21.6% across all prognostic, diagnostic, and predictive models). This finding confirms the traction role of psychosis research, as well as the close link between precision psychiatry and preventive approaches. Psychiatry as a discipline is essentially "Hippocratic," whereby the prediction of outcomes becomes more relevant than the ascertainment of cross-sectional diagnostic categories.[4] The validity of diagnostic categories in psychiatry has always been criticized and it has recently been further questioned by transdiagnostic approaches, which challenged discrete and fixed self-delimitating boundaries across International Classification of Diseases or *Diagnostic and Statistical Manual of Mental Disorders* entities.[15,17] These considerations are particularly valid for early psychosis, where the prediction of outcomes can inform treatment approaches and can explain why diagnostic models were not so frequent (18.2%). Predictive models were even less frequently investigated (13.6%), presumably because these types of studies are inherently more complex to run owing to the intervention-related component. Despite these speculations, accuracy in diagnostic models remained superior to prognostic and predictive models, presumably because diagnostic models rely on more established gold standards to define outcomes.

Despite the substantial progress in developing and validating individualized prediction models for psychiatry, this study also highlighted some important barriers to the advancement of knowledge. The first barrier is that, across the overall pool of prediction models developed and published in the broader psychiatric literature (n = 584), we found only about 15% (n = 88) to be properly validated (n = 61: 10.4% internal validation and n = 27: 4.6% independent external validation). Within those included in the review, about one-third were validated in external databases (supplementary limitations). This finding aligns with a previous review suggesting that external validation of prediction models is infrequent.[115] A growing body of evidence has confirmed a replicability crisis in several areas of scientific knowledge, such as cancer research,[116] economics,[117] behavioral ecology economics,[117] and genetic behavior research.[118] Since precision psychiatry is a relatively emerging paradigm compared to other precision medicine approaches, research to date may have prioritized the development of new models over the external validation of models already established. For example, systematic reviews in chronic obstructive pulmonary disease identified a similar number of prediction models with internal (n = 100) and external (n = 38) validation to the ones reported here.[24] However, several of these models were externally validated between 5 and 17 times.[24] The next generation of prediction modeling in psychiatry should, therefore, consider, along with the development of new prediction models, the replication of existing algorithms across different scenarios. This would necessitate collaborative data-sharing efforts to reach critical mass (studies' sample size ranged from 29[55] to 2 960 929[56] individuals) and the establishment of international clinical research infrastructures, as well as specific support from funders and stakeholders. The current study should also educate editors and reviewers who too often devalue replication studies because they feel that these studies have limited advancement of knowledge compared to the original publications. In reality, focusing on the reproducibility of existing prediction models and updating existing prognostic models, as opposed to dropping these models and developing new ones from scratch, is the recommended procedure to maximize the efficiency of research.[4]

This study also provides relevant methodological evidence. For example, to date, most models (69.5%) are based on clinical predictors and there is no evidence that more complex models encompassing biomarkers or a large number of predictors (which may be more prone to overfitting issues) or advanced analytical methods, such as machine learning, outperform other types of prediction models. These findings align with recent studies indicating that complex machine learning models do not outperform more parsimonious clinically based models developed through standard statistical approaches.[53,119] The current study adds further methodological value by showing that, in psychiatry, for a given apparent accuracy (we found no difference across various accuracy measures), the expected external accuracy can be estimated with a correlating factor of .78 (95% CI: 0.39–0.95; figure 4). Editors and reviewers can use this factor to assess the external accuracy of prediction models that have not been internally/externally validated. However, current guidelines recommend performing at least internal validation,[4] which, if properly performed, can accurately index the true external generalizability of the model (as shown in our meta-regressions).

An associated problem is that 94.3% studies included in the current review—which adopted stringent inclusion criteria focusing on validated studies—were eventually classified at high risk of bias, mostly because of the high risk of bias in the outcomes and analysis domain. These

biases may potentially be even more substantial in the wider literature, limiting the implementation of precision psychiatry. Although the PROBAST threshold for this bias may be too strict, our findings are consistent with an independent review, which applied PROBAST and found that 98.3% of the prediction models were at high risk of bias.[24] Facilitating the external validation of individualized prediction models is also the most robust approach to address the currently largest barrier for precision psychiatry: real-world implementation.

The current systematic review identified only one implementation study, corresponding to 0.2% of the total pool of models developed and published, which did not report data but only described the research protocol of an ongoing project[120] (the full implementation results have been published upon completion of our literature review).[121,122] At the moment, precision psychiatry is severely limited by a translational gap. The implementation pathways of precision psychiatry is a perilous journey,[123] complicated by obstacles related to patients (eg, making their data available or accepting the outputs of the risk calculator), clinicians (eg, adherence to the recommendations made by prediction models and communicating risks), providers (eg, confidentiality and accessibility of data and interpretability and utility of outputs), and funders and organizations (implementing an infrastructure enabling standard prediction procedures). Because of these challenges, most prediction models that are validated are then lost in the dearth of real-world implementation science, even for psychosis research. Implementation science itself, although much needed, is contested and complex, with the unpredictable use of results from routine clinical practice.[124,125] For example, the Consolidated Framework for Implementation Research (CFIR)[30] is rather theoretical[124] and does not offer specific pragmatic guidance to precision psychiatry. A recent systematic review concluded that only 6% of studies acknowledging the CFIR used the CFIR in a meaningful way.[126] Thus, the paucity of implementation studies of individualized prediction models in psychiatry can be secondary to the lack of a general implementation framework and practical guidance. The next generation of empirical research in the field of prediction modeling in psychiatry and psychosis research should primarily aim at filling in the implementation gap by developing a coherent and practical implementation framework, methodological infrastructures, and international implementation infrastructures.

Conclusions

To date, several validated prediction models are available to support the diagnosis and prognosis of psychiatric conditions, in particular, psychotic disorders, or to predict the response to treatments. Advancements of knowledge are mostly limited by the limited replication and lack of implementation research in real-world clinical practice. The next generation of precision psychiatry research is required to address this translational gap.

Supplementary Material

Supplementary material is available at *Schizophrenia Bulletin*.

Funding

This study was supported by the King's College London Confidence in Concept award from the Medical Research Council (MC_PC_16048) to Dr Fusar-Poli. Dr Salazar de Pablo and Dr Vaquerizo-Serrano are supported by the Alicia Koplowitz Foundation. Dr Danese was funded by the Medical Research Council (grant no. P005918) and the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust, and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.
Conflict of interest: The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. Terry SF. Obama's precision medicine initiative. *Genet Test Mol Biomarkers.* 2015;19(3):113–114.
2. Genetics Reference. What is precision medicine? https://ghr.nlm.nih.gov/primer/precisionmedicine/definition. Accessed March 1, 2020.
3. Farhud DD, Zarif Yeganeh M. A brief history of human blood groups. *Iran J Public Health.* 2013;42(1):1–6.
4. Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The science of prognosis in psychiatry: a review. *JAMA Psychiatry.* 2018;75(12):1289–1297.
5. Wolff RF, Moons KGM, Riley RD, et al.; PROBAST Group. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170(1):51–58.
6. Fernandes BS, Williams LM, Steiner J, et al. The new field of "precision psychiatry." *BMC Med.* 2017;15(1):80.
7. Khanra S, Khess CRJ, Munda SK. "Precision psychiatry": a promising direction so far. *Indian J Psychiatry.* 2018;60(3):373–374.
8. Williams LM. Precision psychiatry: a neural circuit taxonomy for depression and anxiety. *Lancet Psychiatry.* 2016;3(5):472–480.
9. Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. *World Psychiatry.* 2017;16(3):251–265.
10. Fusar-Poli P, Cappucciati M, Borgwardt S, et al. Heterogeneity of psychosis risk within individuals at clinical high risk: a meta-analytical stratification. *JAMA Psychiatry.* 2016;73(2):113–120.
11. Fusar-Poli P, Salazar de Pablo G, Correll CU, et al. Prevention of psychosis: advances in detection, prognosis, and intervention. *JAMA Psychiatry.* 2020;77(7):755–765.

12. Tognin S, van Hell HH, Merritt K, et al.; PSYSCAN Consortium. Towards precision medicine in psychosis: benefits and challenges of multimodal multicenter studies-PSYSCAN: translating neuroimaging findings from research into clinical practice. *Schizophr Bull.* 2020;46(2):432–441.
13. Davies C, Cipriani A, Ioannidis JPA, et al. Lack of evidence to favor specific preventive interventions in psychosis: a network meta-analysis. *World Psychiatry.* 2018;17(2):196–209.
14. Correll CU, Galling B, Pawar A, et al. Comparison of early intervention services vs treatment as usual for early-phase psychosis: a systematic review, meta-analysis, and meta-regression. *JAMA Psychiatry.* 2018;75(6):555–565.
15. Fusar-Poli P. TRANSD recommendations: improving transdiagnostic research in psychiatry. *World Psychiatry.* 2019;18(3):361–362.
16. Fusar-Poli P, Davies C, Rutigliano G, et al. Transdiagnostic individualized clinically based risk calculator for the detection of individuals at risk and the prediction of psychosis: model refinement including nonlinear effects of age. *Front Psychiatry.* 2019;10:313.
17. Fusar-Poli P, Solmi M, Brondino N, et al. Transdiagnostic psychiatry: a systematic review. *World Psychiatry.* 2019;18(2):192–207.
18. Chen Y, Yang K, Marušic A, et al.; RIGHT (Reporting Items for Practice Guidelines in Healthcare) Working Group. A reporting tool for practice guidelines in health care: the RIGHT statement. *Ann Intern Med.* 2017;166(2):128–132.
19. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535.
20. Clark GM. Prognostic factors versus predictive factors: examples from a clinical trial of erlotinib. *Mol Oncol.* 2008;1(4):406–412.
21. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 2014;11(10):e1001744.
22. Studerus E, Ramyead A, Riecher-Rössler A. Prediction of transition to psychosis in patients with a clinical high risk for psychosis: a systematic review of methodology and reporting. *Psychol Med.* 2017;47(7):1163–1178.
23. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med.* 2019;170(1):W1–W33.
24. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ.* 2019;367:l5358.
25. Redlich R, Almeida JJ, Grotegerd D, et al. Brain morphometric biomarkers distinguishing unipolar and bipolar depression. A voxel-based morphometry-pattern classification approach. *JAMA Psychiatry.* 2014;71(11):1222–1230.
26. Ioannidis K, Chamberlain SR, Treder MS, et al. Problematic internet use (PIU): associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry. *J Psychiatr Res.* 2016;83:94–102.
27. Carrión RE, Cornblatt BA, Burton CZ, et al. Personalized prediction of psychosis: external validation of the NAPLS-2 psychosis risk calculator with the EDIPPP project. *Am J Psychiatry.* 2016;173(10):989–996.
28. Perkins DO, Jeffries CD, Cornblatt BA, et al. Severity of thought disorder predicts psychosis in persons at clinical high-risk. *Schizophr Res.* 2015;169(1-3):169–177.
29. Corcoran CM, Carrillo F, Fernández-Slezak D, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry.* 2018;17(1):67–75.
30. Fusar-Poli P, Rutigliano G, Stahl D, et al. Deconstructing pretest risk enrichment to optimize prediction of psychosis in individuals at clinical high risk. *JAMA Psychiatry.* 2016;73(12):1260–1267.
31. Fusar-Poli P, Rutigliano G, Stahl D, et al. Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry.* 2017;74(5):493–500.
32. Fusar-Poli P, Werbeloff N, Rutigliano G, et al. Transdiagnostic risk calculator for the automatic detection of individuals at risk and the prediction of psychosis: second replication in an independent national health service trust. *Schizophr Bull.* 2019;45(3):562–570.
33. Addington DE, Beck C, Wang J, et al. Predictors of admission in first-episode psychosis: developing a risk adjustment model for service comparisons. *Psychiatr Serv.* 2010;61(5):483–488.
34. Leighton SP, Krishnadas R, Chung K, et al. Predicting one-year outcome in first episode psychosis using machine learning. *PLoS One.* 2019;14(3):e0212846.
35. Bellon JA, Luna JD, King M, et al. Predicting the onset of major depression in primary care: international validation of a risk prediction algorithm from Spain. *Psychol Med.* 2011;41(10):2075–2088.
36. King M, Walker C, Levy G, et al. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. *Arch Gen Psychiatry.* 2008;65(12):1368–1376.
37. Nigatu YT, Liu Y, Wang JL. External validation of the international risk prediction algorithm for major depressive episode in the US general population: the PredictD-US study. *BMC Psychiatry.* 2016;16:256.
38. Maarsingh OR, Heymans MW, Verhaak PF, Penninx BWJH, Comijs HC. Development and external validation of a prediction rule for an unfavorable course of late-life depression: a multicenter cohort study. *J Affect Disord.* 2018;235:105–113.
39. Wang JL, Patten S, Sareen J, et al. Development and validation of a prediction algorithm for use by health professionals in prediction of recurrence of major depression. *Depress Anxiety.* 2014;31(5):451–457.
40. Klein NS, Holtman GA, Bockting CLH, Heymans MW, Burger H. Development and validation of a clinical prediction tool to estimate the individual risk of depressive relapse or recurrence in individuals with recurrent depression. *J Psychiatr Res.* 2018;104:1–7.
41. Liu Y, Sareen J, Bolton JM, Wang JL. Development and validation of a risk prediction algorithm for the recurrence of suicidal ideation among general population with low mood. *J Affect Disord.* 2016;193:11–17.
42. King M, Bottomley C, Bellón-Saameño JA, et al. An international risk prediction algorithm for the onset of generalized anxiety and panic syndromes in general practice attendees: predictA. *Psychol Med.* 2011;41(8):1625–1639.
43. Nigatu YT, Wang J. External validation of the international risk prediction algorithm for the onset of generalized anxiety and/or panic syndromes (the predict A) in the US general population. *J Anxiety Disord.* 2019;64:40–44.
44. Liu Y, Sareen J, Bolton J, Wang J. Development and validation of a risk-prediction algorithm for the recurrence of panic disorder. *Depress Anxiety.* 2015;32(5):341–348.

45. Ngo DA, Rege SV, Ait-Daoud N, Holstege CP. Development and validation of a risk predictive model for student harmful drinking—A longitudinal data linkage study. *Drug Alcohol Depend.* 2019;197:102–107.
46. Afzali MH, Sunderland M, Stewart S, et al. Machine-learning prediction of adolescent alcohol use: a cross-study, cross-cultural validation. *Addiction.* 2019;114(4):662–671.
47. Gueorguieva R, Wu R, O'Connor PG, et al. Predictors of abstinence from heavy drinking during treatment in COMBINE and external validation in PREDICT. *Alcohol Clin Exp Res.* 2014;38(10):2647–2656.
48. Fazel S, Wolf A, Larsson H, et al. Identification of low risk of violent crime in severe mental illness with a clinical prediction tool (Oxford Mental Illness and Violence tool [OxMIV]): a derivation and validation study. *Lancet Psychiatry.* 2017;4(6):461–468.
49. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry.* 2016;3(3):243–250.
50. Furukawa TA, Kato T, Shinagawa Y, et al. Prediction of remission in pharmacotherapy of untreated major depression: development and validation of multivariable prediction models. *Psychol Med.* 2019;49(14):2405–2413.
51. Perlis RH. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry.* 2013;74(1):7–14.
52. Kautzky A, Dold M, Bartova L, et al. Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study. *Acta Psychiatr Scand.* 2019;139(1):78–88.
53. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12–22.
54. Borenstein M, Hedges L, Higgins J, Rothstein H. *Comprehensive Meta-Analysis Version 3 [Computer Program]*. Englewood, NJ: Biostat; 2013.
55. Zandvakili A, Philip NS, Jones SR, et al. Use of machine learning in predicting clinical response to transcranial magnetic stimulation in comorbid posttraumatic stress disorder and major depression: a resting state electroencephalography study. *J Affect Disord.* 2019;252:47–54.
56. Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry.* 2018;175(10):951–960.
57. Pramparo T, Pierce K, Lombardo MV, et al. Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry.* 2015;72(4):386–394.
58. Kalmady SV, Greiner R, Agrawal R, et al. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *npj Schizophr.* 2019;5(1):2.
59. Rozycki M, Satterthwaite TD, Koutsouleris N, et al. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. *Schizophr Bull.* 2018;44(5):1035–1044.
60. Viviano JD, Buchanan RW, Calarco N, et al.; Social Processes Initiative in Neurobiology of the Schizophrenia(s) Group. Resting-state connectivity biomarkers of cognitive performance and social function in individuals with schizophrenia spectrum disorder and healthy control subjects. *Biol Psychiatry.* 2018;84(9):665–674.
61. Cooper JD, Han SYS, Tomasik J, et al. Multimodel inference for biomarker development: an application to schizophrenia. *Transl Psychiatry.* 2019;9(1):83.
62. Jauhar S, Krishnadas R, Nour MM, et al. Is there a symptomatic distinction between the affective psychoses and schizophrenia? A machine learning approach. *Schizophr Res.* 2018;202:241–247.
63. Lueken U, Straube B, Yang Y, et al. Separating depressive comorbidity from panic disorder: a combined functional magnetic resonance imaging and machine learning approach. *J Affect Disord.* 2015;184:182–192.
64. Parker G, McCraw S, Hadzi-Pavlovic D. The utility of a classificatory decision tree approach to assist clinical differentiation of melancholic and non-melancholic depression. *J Affect Disord.* 2015;180:148–153.
65. Ding XY, Yang YH, Stein EA, Ross TJ. Combining multiple resting-state fMRI features during classification: optimized frameworks and their application to nicotine addiction. *Front Hum Neurosci.* 2017;11:362.
66. Ahn WY, Ramesh D, Moeller FG, Vassileva J. Utility of machine-learning approaches to identify behavioral markers for substance use disorders: impulsivity dimensions as predictors of current cocaine dependence. *Front Psychiatry.* 2016;7:34.
67. Harrington KM, Quaden R, Stein MB, et al.; VA Million Veteran Program and Cooperative Studies Program. Validation of an electronic medical record-based algorithm for identifying posttraumatic stress disorder in U.S. veterans. *J Trauma Stress.* 2019;32(2):226–237.
68. James LM, Belitskaya-Lévy I, Lu Y, et al. Development and application of a diagnostic algorithm for posttraumatic stress disorder. *Psychiatry Res.* 2015;231(1):1–7.
69. Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioral distinction of autism and ADHD. *Transl Psychiatry.* 2016;6(2):e732.
70. Duda M, Haber N, Daniels J, Wall DP. Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Transl Psychiatry.* 2017;7(5):e1133.
71. Bedi G, Carrillo F, Cecchi GA, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophr.* 2015;1:15030.
72. Cannon TD, Yu C, Addington J, et al. An individualized risk calculator for research in prodromal psychosis. *Am J Psychiatry.* 2016;173(10):980–988.
73. Ciarleglio AJ, Brucato G, Masucci MD, et al. A predictive model for conversion to psychosis in clinical high-risk patients. *Psychol Med.* 2019;49(7):1128–1137.
74. Malda A, Boonstra N, Barf H, et al. Individualized prediction of transition to psychosis in 1,676 individuals at clinical high risk: development and validation of a multivariable prediction model based on individual patient data meta-analysis. *Front Psychiatry.* 2019;10:345.
75. Ramyead A, Studerus E, Kometer M, et al. Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. *World J Biol Psychiatry.* 2016;17(4):285–295.
76. Koutsouleris N, Borgwardt S, Meisenzahl EM, et al. Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: results from the FePsy study. *Schizophr Bull.* 2012;38(6):1234–1246.
77. Koutsouleris N, Davatzikos C, Bottlender R, et al. Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophr Bull.* 2012;38(6):1200–1215.

78. Zarogianni E, Storkey AJ, Johnstone EC, Owens DG, Lawrie SM. Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features. *Schizophr Res.* 2017;181:6–12.
79. Chung Y, Addington J, Bearden CE, et al.; North American Prodrome Longitudinal Study Consortium. Adding a neuroanatomical biomarker to an individualized risk calculator for psychosis: a proof-of-concept study. *Schizophr Res.* 2019;208:41–43.
80. Koutsouleris N, Meisenzahl EM, Davatzikos C, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry.* 2009;66(7):700–712.
81. de Wit S, Ziermans TB, Nieuwenhuis M, et al. Individual prediction of long-term outcome in adolescents at ultra-high risk for psychosis: applying machine learning techniques to brain imaging data. *Hum Brain Mapp.* 2017;38(2):704–714.
82. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al.; PRONIA Consortium. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry.* 2018;75(11):1156–1172.
83. Fusar-Poli P. The clinical high-risk state for psychosis (CHR-P), Version II. *Schizophr Bull.* 2017;43(1):44–47.
84. Fond G, Bulzacka E, Boucekine M, et al.; FACE-SZ (FondaMental Academic Centers of Expertise for Schizophrenia) Group. Machine learning for predicting psychotic relapse at 2 years in schizophrenia in the national FACE-SZ cohort. *Prog Neuropsychopharmacol Biol Psychiatry.* 2019;92:8–18.
85. Austin PC, Newman A, Kurdyak PA. Using the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a population-based cohort of adults with schizophrenia in Ontario, Canada. *Psychiatry Res.* 2012;196(1):32–37.
86. Dinga R, Marquand AF, Veltman DJ, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry.* 2018;8(1):241.
87. Rubenstein LV, Rayburn NR, Keeler EB, et al. Predicting outcomes of primary care patients with major depression: development of a depression prognosis index. *Psychiatr Serv.* 2007;58(8):1049–1056.
88. Hafeman DM, Merranko J, Goldstein TR, et al. Assessment of a person-level risk calculator to predict new-onset bipolar spectrum disorder in youth at familial risk. *JAMA Psychiatry.* 2017;74(8):841–847.
89. Bauer IE, Suchting R, Van Rheenen TE, et al. The use of component-wise gradient boosting to assess the possible role of cognitive measures as markers of vulnerability to pediatric bipolar disorder. *Cogn Neuropsychiatry.* 2019;24(2):93–107.
90. Ryu S, Lee H, Lee DK, Park K. Use of a machine learning algorithm to predict individuals with suicide ideation in the general population. *Psychiatry Investig.* 2018;15(11):1030–1036.
91. Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *J Child Psychol Psychiatry.* 2018;59(12):1261–1270.
92. Tran T, Luo W, Phung D, et al. Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry.* 2014;14:76.
93. Kessler RC, Warner CH, Ivany C, et al.; Army STARRS Collaborators. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study To Assess Risk and rEsilience in Servicemembers (Army STARRS). *JAMA Psychiatry.* 2015;72(1):49–57.
94. Russo J, Katon W, Zatzick D. The development of a population-based automated screening procedure for PTSD in acutely injured hospitalized trauma survivors. *Gen Hosp Psychiatry.* 2013;35(5):485–491.
95. Papini S, Pisner D, Shumake J, et al. Ensemble machine learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization. *J Anxiety Disord.* 2018;60:35–42.
96. Rosellini AJ, Dussaillant F, Zubizarreta JR, Kessler RC, Rose S. Predicting posttraumatic stress disorder following a natural disaster. *J Psychiatr Res.* 2018;96:15–22.
97. Galatzer-Levy IR, Ma S, Statnikov A, Yehuda R, Shalev AY. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Transl Psychiatry.* 2017;7(3):e0.
98. Karstoft KI, Galatzer-Levy IR, Statnikov A, et al. Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC Psychiatry.* 2015;15:30.
99. Galatzer-Levy IR, Karstoft KI, Statnikov A, Shalev AY. Quantitative forecasting of PTSD from early trauma responses: a Machine Learning application. *J Psychiatr Res.* 2014;59:68–76.
100. Karstoft KI, Statnikov A, Andersen SB, Madsen T, Galatzer-Levy IR. Early identification of posttraumatic stress following military deployment: application of machine learning methods to a prospective study of Danish soldiers. *J Affect Disord.* 2015;184:170–175.
101. Gueorguieva R, Wu R, Fucito LM, O'Malley SS. Predictors of abstinence from heavy drinking during follow-up in COMBINE. *J Stud Alcohol Drugs.* 2015;76(6):935–941.
102. Hickey N, Yang M, Coid J. The development of the Medium Security Recidivism Assessment Guide (MSRAG): an actuarial risk prediction instrument. *J Forensic Psychiatry Psychol.* 2009;20(2):202–224.
103. Hotzy F, Theodoridou A, Hoff P, et al. Machine learning: an approach in identifying risk factors for coercion compared to binary logistic regression. *Front Psychiatry.* 2018;9:258.
104. Muñoz MA, Jeon N, Staley B, et al. Predicting medication-associated altered mental status in hospitalized patients: development and validation of a risk model. *Am J Health Syst Pharm.* 2019;76(13):953–963.
105. Fernandez A, Salvador-Carulla L, Choi I, et al. Development and validation of a prediction algorithm for the onset of common mental disorders in a working population. *Aust N Z J Psychiatry.* 2018;52(1):47–58.
106. Barker LC, Gruneir A, Fung K, et al. Predicting psychiatric readmission: sex-specific models to predict 30-day readmission following acute psychiatric hospitalization. *Soc Psychiatry Psychiatr Epidemiol.* 2018;53(2):139–149.
107. Serretti A, Olgiati P, Liebman MN, et al. Clinical prediction of antidepressant response in mood disorders: linear multivariate vs. neural network models. *Psychiatry Res.* 2007;152(2–3):223–231.
108. Maciukiewicz M, Marshe VS, Hauschild AC, et al. GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. *J Psychiatr Res.* 2018;99:62–68.

109. Kautzky A, Dold M, Bartova L, et al. Refining prediction in treatment-resistant depression: results of machine learning analyses in the TRD III Sample. *J Clin Psychiatry.* 2018;79(1):16m11385.
110. Koutsouleris N, Kahn RS, Chekroud AM, et al. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *Lancet Psychiatry.* 2016;3(10):935–946.
111. Erguzel TT, Ozekes S, Gultekin S, et al. Neural network based response prediction of rtms in major depressive disorder using QEEG cordance. *Psychiatry Investig.* 2015;12(1):61–65.
112. Niles AN, Wolitzky-Taylor KB, Arch JJ, Craske MG. Applying a novel statistical method to advance the personalized treatment of anxiety disorders: a composite moderator of comparative drop-out from CBT and ACT. *Behav Res Ther.* 2017;91:13–23.
113. Fusar-Poli P, Oliver D, Spada G, et al. Real world implementation of a transdiagnostic risk calculator for the automatic detection of individuals at risk of psychosis in clinical routine: study protocol. *Front Psychiatry.* 2019;10:109.
114. Koutsouleris N, Wobrock T, Guse B, et al. Predicting response to repetitive transcranial magnetic stimulation in patients with schizophrenia using structural magnetic resonance imaging: a multisite machine learning analysis. *Schizophr Bull.* 2018;44(5):1021–1034.
115. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol.* 2015;68(1):25–34.
116. Nosek BA, Errington TM. Making sense of replications. *Elife.* 2017;6:e23383.
117. Ioannidis J, Stanley T, Doucouliagos H. The power of bias in economics research. *Econ J.* 2017;127(605):F236-F265.
118. Jennions MD, Møller AP. A survey of the statistical power of research in behavioral ecology and animal behavior. *Behav Ecol.* 2003;14(3):438–445.
119. Fusar-Poli P, Stringer D, M S Durieux A, et al. Clinical-learning versus machine-learning for transdiagnostic prediction of psychosis onset in individuals at-risk. *Transl Psychiatry.* 2019;9(1):259.
120. Fusar-Poli P, Oliver D, Spada G, et al. Real world implementation of a transdiagnostic risk calculator for the automatic detection of individuals at risk of psychosis in clinical routine: study protocol. *Front Psychiatry.* 2019;10:109.
121. Wang T, Oliver D, Msosa Y, et al. Implementation of a real-time psychosis risk detection and alerting system based on electronic health records using Cogstack. *J Vis Exp.* 2020(159).
122. Oliver D, Spada G, Colling C, et al. Real-world implementation of precision psychiatry: transdiagnostic risk calculator for the automatic detection of individuals at-risk of psychosis. *Schizophr Res.* 2020;S0920-9964(20)30259-0.
123. Chekroud AM, Koutsouleris N. The perilous path from publication to practice. *Mol Psychiatry.* 2018;23(1):24–25.
124. Rapport F, Clay-Williams R, Churruca K, et al. The struggle of translating science into action: Foundational concepts of implementation science. *J Eval Clin Pract.* 2018;24(1):117–126.
125. Damschroder LJ. Clarity out of chaos: Use of theory in implementation research. *Psychiatry Res.* 2020;283:112461.
126. Kirk MA, Kelley C, Yankey N, et al. A systematic review of the use of the consolidated framework for implementation research. *Implement Sci.* 2016;11:72.