

# Individualized Diagnostic and Prognostic Models for Patients With Psychosis Risk Syndromes: A Meta-analytic View on the State of the Art

Rachele Sanfelici, Dominic B. Dwyer, Linda A. Antonucci, and Nikolaos Koutsouleris

## ABSTRACT

**BACKGROUND:** The clinical high risk (CHR) paradigm has facilitated research into the underpinnings of help-seeking individuals at risk for developing psychosis, aiming at predicting and possibly preventing transition to the overt disorder. Statistical methods such as machine learning and Cox regression have provided the methodological basis for this research by enabling the construction of diagnostic models (i.e., distinguishing CHR individuals from healthy individuals) and prognostic models (i.e., predicting a future outcome) based on different data modalities, including clinical, neurocognitive, and neurobiological data. However, their translation to clinical practice is still hindered by the high heterogeneity of both CHR populations and methodologies applied.

**METHODS:** We systematically reviewed the literature on diagnostic and prognostic models built on Cox regression and machine learning. Furthermore, we conducted a meta-analysis on prediction performances investigating heterogeneity of methodological approaches and data modality.

**RESULTS:** A total of 44 articles were included, covering 3707 individuals for prognostic studies and 1052 individuals for diagnostic studies (572 CHR patients and 480 healthy control subjects). CHR patients could be classified against healthy control subjects with 78% sensitivity and 77% specificity. Across prognostic models, sensitivity reached 67% and specificity reached 78%. Machine learning models outperformed those applying Cox regression by 10% sensitivity. There was a publication bias for prognostic studies yet no other moderator effects.

**CONCLUSIONS:** Our results may be driven by substantial clinical and methodological heterogeneity currently affecting several aspects of the CHR field and limiting the clinical implementability of the proposed models. We discuss conceptual and methodological harmonization strategies to facilitate more reliable and generalizable models for future clinical practice.

**Keywords:** Biomarkers, Clinical psychobiology, Machine learning, Predictive psychiatry, Psychosis, Translational medicine

<https://doi.org/10.1016/j.biopsych.2020.02.009>

Psychotic disorders are among the most disabling mental illnesses and represent one of the top 20% causes of socioeconomic burden worldwide (1). Therefore, psychiatric research has substantially invested in better early detection strategies for these disorders (2). The clinical high risk (CHR) concept (3) describes a mental state characterized by sub-threshold psychotic symptoms that differ quantitatively in their intensity from those of a full-blown psychosis (Supplement and Table 1). The CHR paradigm has become a well-established clinical avenue to early detect and potentially treat the psychosis high-risk states. Based on the CHR paradigm, researchers have investigated the nature of the prepsychotic phase from both pathophysiological and epidemiological perspectives (4,5). However, these efforts have been challenged by a constantly declining incidence rate of psychosis among CHR patients (4,6), with roughly one third of not-transitioned CHR cases still experiencing subthreshold symptoms,

psychosocial impairments (7), and lower level of quality of life (8). Thus, the CHR designation delineates a mental condition that is burdensome per se and, in addition, is associated with a known set of comorbidities (e.g., depression, substance abuse, anxiety disorders) (9). Therefore, predictive psychiatry has gradually broadened its scope from detecting disease transition to encompassing adverse outcomes more broadly [e.g., functional deficits (10), treatment response (11), persisting negative symptoms (12), psychiatric comorbidities (13)].

Considering that clinical CHR instruments alone detect only about 47% of transitions after 3 years (14), efforts have been made to identify potential risk factors for psychosis in several symptomatological and biological readouts, or biomarkers, of the disorder (15) so that individualized prognostication may be enhanced. The presence of environmental adverse events (16), cognitive impairments (17), neuromorphological (18), and electrophysiological (19) and hematological (20) alterations, as well

as resting-state (21) and task-related (22) neural activity and connectivity anomalies, has been consistently reported in people at risk for psychosis compared with healthy individuals. Some of these phenotypes have been associated with both disease course and transition to the overt disease (4). Therefore, the identification of reliable markers able to distinguish between at-risk and healthy populations may be potentially useful in clinical practice to monitor disease development and treatment outcome (23) and to obviate time-consuming CHR assessments. The two prevailing statistical approaches to address the challenge of single-subject prediction are machine learning (ML) methods (e.g., support vector machine, LASSO [least absolute shrinkage and selection operator] regression, random forest), which can handle large databases and different data domains (24,25), and Cox proportional hazard regression, a form of multivariate survival analysis (26) able to investigate time-to-conversion trajectories. Recent research applying these methods has produced prognostic models able to stratify CHR patients into different risk classes according to their pretest risk enrichment (27) or a set of combined predictors (28,29), or to predict patients' functional outcomes based on different data modalities with performance accuracies of up to 83% (10,30). Despite the great potential of these models, their applicability is still hindered by the methodological heterogeneity in the field. Indeed, CHR patients are identified by several clinical instruments and are characterized by subtypes with different levels of risk (14). Moreover, models' generalizability has been assessed through discrepant validation strategies across studies, ranging from the less replicable (i.e., single-site cross-validation [CV]) to the most robust (i.e., validation to external samples) (25). Thus, methodological approaches still lack standardized validation strategies testing clinical applicability under real-world conditions. One way to tackle these issues is to use a meta-analytic approach to quantitatively investigate models' performance across different outcomes, algorithms, and data modalities. Although important contributions to this goal have been made (5,29,31), to the best of our knowledge, the field is still lacking such an analysis. Investigating the field's heterogeneity would allow a comprehensive assessment of accuracy and validity of the existing diagnostic and prognostic models, an important prerequisite for establishing reliable tools for psychosis risk quantification in clinical care.

Our aim was to review the literature on ML-based and Cox regression-based diagnostic models (i.e., discriminating CHR individuals from healthy individuals) and prognostic models (i.e., predictive approaches for transition or negative outcomes). Furthermore, we performed a meta-analysis of models' performance, with the aim of investigating the effects of 1) data modality, 2) type of algorithm, and 3) validation paradigms. We expected that our results would elucidate the complexity of methods and data domains currently used in the predictive analytics arm of CHR research. This will facilitate a deeper understanding of the state of the art within the field and may clarify the bottlenecks impeding clinical translation.

## METHODS AND MATERIALS

### Literature Search

We conducted a systematic search of published original articles in English through June 30, 2019, using a range of search

terms in PubMed and Scopus as well as reference lists of the included articles (Supplement). We selected studies that reported prognostic or diagnostic models constructed using ML or Cox proportional hazard regression. Concerning diagnostic models, we included only those that used healthy control subjects (HCs) as a reference group to enlarge the sample size by selecting comparable classification models across studies. CHR included patients with a psychosis risk syndrome categorized as CHR, ultra high risk (UHR), or at-risk mental states (Table 1) as well as those with a familial risk (FR) or 22q11.2 deletion syndrome (22q11.2DS). Studies were included if measures of performance accuracy were reported (i.e., true positives [TP], false positives [FP], true negatives [TN], and false negatives [FN]) or if they could be extracted. Results of the literature search are illustrated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flowchart (32) (Supplemental Figure S1).

A comprehensive list of all variables extracted by each study is reported in the Supplement (second section). Performance accuracy measures used for analyses comprised TP, FN, TN, FP, sensitivity (SE) [TP/(TP + FN)], and specificity (SP) [TN/(TN + FP)].

### Data Analysis

The meta-analysis of diagnostic models was conducted following previous work (33). Extracted SE and SP were converted to a confusion matrix tabulated across studies. Publication bias was assessed with both overall diagnostic odds ratio and SE. The Deeks *et al.* (34) method was used to account for biases associated with unequal proportions of TP and TN cases (Supplement).

Models were built using the bivariate random effects modeling of Reitsma *et al.* (2005) (35) in the mada R package (version 0.5.8), which permits the analysis of SE and SP separately by explicitly accounting for correlations between each measure, incorporating precision estimates arising from sample size differences (i.e., more precision with higher weight), and modeling normal distributions of each with a random effects approach. This bivariate method was used to produce summary estimates of SE, SP, and confidence intervals (CIs) that were used in forest plots, in addition to the analysis of moderators using mixed modeling. Moderators were age, sex, data modality, algorithm, presence of CV, type of CHR, being a multisite study, and year of publication. For prognostic studies, we also investigated follow-up time and prognostic target. Moderator analyses were conducted if a minimum of 10 models for variable were available to decrease the standard error and maximize power in case of high between-study variance (36) and to control for sample size and CV scheme—the latter factor overlapping with algorithm used. Results were corrected for false discovery rate. Likelihood ratios and diagnostic odds ratios were produced using a Markov chain Monte Carlo approach within the mada toolbox. All analyses were conducted with R (version 3.6.0).

## RESULTS

The systematic literature search detected 881 articles, from which 44 were considered eligible after screening for exclusion criteria, for a total of 12 diagnostic models (Table 2 and

## Meta-analysis of Psychosis Risk Syndrome Prediction Models

**Table 1. Definitions of Different Psychosis Risk Syndromes Commonly Referred to as CHR States and Descriptions of the Abbreviations and Respective Clinical Diagnostic Instruments**

Concept	Description	Instruments
CHR	Clinical high risk: psychosis risk syndrome operationalized by UHR, BS, or both diagnostic criteria	All instruments below
ARMS	At-risk mental state: same as the CHR state	
UHR	Ultra high risk: psychosis risk syndrome described by the fulfillment of APS, BLIP, or GRDS criteria	SIPS, SOPS, CAARMS
APS	Attenuated psychotic symptoms: subthreshold psychotic symptoms	
BLIPS	Brief limited intermittent psychotic symptoms: full-blown psychotic symptoms present for a maximum of a week	
GRDS	Genetic risk and deterioration syndrome: family history of psychosis or schizotypal personality and drop in functioning or sustained low functioning <sup>a</sup>	
BS	Basic symptoms: subjective disturbances of cognitive, affective, and perceptual nature	BSABS
COGDIS	Cognitive disturbances: 9 BS describing disturbances of cognitive nature	SPI-A/SPI-CY
COPER	Cognitive-perceptual symptoms: 10 BS describing disturbances of a cognitive-perceptual nature	
UPS	Unspecific prodromal symptoms: unspecific attenuated symptoms characterizing a low-risk state	BSIP

BSABS, Bonn Scale for the Assessment of Basic Symptoms; BSIP, Basel Screening Instrument for Psychosis; CAARMS, Comprehensive Assessment of the At-Risk Mental State; SIPS, Structured Interview for the Prodromal Syndrome; SOPS, Scale of Prodromal Symptoms; SPI-A/SPI-CY; Schizophrenia Proneness Instrument–Adult version/Schizophrenia Proneness Instrument–Child and Youth version.

<sup>a</sup>Drop in functioning is described 1) in the CAARMS as a Social and Occupational Functioning Assessment Scale (SOFAS) score  $\leq 30\%$  compared with the previous functioning, within the last year, and for at least 1 month and 2) in the SIPS/SOPS as a 30% decrease in the Global Assessment of Functioning scale score from premorbid baseline. A sustained low functioning is defined only in the CAARMS as a SOFAS score  $\leq 50$  in the past year or longer.

Supplemental Figure S1) and 32 prognostic models (Table 3 and Supplemental Figure S1). The final sample comprised 3707 patients for prognostic studies (mean age = 20.41 years; ~58% male), of which 320 (~9%) were CHR patients investigated for nontransition outcomes (mean age = 19.25 years; 56% male) and 1052 were used for diagnostic classification (mean age = 23.42 years; ~59% male), of which 480 (45%) were HCs. In addition, 26 studies used ML (all diagnostic studies) and 18 were conducted with Cox regression (Tables 2 and 3 and Supplemental Table S1).

**Meta-analytic Results**

CHR individuals could be classified against HCs with an overall SE of 78% (95% CI = 73%–83%) and an SP of 77% (95% CI = 68%–84%), while across all prognostic

models SE reached 67% (95% CI = 63%–70%) and SP reached 78% (95% CI = 73%–82%). Prognostic studies showed a publication bias ( $R^2 = .26$ ,  $p < .001$ ), whereas diagnostic studies did not ( $R^2 = .07$ ,  $p > .05$ ) (Supplemental Figure S2). Performances of both models' categories are illustrated in two summary receiving operating characteristic curves (Figures 1 and 2) and forest plots (Figures 3 and 4). Within diagnostic models, moderator effects of type of CHR and algorithm, data modality, presence of CV, and being a multisite study were not investigated because less than 10 models per factor were available (36). We found no effects of moderator variables in either application domain ( $p > .10$ ) (Supplemental Table S2) even when splitting the sample based on CV (Supplement).

**Table 2. Summary of Diagnostic Studies Included in the Current Meta-analysis**

Study	CHR Type	Data Modality	Algorithm	Outcome	SE	FPR
Bendfeldt <i>et al.</i> (37)	UHR, UPS	Biological: fMRI	SVM	Diagnosis	74	0.42
Guo <i>et al.</i> (39)	FR	Biological: fMRI	SVM	Diagnosis	60	0.6
Koutsouleris <i>et al.</i> (41)	UHR, BS	Clinical: cognition	SVM	Diagnosis	96	0.2
Koutsouleris <i>et al.</i> (43)	UHR, BS	Biological: sMRI	SVM	Diagnosis	89	0.2
Liu <i>et al.</i> (53)	FR	Biological: fMRI	SVM	Diagnosis	72	0.14
Pettersson-Yeo <i>et al.</i> (47)	UHR	Biological: sMRI	SVM	Diagnosis	80	0.27
Scariati <i>et al.</i> (48)	22q11.2DS	Biological: fMRI	SVM	Diagnosis	81	0.12
Studerus <i>et al.</i> (99)	UHR, UPS	Clinical: cognition	Random forest	Diagnosis	73	0.23
Tylee <i>et al.</i> (49)	22q11.2DS	Biological: DTI	SVM	Diagnosis	85	0.18
Valli <i>et al.</i> (50)	UHR	Biological: sMRI	SVM	Diagnosis	68	0.24
Wang <i>et al.</i> (42)	UHR	Biological: fMRI	SVM	Diagnosis	82	0.31
Zhu <i>et al.</i> (46)	UHR	Biological: fMRI	SVM	Diagnosis	72	0.53

22q11.2DS, 22q11.2 deletion syndrome; BS, basic symptoms; CHR, clinical high risk; DTI, diffusion tensor imaging; fMRI, functional magnetic resonance imaging; FPR, false positive rate; FR, familial risk; SE, sensitivity; sMRI, structural magnetic resonance imaging; SVM, support vector machine; UHR, ultra high risk; UPS, unspecific prodromal symptoms.

**Table 3. Summary of Prognostic Studies Included in the Current Meta-analysis**

Study	CHR Type	Data Modality	Algorithm	Outcome	SE	FPR
Amminger <i>et al.</i> (11)	UHR	Biological: lipids	GPC	Functioning	83	0.25
Bedi <i>et al.</i> (54)	UHR	Clinical: speech	Convex Hull	Transition	100	0
Buchy <i>et al.</i> (100)	UHR	Clinical: substance use	Cox regression	Transition	69	0.19
Cannon <i>et al.</i> (101)	UHR	Clinical: symptoms, family risk, functioning	Cox regression	Transition	67	0.47
Cannon <i>et al.</i> (28)	UHR	Multimodal: symptoms, environment, genetic, cognition	Cox regression	Transition	67	0.28
Carrion <i>et al.</i> (74)	UHR	Multimodal: symptoms, environment, genetic, cognition	Cox regression	Transition	58	0.27
Chan <i>et al.</i> (56)	UHR, UPS	Biological: serum	LASSO regression	Transition	89	0.34
		Clinical: positive symptoms			78	0.4
		Multimodal: serum, symptoms			89	0.21
Cornblatt <i>et al.</i> (58)	UHR	Multimodal: clinical, demographics, cognition	Cox regression	Transition	60	0.03
Das <i>et al.</i> (55)	UHR, UPS	Biological: cortical gyrification	Randomized trees	Transition	66	0.03
de Wit <i>et al.</i> (30)	UHR, BS	Biological: sMRI, gyrification	SVM	Functioning	67	0.25
		Clinical: disorganized speech			76	0.25
		Multimodal: sMRI, clinical, combination			68	0.19
DeVylder <i>et al.</i> (102)	UHR	Clinical: disorganized communication	Cox regression	Functioning	58	0.4
Dragt <i>et al.</i> (64)	UHR	Clinical: disorganized communication	Cox regression	Transition	50	0.09
Francesconi <i>et al.</i> (59)	UHR	Clinical: thought content, ToM, processing, NSS	Cox regression	Transition	67	0.03
Fusar-Poli <i>et al.</i> (60)	UHR-BLIPS	Clinical: disorganizing symptoms	LASSO Cox regression	Transition	24	0.37
Gothelf <i>et al.</i> (38)	22q11.2DS	Biological: sMRI	SVM	Transition	90	0
Hoffman <i>et al.</i> (73)	UHR	Clinical: cognition	Cox regression	Transition	89	0.11
Kambeitz-Ilankovic <i>et al.</i> (40)	UHR, BS	Biological: cortical surface area	SVM	Functioning	79	0.15
Koutsouleris <i>et al.</i> (41)	UHR, BS	Clinical: cognition	SVM	Transition	80	0.25
Koutsouleris <i>et al.</i> (10)	UHR, BS	Biological: sMRI	SVM	Functioning (role)	67	0.53
		Clinical: functioning			61	0.25
		Multimodal: sMRI and functioning			59	0.3
Koutsouleris <i>et al.</i> (10)	UHR, BS	Biological: sMRI	SVM	Functioning (social)	80	0.28
		Clinical: functioning			70	0.16
		Multimodal: sMRI and functioning			83	0.18
Koutsouleris <i>et al.</i> (44)	UHR, BS	Biological: sMRI	SVM	Transition	76	0.15
Lavoie <i>et al.</i> (71)	UHR	Biological: blood antioxidant	Cox regression	Transition	91	0.33
Mechelli <i>et al.</i> (45)	UHR	Clinical: disorders of thought content, attenuated positive symptoms, functioning	SVM	Transition	69	0.39
				Functioning	63	0.37
Michel <i>et al.</i> (61)	UHR, BS	Clinical: SIPS, SPI-A, cognition	Cox regression	Transition	57	0.45
Nieman <i>et al.</i> (62)	UHR, BS	Multimodal: symptoms and ERPs	Cox regression	Transition	78	0.12
Perkins <i>et al.</i> (20)	UHR	Biological: blood plasma analytes	Greedy algorithm	Transition	60	0.1
Ramyead <i>et al.</i> (57)	UHR, UPS	Biological: EEG	LASSO	Transition	58	0.17
Ruhrmann <i>et al.</i> (65)	UHR, BS	Clinical: symptoms, sleep, schizotypy, functioning, education	Cox regression	Transition	42	0.02
Tarbox <i>et al.</i> (66)	UHR	Clinical: alogia, anhedonia/asociality, suspiciousness	Cox regression	Transition	62	0.39
Thompson <i>et al.</i> (67)	UHR	Clinical: unusual thought content, functioning, family history, functional decline	Cox regression	Transition	30	0.11
van Tricht <i>et al.</i> (68)	UHR	Biological: EEG	Cox regression	Transition	46	0.13
van Tricht <i>et al.</i> (72)	UHR, BS	Biological: EEG	Cox regression	Transition	83	0.21

**Table 3. Continued**

Study	CHR Type	Data Modality	Algorithm	Outcome	SE	FPR
Zarogianni <i>et al.</i> (51)	UHR, BS	Multimodal: sMRI and cognition	SVM	Transition	63	0.16
Zarogianni <i>et al.</i> (52)	FR	Biological: sMRI	SVM	Transition	76	0.23
		Multimodal: sMRI and cognition			100	0.17

22q11.2DS, 22q11.2 deletion syndrome; BLIPS, brief limited intermittent psychotic symptoms; BS, basic symptoms; CHR, clinical high risk; EEG, electroencephalography; ERP, evoked response potential; FPR, false positive rate; FR, familial risk; GPC, Gaussian process classification; LASSO, least absolute shrinkage and selection operator; NSS, neurological soft signs; SE, sensitivity; SIPS, Structured Interview for Prodromal Syndromes; sMRI, structural magnetic resonance imaging; SPI-A, Schizophrenia Proneness Instrument–Adult version; SVM, support vector machine; ToM, theory of mind; UHR, ultra high risk; UPS, unspecific prodromal symptoms.

### Effect of Algorithm Choice

A total of 19 ML studies (73%) employed a support vector machine algorithm (10,30,37–53), while the rest used Gaussian process (11) or convex hull classification (54), randomized trees (55), greedy algorithm (20), random forest (5), or LASSO regression (56,57). All ML models were computed with CV, whereas studies using Cox regression applied bootstrapping (28,58–62), reported apparent results (i.e., the model is tested in the same sample from which it was derived) (63–68), or lacked a validation procedure. Among the cross-validated studies, 58% applied leave-one-out CV, 3 of which nested and 7 of which used  $k$ -fold CV (3 in its repeated nested form). Only 1 study applied a leave-site-out CV (10), that is, a form of internal–external validation (69). Within prognostic studies, we found a main effect of CV/algorithm on SE ( $p = .009$ ;  $\chi^2_2 = 6.96$ ,  $p = .031$ ); that is, cross-validated ML models reached a higher SE (71%, 95% CI = 67%–74%) than Cox regression ones (61%, 95% CI = 54%–68%) (Figure 4).

### Effect of Data Modality

Diagnostic models included the use of functional (37,39,43,47,48) and structural (46,50,70) magnetic resonance imaging (MRI) and diffusion tensor imaging (49), and behavioral models were based on neurocognitive functions (42,43).

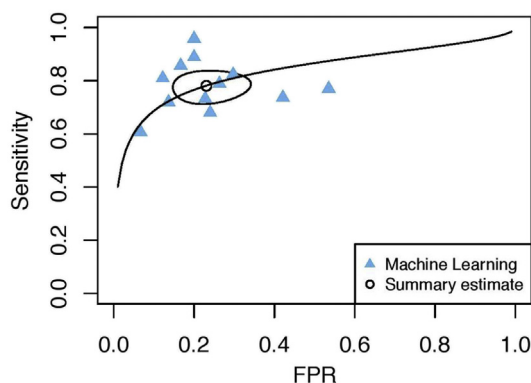
Models for prediction of transition to psychosis involved blood-based (20,56,71), electrophysiological (57,68,72), and neuroanatomical data using white and/or gray matter volume (38,44,51) or gyrification measures (55). Clinical models were trained on prodromal positive and negative symptoms,

functioning, and family risk associated with functional decline; the neurocognitive modality was based on executive functions and verbal IQ (41) or speech features (54,73). Multimodal approaches included different combinations of clinical, neuropsychological, and demographic variables as well as genetic risk (28,51,52,54,74). One model was built on P300 amplitude from event-related potentials and sociopersonal adjustment (62). Functional outcomes were predicted with neuroanatomical (63,9,19) and blood-based biomarkers (11), and 2 studies combined clinical and MRI measures (10,30). There were no effects of data modality on SE ( $p = .172$ ) or false positive rate ( $p = .606$ ) (Supplemental Table S2).

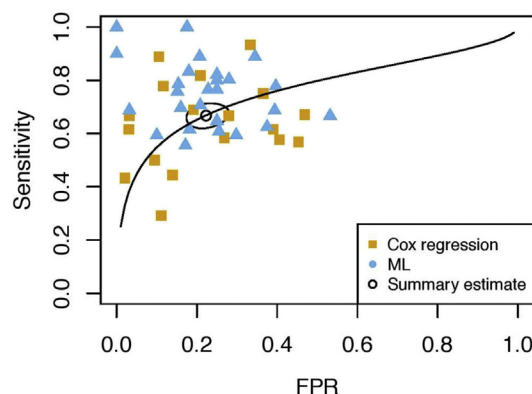
### Effect of Sample Characteristics

Performance accuracies were not influenced by age and sex of individuals ( $p > .10$ ) (Supplemental Table S2). CHR in 86% of the studies fulfilled the UHR criteria (75), while 6 models were based on the genetic risk syndromes 22q11.2DS (38,48,49) or FR (39,52,53). Because of this imbalance, we could not statistically test the effects of this variable, yet results did not change when excluding patients with 22q11.2DS and FR (Supplement).

Furthermore, individuals differed in their outcome definitions. Poor functional outcome was defined on the Global Assessment of Functioning scale (GAF) (cutoff: 70) (40), the Social and Occupational Functioning Assessment Scale (score  $\leq 50$ ) (45), the GAF modified version (76) defining nonresilience through a cutoff of  $\leq 65$  (30), or the Global Functioning social/role scale ( $< 8$ ) (10). In one case (11), treatment response was operationalized as an increase of  $\geq 15$  points in the GAF. There

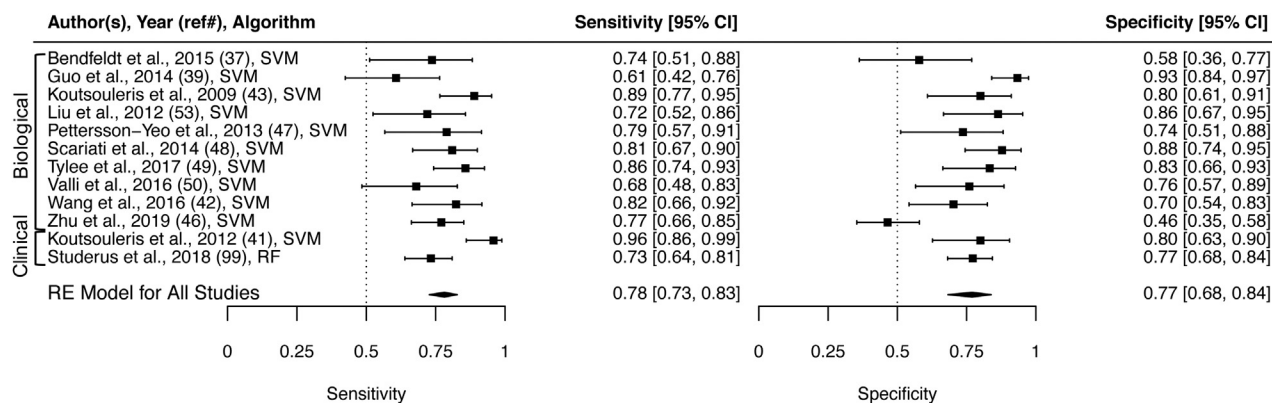


**Figure 1.** Summary receiver operating characteristic curve of diagnostic studies. FPR, false positive rate.



**Figure 2.** Summary receiver operating characteristic curve of prognostic studies. FPR, false positive rate; ML, machine learning.





**Figure 3.** Forest plot of sensitivity and specificity for all diagnostic studies divided by data modality. CI, confidence interval; RE, random effects; RF, random forest; SVM, support vector machine.

were no significant effects on SE or false positive rate driven by prognostic target ( $p = .570$  or  $.085$ , respectively) or the duration of time-to-follow-up examination ( $p = .637$  or  $.305$ , respectively).

## DISCUSSION

We conducted a systematic review and meta-analysis on 44 studies reporting prognostic and diagnostic models for a total of 3707 and 572 CHR individuals, respectively, with the aim to quantitatively assess their accuracy, validity, and heterogeneity. Our results point to good model performance overall and to a higher SE of ML models compared with Cox regression in prognostic studies. This effect was fully collinear with that of CV, mainly due to the complete overlap of this factor with algorithm type. Notably, there were no significant effects of data modality, CHR or CV type, prognostic target, or any other potential confounding variables (e.g., age distribution, sex, year of publication, follow-up interval time) on accuracy performance in our data. It is noteworthy that in prognostic studies we observed a publication bias, that is, the tendency for studies with smaller sample sizes to report higher, and potentially inflated, prediction accuracies (77). This might have affected our results (77) so that we cannot draw robust conclusions from our meta-analytical findings.

### Methodological Differences and Pitfalls

Prognostic models employing ML outperformed those using Cox regression by 10% SE. This finding may have resulted from a complex interplay of cohort-related and methodological heterogeneity. Notably, there was a complete overlap between the statistical method chosen and implementation of CV, that is, all ML models were cross-validated, while only 6 Cox regression studies applied bootstrapping as the validation procedure. Because the choice of a reliable validation method strongly determines both performance and generalizability of models (25), this methodological discrepancy may have biased our findings. Validation issues were also present in studies employing ML for prognostic modeling. First, 53% of these studies applied CV without nesting and repetitions, which is known to generate overoptimistic results due to high variability

and information leakage between training and testing data during model optimization (78). The extended use of this validation scheme may explain the higher SE found in ML studies.

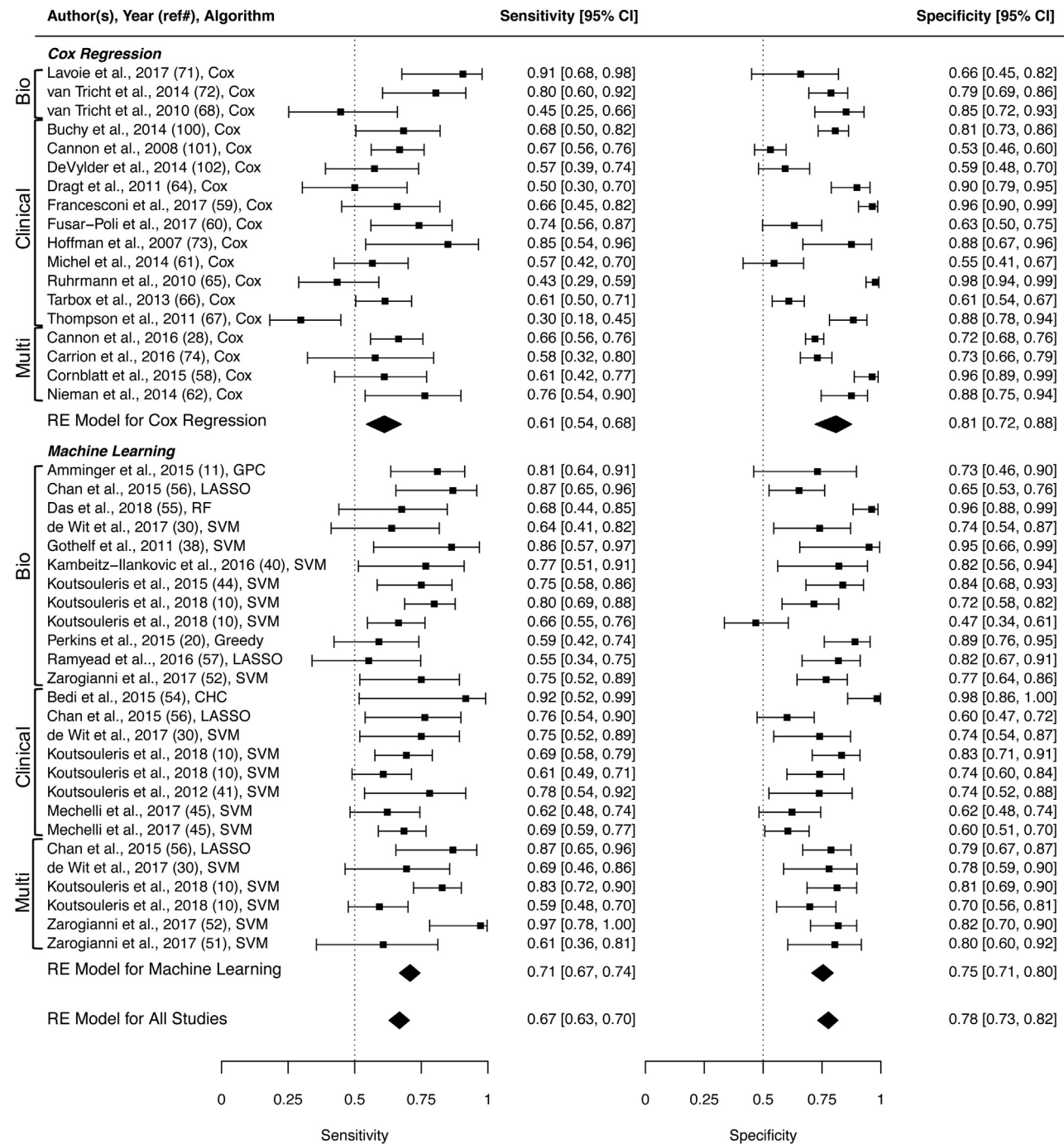
Second, several Cox regression studies included in this meta-analysis either did not report probability thresholds or chose a priori optimal thresholds from the data. While ML's lack of homogeneous thresholds is mainly handled via CV schemes averaging performances across folds and repetitions, the use of  $p$  values or data-derived thresholds without a proper training-test separation might have inflated Cox regression models' performance (63).

Third, preprocessing approaches varied across studies. In 3 cases, for instance, prognostic features were derived from univariate group comparisons or by applying principal component analysis outside the CV scheme (20,43,53), which is a known source of information leakage, because variance from the training sample data is carried into the test sample (25). One model was constructed on a nonrandom sampling of the training set (49), while another model classified patients at UHR from HCs based on the brain pattern shared by patients at UHR and with first-episode psychosis (46). These approaches, as well as the use of stepwise methods in Cox regression models, entail sample-driven variance and, therefore, could lead to good predictive performance, but arguably they should be tested for generalizability in an external dataset. Valuable alternatives are literature-based feature selection and embedded feature optimization, where the intrinsic optimal feature configuration is learned by the model itself (79).

It should be noted that some of the studies included in our meta-analytic contribution had very low sample sizes. One study had  $N < 20$ , while 2 diagnostic and 21 prognostic models had, respectively, less than 20 CHR individuals or CHR with poor outcome. Findings from these studies might be consistent with literature demonstrating a publication bias toward increased accuracy with reduced sample size (80), possibly caused by overfitting. This indicates the need for future ML research to employ larger, preferably multisite samples for both diagnostic and prognostic purposes (80).

Taken together, these issues may mirror the heterogeneity of methodological procedures within the field. Arguably, the

Meta-analysis of Psychosis Risk Syndrome Prediction Models



**Figure 4.** Forest plot of sensitivity and specificity for all prognostic studies divided by algorithm and data modality. CHC, convex hull classification; CI, confidence interval; GPC, Gaussian process classifier; LASSO, least absolute shrinkage and selection operator regularized regression; RE, random effects; RF, random forest; SVM, support vector machine.

application of ML techniques to diagnosis and prognosis in psychiatry is still relatively young (24), so conventions and standard operating procedures facilitating model comparability and replicability have not become generally accepted. Our findings highlight the urgency to develop such guidelines for

the construction of prognostic and diagnostic models (81). As indicated in Table 4, the most important ones are 1) the implementation of repeated nested CV, internal-external, or external validation schemes and 2) the full and strict embedding of all preprocessing or feature engineering procedures

**Table 4. Conceptual and Methodological Guidelines for Construction of Diagnostic and Predictive Models Implementable in Real-Life Clinical Practice**

Guidelines	Practical Suggestions
<b>Conceptual Guidelines</b>	
Harmonization of the CHR definition and diagnostic instruments	Create a harmonized early recognition instrument that encompasses those at-risk definitions and criteria from the existing diverse inventories that parsimoniously delineate the CHR state and also are predictive of its adverse outcomes
Broaden the scope of prediction to nontransition outcomes	Harmonize social and occupational outcomes, pharmacological and nonpharmacological treatment response criteria, and definitions of persistence or remission of symptoms and use these end points in future predictive studies
<b>Methodological Guidelines</b>	
Increase in sample size	Facilitate collaborative science approaches that enable the harmonization of end-point definitions and the external validation of predictive models Get access to open-source databases
Study design harmonization	Employ reliable methodologies (CV and external validation are recommended); avoid leave-one-out CV; implement <i>k</i> -fold CV Embed all preprocessing or feature engineering procedures within the chosen CV scheme Enforce preregistration processes (as in clinical trials) to facilitate monitoring of standardized data acquisition, model discovery, and validation plan
Common modeling platforms and open-source model libraries	Large-scale, consortium-wide international model benchmarking

CHR, clinical high risk; CV, cross-validation.

within the CV scheme. Researchers, funding organizations, and journals should support efforts to standardize approaches, favoring the importance of thorough validation over model performance per se.

### Type of Data Modality

Overall, most models were constructed using biological (44%) and clinical (38%) data, with only 10 prognostic models based on more than one data modality. Most diagnostic models used MRI data (83%), whereas prognostic models showed a higher variability. Prognostic models of psychosis transition included molecular, neuroanatomical, electrophysiological, neuropsychological, and clinical data modalities, most of the latter trained on prodromal positive and negative symptoms, functioning, and FR associated with functional decline. We found no significant differences in predictive accuracy when comparing data modalities within and between algorithms.

This result may mirror a real lack of significant differences in biomarker type when distinguishing the CHR state from the norm or predicted outcome. However, because only 4 prognostic studies tested the relative and combined predictive ability of different data modalities on the same individuals (10,30,52,56), and because data modalities are overall under- or overrepresented, the currently available studies do not allow this conclusion to be drawn. Further research directly comparing performance across data modalities, followed by meta-analytic evaluation, is warranted.

Alternatively, our results may reflect the complexity of the multifaceted architecture of psychosis risk (82), which might be only partly captured by single data modalities. Indeed, a neuroanatomical biomarker might be informative for genetically or pathophysiologically driven mechanisms given that genes' effect may be closer to brain than to behavior (83); on the other hand, neurocognitive performance might explain more environmentally driven variance relating, for example, to socioeconomic status (84). Hence, a multimodal approach may be a

viable way to reconcile and leverage information from single risk domains. Powerful new methodologies able to combine multiple sources of data, such as similarity network fusion (85), might be suitable for this purpose. Indeed, research has shown that a combination of clinical variables and structural brain imaging data might represent a promising multimodal framework for psychosis prediction (10,23,31). Along these lines, Schmidt *et al.* (29) devised a 3-stage sequential testing paradigm, which in theory reaches nearly perfect positive predictive value when individuals are tested on one multimodal modality (i.e., clinical and electroencephalography) and two biological data modalities (i.e., structural MRI and blood based). However, these findings are simulated, have not been confirmed in empirical studies yet, and did not follow a thorough meta-analytical approach like the one implemented here.

Alternatively, similar performance of tested data modalities may have resulted from the variability induced by higher-order algorithm–data validation interactions. To thoroughly compare models originating from different data spaces, methodological consensus guidelines are urgently needed in the precision psychiatry field. A strict cross-study standardization, in terms of both data definitions and algorithm implementations, may shed light on real phenotypic and neurobiological differences and thus lead to unique insights into the pathology of emerging psychosis.

### At-Risk State/Sample Differences

Another source of heterogeneity affecting our results may be due to clinical sample definitions. Most of the at-risk individuals in our sample fulfilled the UHR criteria, while a minority (5.7%) had an FR or a 22q11.2DS diagnosis, which prevented us from quantitatively estimating the effects of risk group designation. However, it is noteworthy that two of the instruments operationalizing UHR criteria (i.e., SIPS [Structured Interview for Prodromal Syndromes] and CAARMS [Comprehensive Assessment of At-Risk Mental States])



## Meta-analysis of Psychosis Risk Syndrome Prediction Models

include a genetic risk group (i.e., the genetic risk deterioration syndrome) and that two studies in our sample included FRs and deletion syndrome patients with subthreshold psychotic symptoms (49,52). This diagnostic overlap might create, on the one hand, a further source of variability and, on the other, a tangible bridge to the well-known heterogeneity among CHR individuals. This issue was tackled by a recent study (14) that provided evidence of a differential risk level within the subcategories of the CHR construct. Hence, further research should put effort into revising the CHR paradigm toward a more parsimonious definition based on one gold-standard clinical instrument and clear-cut biological underpinnings.

Furthermore, in our sample, criteria to define transition to psychosis or poor functional outcome differed both in their operationalization and in the threshold used within a specific diagnostic instrument. Another issue in the variability of outcome definition is dichotomization of continuous variables such as GAF and global functioning, which has proven to be a potential source of bias in prognostic models (63,86). It is noteworthy that 1 study (45) addressed this point by conducting an additional analysis to investigate the continuous nature of functioning by using a support vector regression algorithm. The predictability of nontransition outcomes in at-risk individuals is still relatively unexplored. Therefore, there is a need for clinical consensus on relevant nontransition outcomes and how they should be assessed. Additionally, adopting adaptive risk models, which capture the high extent of variability of symptoms and risk factors over time (87), may tackle this complexity and provide more precise measurements of developing negative outcomes, as proposed by digital phenotyping approaches (88).

Notably, CHR populations differ not only in their clinical picture but also along demographic and sociocultural dimensions (89). For instance, American CHR individuals are usually younger (~16–18 years) than their European counterparts (~22–24 years). Interestingly, recent research has shown that neuroanatomical development and risk for developing psychosis are interconnected (90,91). This evidence might also reveal neurobiological processes leading to neurocognitive changes in the CHR state (92,93). Overall, our findings suggest that the gestalt of the CHR state might be successfully modeled only if multiple behavioral and neurobiological moderators are conjointly considered using standardized multivariate methods, thereby fully embracing the complexity of this risk paradigm.

### Limitations

Our meta-analysis was driven by the primary aim to evaluate the potential applicability of diagnostic and prognostic models in real-life clinical practice. Therefore, we focused only on the two currently prevailing methodological approaches (i.e., ML and Cox regression). Importantly, we might have missed significant results by excluding other more traditional statistical methods such as logistic regression (15,63), which has often been implemented for prognostic purposes (15,63), eventually showing higher performance than ML (94). Nevertheless, ML approaches enable the investigation of the intrinsic complexity

of specific data types (e.g., brain features) and are devised for better generalizability.

Another limitation might be the lack of investigation into symptomatology, treatment, substance use, or additional comorbidities, which was due to missing or inconsistent information for several studies. Indeed, already in patients with first-episode psychosis, antipsychotic treatment has been shown to have neuroanatomical effects (95), and continuous cannabis use has been shown to lead to worse outcomes (96). It is also plausible that the high variability of symptoms and clinical comorbidities in the CHR population (13) has further introduced spurious variance in our analyses.

Furthermore, the CHR paradigm has proven to have intrinsic limitations. On the one hand, its predictive power might be partly driven by the so-called pretest risk enrichment; that is, the assessment of at-risk criteria in a specific constellation of help-seeking individuals (97,98). On the other hand, it might not capture the full extent of risk in the population, as a recent study pointed out by reporting that most transitions occurred in patients with an unclear psychiatric diagnosis or no CHR status (9). Because most prognostic models have been developed for the CHR state, their usefulness outside of this category should be intensively investigated.

Lastly, given the heterogeneity of our data and the publication bias detected, our meta-analysis is inherently limited to a description of, not an ultimate decision on, which diagnostic and prognostic models are sufficiently reliable to be applied in clinical settings.

### Conclusions

A comprehensive paradigm shift is required to enable the clinical application of diagnostic and prognostic models for the CHR state. First, the field requires study design harmonization, which demands reliable methodological approaches such as CV or external validation to ensure generalizability. An approach to enhance the studies' potential for real-life implementation could be a preregistration process similar to clinical trials, during which their validity in terms of standardized data acquisition, model discovery, and validation could be monitored. Furthermore, large-scale international model benchmarking at the level of external model validation can be achieved only by constructing common modeling platforms and open source model libraries. The National Institute of Mental Health's Harmonization of At-Risk Multisite Observational Networks for Youth (HARMONY) is a first step in the above direction. Consortium-wise coordinated work will also allow strategic methodological testing; that is, controlled comparison of algorithms, preprocessing and feature optimization pipelines, and multiple data modalities (for an overview of conceptual and methodological guidelines, see Table 4). Multimodal ML carries the challenging responsibility to better disentangle the complex architecture of psychosis risk within a clinical consensus environment. This should involve efforts to unify the CHR definition, both theoretically and practically, and also to embrace relevant nontransition outcomes to broaden the prognostic scope. Future studies are warranted to investigate whether harmonizing procedures within precision psychiatry will lead to more reliable and reproducible translational research in the field.

## ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by a EU-FP7-HEALTH grant for the project “PRONIA” (Personalized Prognostic Tools for Early Psychosis Management) (Grant No. 602152) and by the National Institute of Mental Health (NIMH) for the project “HARMONY” (Harmonization of At Risk Multisite Observational Networks for Youth) (Grant No. MH081928). PRONIA, BMBF (Federal Ministry of Education and Research), and the Max Planck Society funded RS.

NK received honoraria for two lectures from Otsuka. He has a patent issued related to adaptive pattern recognition for psychosis risk modeling (U.S. patent 20160192889A1). RS received honoraria for one lecture from Lundbeck. The other authors report no biomedical financial interests or potential conflicts of interest.

## ARTICLE INFORMATION

From the Department of Psychiatry and Psychotherapy (RS, DBD, LAA, NK), Ludwig-Maximilian-University Munich, Max Planck Institute of Psychiatry Munich (NK), Munich, and Max Planck School of Cognition (RS), Leipzig, Germany; and Department of Education, Psychology, and Communication (LAA), University of Bari “Aldo Moro,” Bari, Italy.

RS and DBD contributed equally to this work.

Address correspondence to Nikolaos Koutsouleris, M.D., Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Nussbaumstr. 7, D-80336 Munich, Germany; E-mail: [nikolaos.koutsouleris@med.uni-muenchen.de](mailto:nikolaos.koutsouleris@med.uni-muenchen.de).

Received Sep 30, 2019; revised Jan 25, 2020; accepted Feb 6, 2020.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.biopsych.2020.02.009>.

## REFERENCES

- Vigo D, Thornicroft G, Atun R (2016): Estimating the true global burden of mental illness. *Lancet Psychiatry* 3:171–178.
- Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW (2018): The science of prognosis in psychiatry: A review. *JAMA Psychiatry* 75:1280–1288.
- Fusar-Poli P, Borgwardt S, Bechdolf A, Addington J, Riecher-Rössler A, Schultze-Lutter F, *et al.* (2013): The psychosis at risk state: A comprehensive state-of-the-art review. *JAMA Psychiatry* 70:107–120.
- Riecher-Rössler A, Studerus E (2017): Prediction of conversion to psychosis in individuals with an at-risk mental state: A brief update on recent developments. *Curr Opin Psychiatry* 30:209–219.
- Studerus E, Rameyead A, Riecher-Rössler A (2017): Prediction of transition to psychosis in patients with a clinical high risk for psychosis: A systematic review of methodology and reporting. *Psychol Med* 47:1163–1178.
- Fusar-Poli P, Bonoldi I, Yung AR, Borgwardt S, Kempton MJ, Valmaggia L, *et al.* (2012): Predicting psychosis. *Arch Gen Psychiatry* 69:220–229.
- Beck K, Andreou C, Studerus E, Heitz U, Ittig S, Leanza L, Riecher-Rössler A (2019): Clinical and functional long-term outcome of patients at clinical high risk (CHR) for psychosis without transition to psychosis: A systematic review. *Schizophr Res* 210:39–47.
- Fusar-Poli P, Rocchetti M, Sardella A, Avila A, Brandizzi M, Caverzasi E, *et al.* (2015): Disorder, not just state of risk: Meta-analysis of functioning and quality of life in people at high risk of psychosis. *Br J Psychiatry* 207:198–206.
- Fusar-Poli P, Rutigliano G, Stahl D, Davies C, Bonoldi I, Reilly T, McGuire P (2017): Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry* 74:493–500.
- Koutsouleris N, Kambeitz-Illankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, *et al.* (2018): 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* 75:1156–1172.
- Amminger GP, Mechelli A, Rice S, Kim SW, Klier CM, McNamara RK, *et al.* (2015): Predictors of treatment response in young people at ultra-high risk for psychosis who received long-chain omega-3 fatty acids. *Transl Psychiatry* 5:3–9.
- Yung AR, Nelson B, McGorry PD, Wood SJ, Lin A (2019): Persistent negative symptoms in individuals at ultra high risk for psychosis. *Schizophr Res* 206:355–361.
- Rutigliano G, Valmaggia L, Landi P, Frascarelli M, Cappucciati M, Sear V, *et al.* (2016): Persistence or recurrence of non-psychotic comorbid mental disorders associated with 6-year poor functional outcomes in patients at ultra high risk for psychosis. *J Affect Disord* 203:101–110.
- Fusar-Poli P, Cappucciati M, Borgwardt S, Woods SW, Addington J, Nelson B, *et al.* (2016): Heterogeneity of psychosis risk within individuals at clinical high risk: A meta-analytical stratification. *JAMA Psychiatry* 73:113–120.
- Addington J, Farris M, Stowkowy J, Santesteban-Echarri O, Metzack P, Kalathil MS (2019): Predictors of transition to psychosis in individuals at clinical high risk. *Curr Psychiatry Rep* 21:39.
- Fusar-Poli P, Tantardini M, De Simone S, Ramella-Cravaro V, Oliver D, Kingdon J, *et al.* (2017): Deconstructing vulnerability for psychosis: Meta-analysis of environmental risk factors for psychosis in subjects at ultra high-risk. *Eur Psychiatry* 40:65–75.
- Seidman LJ, Shapiro DI, Stone WS, Woodberry KA, Ronzio A, Cornblatt BA, *et al.* (2016): Association of neurocognition with transition to psychosis: Baseline functioning in the second phase of the North American Prodrome Longitudinal Study. *JAMA Psychiatry* 73:1239–1248.
- Gifford G, Crossley N, Fusar-Poli P, Schnack HG, Kahn RS, Koutsouleris N, *et al.* (2017): Using neuroimaging to help predict the onset of psychosis. *NeuroImage* 145:209–217.
- Perez VB, Woods SW, Roach BJ, Ford JM, McGlashan TH, Srihari VH, Mathalon DH (2014): Automatic auditory processing deficits in schizophrenia and clinical high-risk patients: Forecasting psychosis risk with mismatch negativity. *Biol Psychiatry* 75:459–469.
- Perkins DO, Jeffries CD, Addington J, Bearden CE, Cadenhead KS, Cannon TD, *et al.* (2015): Towards a psychosis risk blood diagnostic for persons experiencing high-risk symptoms: Preliminary results from the NAPLS project. *Schizophr Bull* 41:419–428.
- Anticevic A, Haut K, Murray JD, Repovs G, Yang GJ, Diehl C, *et al.* (2015): Association of thalamic dysconnectivity and conversion to psychosis in youth and young adults at elevated clinical risk. *JAMA Psychiatry* 72:882–891.
- Antonucci LA, Penzel N, Pergola G, Kambeitz-Illankovic L, Dwyer D, Kambeitz J, *et al.* (2020): Multivariate classification of schizophrenia and its familial risk based on load-dependent attentional control brain functional connectivity. *Neuropsychopharmacology* 45:613–621.
- Antonucci LA, Pergola G, Pigoni A, Dwyer D, Kambeitz-Illankovic L, Penzel N, *et al.* (2019): A pattern of cognitive deficits stratified for gene-and-environment risk reliably classifies patients with schizophrenia from healthy controls [published online ahead of print Nov 15]. *Biol Psychiatry*.
- Bzdok D, Meyer-Lindenberg A (2018): Machine learning for precision psychiatry: Opportunities and challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging* 3:223–230.
- Dwyer DB, Falkai P, Koutsouleris N (2018): Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol* 14:91–118.
- Cox DR (1972): Regression models and life-tables. *J R Stat Soc Ser B: Methodological* 34:187–220.
- Fusar-Poli P, Rutigliano G, Stahl D, Schmidt A, Ramella-Cravaro V, Hitesh S, McGuire P (2016): Deconstructing pretest risk enrichment to optimize prediction of psychosis in individuals at clinical high risk. *JAMA Psychiatry* 73:1260–1267.
- Cannon TD, Yu C, Addington J, Bearden CE, Cadenhead KS, Cornblatt BA, *et al.* (2016): An individualized risk calculator for research in prodromal psychosis. *Am J Psychiatry* 173:980–988.
- Schmidt A, Cappucciati M, Radua J, Rutigliano G, Rocchetti M, Dell’Osso L, *et al.* (2017): Improving prognostic accuracy in subjects

## Meta-analysis of Psychosis Risk Syndrome Prediction Models

- at clinical high risk for psychosis: Systematic review of predictive models and meta-analytical sequential testing simulation. *Schizophr Bull* 43:375–388.
30. de Wit S, Ziermans TB, Nieuwenhuis M, Schothorst PF, van Engeland H, Kahn RS, *et al.* (2017): 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* 38:704–714.
  31. Strobl EV, Eack SM, Swaminathan V, Visweswaran S (2012): Predicting the risk of psychosis onset: Advances and prospects. *Early Interv Psychiatry* 6:368–379.
  32. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009): Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6:1000097.
  33. Kambeitz J, Cabral C, Sacchet MD, Gotlib IH, Zahn R, Serpa MH, *et al.* (2017): Detecting neuroimaging biomarkers for depression: A meta-analysis of multivariate pattern recognition studies. *Biol Psychiatry* 82:330–338.
  34. Deeks JJ, Macaskill P, Irwig L (2005): The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 58:882–893.
  35. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH (2005): Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 58:982–990.
  36. Borenstein M, Hedges LV (2009): *Introduction to Meta-analysis*. Hoboken, NJ: John Wiley.
  37. Bendfeldt K, Smieskova R, Koutsouleris N, Klöppel S, Schmidt A, Walter A, *et al.* (2015): Classifying individuals at high-risk for psychosis based on functional brain activity during working memory processing. *NeuroImage Clin* 9:555–563.
  38. Gothelf D, Hoeffl F, Ueno T, Sugiura L, Lee AD, Thompson P, Reiss AL (2011): Developmental changes in multivariate neuroanatomical patterns that predict risk for psychosis in 22q11.2 deletion syndrome. *J Psychiatr Res* 45:322–331.
  39. Guo S, Palaniyappan L, Yang B, Liu Z, Xue Z, Feng J (2014): Anatomical distance affects functional connectivity in patients with schizophrenia and their siblings. *Schizophr Bull* 40:449–459.
  40. Kambeitz-Illankovic L, Meisenzahl EM, Cabral C, von Saldern S, Kambeitz J, Falkai P, *et al.* (2016): Prediction of outcome in the psychosis prodrome using neuroanatomical pattern classification. *Schizophr Res* 173:159–165.
  41. Koutsouleris N, Davatzikos C, Bottlender R, Patschurek-Kliche K, Scheuerecker J, Decker P, *et al.* (2012): Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophr Bull* 38:1200–1215.
  42. Wang S, Wang G, Lv H, Wu R, Zhao J, Guo W (2016): Abnormal regional homogeneity as potential imaging biomarker for psychosis risk syndrome: A resting-state fMRI study and support vector machine analysis. *Sci Rep* 6:27619.
  43. Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, *et al.* (2009): Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry* 66:700–712.
  44. Koutsouleris N, Riecher-Rössler A, Meisenzahl EM, Smieskova R, Studerus E, Kambeitz-Illankovic L, *et al.* (2015): Detecting the psychosis prodrome across high-risk populations using neuroanatomical biomarkers. *Schizophr Bull* 41:471–482.
  45. Mechelli A, Lin A, Wood S, McGorry P, Amminger P, Tognin S, *et al.* (2017): Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. *Schizophr Res* 184:32–38.
  46. Zhu F, Liu Y, Liu F, Yang R, Li H, Chen J, *et al.* (2019): Functional asymmetry of thalamocortical networks in subjects at ultra-high risk for psychosis and first-episode schizophrenia. *Eur Neuropsychopharmacol* 29:519–528.
  47. Petterson-Yeo W, Benetti S, Marquand AF, Dell'Acqua F, Williams SCR, Allen P, *et al.* (2013): Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high-risk and first-episode psychosis at the individual level. *Psychol Med* 43:2547–2562.
  48. Scariati E, Schaer M, Richiardi J, Schneider M, Debbané M, Van De Ville D, Eliez S (2014): Identifying 22q11.2 deletion syndrome and psychosis using resting-state connectivity patterns. *Brain Topogr* 27:808–821.
  49. Tylee DS, Kikinis Z, Quinn TP, Antshel KM, Fremont W, Tahir MA, *et al.* (2017): Machine-learning classification of 22q11.2 deletion syndrome: A diffusion tensor imaging study. *Neuroimage Clin* 15:832–842.
  50. Valli I, Marquand AF, Mechelli A, Raffin M, Allen P, Seal ML, McGuire P (2016): Identifying individuals at high risk of psychosis: Predictive utility of support vector machine using structural and functional MRI data. *Front Psychiatry* 7:52.
  51. Zarogianni E, Storkey AJ, Borgwardt S, Smieskova R, Studerus E, Riecher-Rössler A, Lawrie SM (2019): Individualized prediction of psychosis in subjects with an at-risk mental state. *Schizophr Res* 214:18–23.
  52. Zarogianni E, Storkey AJ, Johnstone EC, Owens DGC, Lawrie SM (2017): Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features. *Schizophr Res* 181:6–12.
  53. Liu M, Zeng L-L, Shen H, Liu Z, Hu D (2012): Potential risk for healthy siblings to develop schizophrenia: Evidence from pattern classification with whole-brain connectivity. *NeuroReport* 23:265–269.
  54. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, *et al.* (2015): Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophr* 1:15030.
  55. Das T, Borgwardt S, Hauke DJ, Harrisberger F, Lang UE, Riecher-Rössler A, *et al.* (2018): Disorganized gyrification network properties during the transition to psychosis. *JAMA Psychiatry* 75:613–622.
  56. Chan MK, Krebs MO, Cox D, Guest PC, Yolken RH, Rahmoune H, *et al.* (2015): Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Transl Psychiatry* 5: e601.
  57. Rameyad A, Studerus E, Kometer M, Uttinger M, Gschwandtner U, Fuhr T, Riecher-Rössler A (2016): Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. *World J Biol Psychiatry* 17:285–295.
  58. Cornblatt BA, Auther A, Mclaughlin D, Olsen RH, John M, Christoph U, *et al.* (2015): Psychosis prevention: A modified clinical high risk perspective from the recognition and prevention (RAP) program. *Am J Psychiatry* 172:986–994.
  59. Francesconi M, Minichino A, Carrión RE, Delle Chiaie R, Bevilacqua A, Parisi M, *et al.* (2017): Psychosis prediction in secondary mental health services: A broad, comprehensive approach to the “at risk mental state” syndrome. *Eur Psychiatry* 40:96–104.
  60. Fusar-Poli P, Cappucciati M, De Micheli A, Rutigliano G, Bonoldi I, Tognin S, *et al.* (2017): Diagnostic and prognostic significance of brief limited intermittent psychotic symptoms (BLIPS) in individuals at ultra high risk. *Schizophr Bull* 43:48–56.
  61. Michel C, Ruhrmann S, Schimmelmann BG, Klosterkötter J, Schultze-Lutter F (2014): A stratified model for psychosis prediction in clinical practice. *Schizophr Bull* 40:1533–1542.
  62. Nieman DH, Ruhrmann S, Dragt S, Soen F, Van Tricht MJ, Koelman JHTM, *et al.* (2014): Psychosis prediction: Stratification of risk estimation with information-processing and premorbid functioning variables. *Schizophr Bull* 40:1482–1490.
  63. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* (2014): Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Med* 11:1001744.
  64. Dragt S, Nieman DH, Veltman D, Becker HE, van de Fliert R, de Haan L, Linszen DH (2011): Environmental factors and social adjustment as predictors of a first psychosis in subjects at ultra high risk. *Schizophr Res* 125:69–76.
  65. Ruhrmann S, Schultze-Lutter F, Salokangas RKR, Heinimaa M, Linszen D, Dingemans P, *et al.* (2010): Prediction of psychosis in

- adolescents and young adults at high risk: Results from the prospective European Prediction of Psychosis Study. *JAMA Psychiatry* 67:241–251.
66. Tarbox SI, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Perkins DO, *et al.* (2013): Premorbid functional development and conversion to psychosis in clinical high-risk youths. *Dev Psychopathol* 25:1171–1186.
  67. Thompson A, Nelson B, Yung A (2011): Predictive validity of clinical variables in the “at risk” for psychosis population: International comparison with results from the North American Prodrome Longitudinal Study. *Schizophr Res* 126:51–57.
  68. Van Tricht MJ, Nieman DH, Koelman JHTM, Van Der Meer JN, Bour LJ, De Haan L, Linszen DH (2010): Reduced parietal P300 amplitude is associated with an increased risk for a first psychotic episode. *Biol Psychiatry* 68:642–648.
  69. Steyerberg EW, Harrell FE (2016): Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 69:245–247.
  70. Koutsouleris N, Gaser C, Bottlender R, Davatzikos C, Decker P, Jäger M, *et al.* (2010): Use of neuroanatomical pattern regression to predict the structural brain dynamics of vulnerability and transition to psychosis. *Schizophr Res* 123:175–187.
  71. Lavoie S, Berger M, Schlägelhofer M, Schäfer MR, Rice S, Kim SW, *et al.* (2017): Erythrocyte glutathione levels as long-term predictor of transition to psychosis. *Transl Psychiatry* 7:6–10.
  72. van Tricht MJ, Ruhrmann S, Arns M, Müller R, Bodatsch M, Velthorst E, *et al.* (2014): Can quantitative EEG measures predict clinical outcome in subjects at clinical high risk for psychosis? A prospective multicenter study. *Schizophr Res* 153:42–47.
  73. Hoffman RE, Woods SW, Hawkins KA, Pittman B, Tohen M, Preda A, *et al.* (2007): Extracting spurious messages from noise and risk of schizophrenia-spectrum disorders in a prodromal population. *Br J Psychiatry* 191:355–356.
  74. Carrión RE, Cornblatt BA, Burton CZ, Tso IF, Auther AM, Adelsheim S, *et al.* (2016): Personalized prediction of psychosis: External validation of the NAPLS-2 psychosis risk calculator with the EDIPPP project. *Am J Psychiatry* 173:989–996.
  75. Yung AR, McGorry PD (1996): The initial prodrome in psychosis: Descriptive and qualitative aspects. *Aust N Z J Psychiatry* 30:587–599.
  76. Hall RCW (1995): Global Assessment of Functioning: A modified scale. *Psychosomatics* 36:267–275.
  77. Sterne JAC, Egger M, Smith GD (2001): Investigating and dealing with publication and other biases in meta-analysis. *Br J Med* 323:101–105.
  78. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B (2017): Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage* 145B:166–179.
  79. Snoek J, Larochelle H, Adams RP (2012): Practical Bayesian optimization of machine learning algorithms [published online ahead of print Jun 13]. *arXiv*.
  80. Schnack HG, Kahn RS (2016): Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. *Front Psychiatry* 7: 50.
  81. Poldrack RA, Huckins G, Varoquaux G (2019): Establishment of best practices for evidence for prediction: A review [published online ahead of print Nov 27]. *JAMA Psychiatry*.
  82. Howes OD, Murray RM (2014): Schizophrenia: An integrated sociodevelopmental-cognitive model. *Lancet* 383:1677–1687.
  83. Bertolino A, Blasi G (2009): The genetics of schizophrenia. *Neuroscience* 164:288–299.
  84. Cuesta MJ, Sánchez-Torres AM, Cabrera B, Bioque M, Merchán-Naranjo J, Corripio I, *et al.* (2015): Premorbid adjustment and clinical correlates of cognitive impairment in first-episode psychosis: The PEPsCog Study. *Schizophr Res* 164:65–73.
  85. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, *et al.* (2014): Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 11:333–337.
  86. Royston P, Moons KGM, Altman DG, Vergouwe Y (2009): Prognosis and prognostic research: Developing a prognostic model. *BMJ* 338:b604.
  87. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, Reininghaus U (2018): Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry* 17:123–132.
  88. Insel TR (2018): Digital phenotyping: A global tool for psychiatry. *World Psychiatry* 17:276–277.
  89. Li H, Shapiro DI, Seidman LJ (2019): Handbook of Attenuated Psychosis Syndrome Across Cultures: International Perspectives on Early Identification and Intervention. Li H, Shapiro DI, Seidman LG, editors. Basel, Switzerland: Springer Nature.
  90. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, *et al.* (2018): Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. *JAMA Psychiatry* 75:960–968.
  91. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, *et al.* (2019): Adding a neuroanatomical biomarker to an individualized risk calculator for psychosis: A proof-of-concept study. *Schizophr Res* 208:41–43.
  92. Kambeitz-Illankovic L, Haas SS, Meisenzahl E, Dwyer DB, Weiske J, Peters H, *et al.* (2019): Neurocognitive and neuroanatomical maturation in the clinical high-risk states for psychosis: A pattern recognition study. *NeuroImage Clin* 21:101624.
  93. Koutsouleris N, Davatzikos C, Borgwardt S, Gaser C, Bottlender R, Frodl T, *et al.* (2014): Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders. *Schizophr Bull* 40:1140–1153.
  94. Perlis RH (2013): A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry* 74:7–14.
  95. Lesh TA, Tanase C, Geib BR, Niendam TA, Yoon JH, Minzenberg MJ, *et al.* (2015): A multimodal analysis of antipsychotic effects on brain structure and function in first-episode schizophrenia. *JAMA Psychiatry* 72:226–234.
  96. Schoeler T, Petros N, Di Forti M, Klamerus E, Foglia E, Murray R, Bhattacharyya S (2017): Poor medication adherence and risk of relapse associated with continued cannabis use in patients with first-episode psychosis: A prospective analysis. *Lancet Psychiatry* 4:627–633.
  97. Fusar-Poli P, Schultze-Lutter F, Cappucciati M, Rutigliano G, Bonoldi I, Stahl D, *et al.* (2016): The dark side of the moon: Meta-analytical impact of recruitment strategies on risk enrichment in the clinical high risk state for psychosis. *Schizophr Bull* 42:732–743.
  98. Fusar-Poli P (2017): Why ultra high risk criteria for psychosis prediction do not work well outside clinical samples and what to do about it. *World Psychiatry* 16:212–213.
  99. Studerus E, Corbisiero S, Mazzariello N, Ittig S, Leanza L, Egloff L, *et al.* (2018): Can neuropsychological testing facilitate differential diagnosis between at-risk mental state (ARMS) for psychosis and adult attention-deficit/hyperactivity disorder (ADHD)? *Eur Psychiatry* 52:38–44.
  100. Buchy L, Perkins D, Woods SW, Liu L, Addington J (2014): Impact of substance use on conversion to psychosis in youth at clinical high risk of psychosis. *Schizophr Res* 156:277–280.
  101. Cannon TD, Cadenhead K, Cornblatt B, Woods SW, Addington J, Walker E, *et al.* (2008): Prediction of psychosis in youth at high clinical risk: A multisite longitudinal study in North America. *Arch Gen Psychiatry* 65:28–37.
  102. DeVlyder JE, Muchomba FM, Gill KE, Ben-David S, Walder DJ, Malaspina D, Corcoran CM (2014): Symptom trajectories and psychosis onset in a clinical high-risk cohort: The relevance of sub-threshold thought disorder. *Schizophr Res* 159:278–283.