Prediction models in first episode psychosis: a systematic review and critical appraisal

Rebecca Lee*, Samuel P Leighton*, Lucretia Thomas, Georgios V Gkoutos, Stephen J Wood, Sarah-Jane H Fenton, Fani Deligianni, Jonathan Cavanagh, Pavan K Mallikarjun

*Joint first authors

Abstract

Background: People presenting with first episode psychosis (FEP) have heterogenous outcomes. More than 40% fail to achieve symptomatic remission. Accurate prediction of individual outcome in FEP could facilitate early intervention to change the clinical trajectory and improve prognosis.

Aims: We aim to systematically review evidence for prediction models developed for predicting poor outcome in FEP.

Methods: A protocol for this study was published on the International Prospective Register of Systematic Reviews (PROSPERO), registration number CRD42019156897. Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance, we systematically searched six databases from inception to 28th January 2021. We used the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and the Prediction model Risk Of Bias Assessment Tool (PROBAST) to extract and appraise the outcome prediction models. We considered study characteristics, methodology and model performance.

Results: Thirteen studies reporting 31 prediction models across a range of clinical outcomes met criteria for inclusion. Eleven studies employed logistic regression with clinical and sociodemographic predictor variables. Just two studies were found to be at low risk of bias. Methodological limitations identified included a lack of appropriate validation, small sample sizes, poor handling of missing data and inadequate reporting of calibration and discrimination measures. To date, no model has been applied to clinical practice.

Conclusions: Future prediction studies in psychosis should prioritise methodological rigour and external validation in larger samples. The potential for prediction modelling in FEP is yet to be realised.

Introduction

Psychosis is a mental illness characterised by hallucinations, delusions and thought disorder. The median lifetime prevalence of psychosis is around eight per 1000 of the global population.[1] Psychotic disorders, including schizophrenia, are in the top 20 leading causes of disability worldwide.[2] People with psychosis have heterogeneous outcomes. More than 40% fail to achieve symptomatic remission.[3] At present, clinicians struggle to predict long term outcome in individuals with first episode psychosis (FEP).

Prediction modelling has the potential to revolutionise medicine by predicting individual patient outcome.[4] Early identification of those with good and poor outcomes would allow for a more personalised approach to care, matching interventions and resources to those most at need. This is the basis of precision medicine. Risk prediction models have been successfully employed clinically in many areas of medicine; for example, the QRISK tool predicts cardiovascular risk in individual patients.[5] However, within psychiatry, precision medicine is not yet established within clinical practice. In first episode psychosis, precision medicine could enable rapid stratification and targeted intervention thereby decreasing patient suffering and limiting treatment associated risks such as medication side effects and intrusive monitoring.

Salazar de Pablo *et al* recently undertook a broad systematic review of individualised prediction models in psychiatry. They found clear evidence that precision psychiatry has developed into an important area of research, with the greatest number of prediction models focussing on outcomes in psychosis. However, the field is hindered by methodological flaws, for example lack of validation. Further, there is a translation gap with only one study considering implementation into clinical practice. Systematic guidance for the development, validation and presentation of prediction models is available.[6] Further, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement sets standards for reporting.[7] Models that do not adhere to these guidelines result in unreliable predictions, which may cause more harm than good in

guiding clinical decisions.[8] Salazar de Pablo *et al* 's review was impressive in scope but necessarily limited in detailed analysis of the specific models included.[9] Systematic reviews focussing on the predicting the transition to psychosis,[10,11] and predicting relapse in psychosis have also been published.[12] In our present review, we focus on FEP with the aim to systematically review and critically appraise the prediction models for the prediction of poor outcomes.

Methods

We designed this systematic review in accordance with the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS).[13] A protocol for this study was published on the International Prospective Register of Systematic Reviews (PROSPERO), registration number CRD42019156897.

We developed the eligibility criteria under the Population, Index, Comparator, Outcome, Timing and Setting (PICOTS) guidance (**see supplementary materials**). A study was eligible for inclusion if it utilised a prospective design, including patients diagnosed with FEP, and developed, updated, or validated prognostic prediction models for any possible outcome, in any setting. We excluded non-English language studies, those where the full text was not available, those involving diagnostic prediction models, and those where the outcome predicted was less than or equal to three months from baseline because we were interested in longer term prediction.

We searched PubMed, PsychINFO, EMBASE, CINAHL Plus, Web of Science Core Collection and Google Scholar from inception up to 28th January 2021. In addition, we manually checked references cited in the systematically searched articles. The search terms were based around three themes – 'Prediction', 'Outcome' and 'First Episode Psychosis' terms. The full search strategy is available in the **supplementary materials**. Two reviewers (RL and LT) independently screened the titles and abstracts. Full text screening was completed by three independent reviewers (RL, PM and SPL). Disagreements were resolved by consensus.

Data extraction was conducted independently by two reviewers (RL and SPL) following recommendations in the CHARMS checklist.[13] From all eligible studies, we collected information on study characteristics, methodology and performance. Study characteristics collected included first author name, year, region, whether multicentre, study type, setting, participant description, outcome, outcome timing, predictor categories and number of models presented. Methodology considered sample size, events per variable (EPV), number of events in validation dataset, number of candidate and retained predictors, methods of variable selection, presence and handling of missing data, modelling strategies, shrinkage, validation strategies (see below), whether models were recalibrated, if clinical utility was assessed and whether the full models were presented. Steyerberg and Harrell outline a hierarchy of validation strategies from apparent (which assesses model performance on the data used to develop it and will be severely optimistic), to internal (via cross validation or bootstrapping), internal-external (for example, validation across centres in the same study) and external validation (to assess if models generalise to related populations in different settings).[14] Apparent, internal and internal-external validation use the derivation dataset only, while external validation requires the addition of a validation dataset. Performance for the best performing model per outcome in each article was considered by model validation strategy, including model discrimination (reported as the C-statistic which is equal to the area under the receiver operating characteristic (ROCAUC) curve for binary outcomes), calibration, other global performance measures, and classification metrics. If not reported, where possible, the balanced accuracy (sensitivity + specificity / 2) and the prognostic summary index (positive + negative predictive value - 1) were calculated.

Two reviewers (RL and SPL) independently assessed the risk of bias (ROB) in included studies using the Prediction model Risk Of Bias Assessment Tool (PROBAST), a risk of bias assessment tool designed for systematic reviews of diagnostic or prognostic prediction models.[15,16] We considered all models reported in each article and assigned to the article an overall rating. PROBAST uses a structured approach with signalling questions across four domains: ‘participants’, ‘predictors’,

'outcome' and 'statistical analysis'. Signalling questions are answered 'yes', 'probably yes', 'no', 'probably no' or 'no information'. Answering 'yes' indicates a low ROB, while 'no' indicates high ROB. A domain where all signalling questions are answered as 'yes' or 'probably yes' indicates low ROB. Answering 'no' or 'probably no' flags the potential for the presence of bias and reviewers should use their personal judgement to determine whether issues identified have introduced bias. Applicability of included studies to the review question is also considered in PROBAST.

We reported our results according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (**see supplementary materials**).[17]

Results

Systematic review of the literature yielded 2353 records from database searches and 67 from additional sources. After removal of duplicates, 1543 records were screened. Of these, 82 full texts were reviewed, which resulted in 13 studies meeting criteria for inclusion in our qualitative synthesis (**Figure 1**).[18,19,28–30,20–27]

Study characteristics are summarised in **Table 1**. The 13 included studies, comprising a total of 19 different patient cohorts, reported 31 different prediction models. Dates of publication ranged from 2006 and 2021. Twelve studies (92%) recruited participants from Europe, with two studies (15%) also recruiting participants from Israel and one study (8%) from Singapore. Over two-thirds (n=9) of studies were multicentre. Ten studies (77%) included participants from cohort studies, three studies (23%) included participants from randomised controlled trials and two studies (15%) included participants from case registries. Two studies (15%) included only out-patients, four (31%) included in-patients and out-patients and the rest did not specify their setting. Cohort sample size ranged from 47 to 1663 patients. The average age of patients ranged from 21 to 28 years, and 49% to 77% of the cohorts were male. Where specified, the average duration of untreated psychosis ranged from 34 to 106 weeks. Ethnicity was reported in 8 studies (62%) with the percentage non-white

patients in the cohorts ranging from 4% to greater than 75%. The definition of FEP was primarily non-affective psychosis in the majority of patient cohorts, with the minority also including affective psychosis and two cohorts also including drug-induced psychosis patients. All but one study (92%) considered solely sociodemographic and clinical predictors. A wide range of outcomes were assessed across the 13 included studies including symptom remission in five studies (38%), global functioning in five studies (38%), vocational functioning in three studies (23%), treatment resistance in two studies (15%), rehospitalisation in two studies (15%), and quality of life in one study (8%). All the outcomes were binary. The follow-up period of included studies ranged from 1 to 10 years.

Study prediction modelling methodologies are outlined in **Table 2**. Nine (69%) studies pertained solely to model development with the highest level of validation reported being apparent validity in four of the studies, internal validity in three of the studies and internal-external validity (via leave one-site out cross-validation) in two of the studies. The remaining four (31%) studies also included a validation cohort and reported external validity. High dimensionality was common across the study cohorts, with the majority having a very low events per variable (EPV) ratio and up to 258 candidate predictors considered. Some form of variable selection was employed in the majority (62%) of studies. The number of events in the external validation cohort ranged from 23 to 173. All the studies had missing data. Six studies (46%) used complete case analysis, five (38%) used single imputation and the remaining two (15%) applied multiple imputation.

The most common modelling methodology was logistic regression fitted by maximum likelihood estimation, then logistic regression with regularisation. Only two studies employed machine learning based methods, both via support vector machines. Just over half of studies (54%) did not use any variable shrinkage and only three studies (23%) recalibrated their models based on validation to improve performance. The full model was presented in seven (54%) studies. Only two studies (15%) assessed clinical utility.

The performance of the best model per study outcome grouped by method of validation to allow for appropriate comparisons is reported in **Table 3**. For the five studies (38%) reporting only apparent validity, two reported a measure of discrimination and only one considered calibration. For the seven studies (54%) reporting internal validation performance, four reported discrimination with a C-statistic ranging from 0.66 to 0.77 and four reported calibration. For the three studies (23%) reporting internal-external validation only one study considered discrimination with a C-statistic which ranged from 0.703 to 0.736 across each of its four models. None of the studies reporting internal-external validation considered any measure of calibration. All four studies (31%) reporting external validation considered model discrimination with C-statistics ranging from 0.556 to 0.876. However, only two of these studies considered calibration. **Table 3** also records any global performance metrics which included the Brier score and McFadden's pseudo-R^2, both of which incorporate aspects of discrimination and calibration. Various classification metrics were reported across the study models, but it is difficult to make any meaningful comparisons between these alone, without considering the models' corresponding discrimination and calibration metrics which were not universally reported.

We applied the PROBAST tool to the 31 different prediction models across the 13 studies in our systematic review and determined an overall risk of bias rating for each study as summarised in **Supplementary Table 1**. The majority (85%) of studies had an overall 'high' ROB. In each of these studies, the ROB was rated 'high' in the analysis domain with one study also having a 'high' ROB in the predictors domain. The main reasons for the 'high' ROB in the analysis domain were insufficient participant numbers and consequently low EPV, inappropriate methods of variable selection including via univariable analysis, a lack of appropriate validation with only apparent validation, an absence of reported measures of discrimination and calibration, and inappropriate handling of missing data by either complete case analysis or single imputation. Two studies, Leighton et al 2021[29] and Puntis et al 2021,[30] were rated overall 'low' ROB. These studies considered symptom remission and psychiatric rehospitalisation outcomes, respectively. Both studies externally validated

their prediction model and considered its clinical utility. However, neither study considered the implementation of the prediction model into actual clinical practice. When we assessed the 13 included studies according to PROBAST applicability concerns, all the studies were considered overall 'low' concern. This is indicative of the broad scope of our systematic review.

Discussion

Our systematic review identified 13 studies reporting 31 prognostic prediction models for the prediction of a wide range of clinical outcomes. The majority of models were developed via logistic regression. There were several methodological limitations identified including a lack of appropriate validation, issues with handling missing data and a lack of reporting of calibration and discrimination measures. We identified two studies with models at low risk of bias as assessed with PROBAST, both of which externally validated their models.

Principal Findings in Context

Our systematic review found no consistent definition of FEP across the different cohorts used for developing and validating prediction models. A lack of an operational definition for FEP within clinical and research settings has previously been identified as major a barrier to progress.[31] The majority of cohorts in our systematic review included only individuals with non-affective psychosis with a minority also including affective psychosis. In contrast, early intervention services typically do not make a distinction between affective and non-affective psychosis in those whom they accept into their service.[32] As such, there may be issues with generalisability of prediction models developed in cohorts with solely non-affective psychosis to real-world clinical practice.

A wide range of different outcomes were predicted by the FEP models including symptom remission, global functioning, vocational functioning, treatment resistance, rehospitalisation and quality of life outcomes. This is reflective of the fact that recovery from FEP is not readily distilled down to a single factor like symptom remission. Meaningful recovery is represented by a constellation of

multidimensional outcomes unique to each individual.[33] We should engage people with lived experience, to ensure that prediction models are welcomed and are predicting outcomes most relevant to the people they are for.

All the prediction models were developed in populations from high-income developed countries and only three studies included participants from countries outside of Europe, an issue not unique to FEP research. Consequently, it is currently unknown how prediction models for FEP would generalise to low-income developing countries. Prediction models may have considerable benefit in developing countries where almost 80% of patients with FEP live but where mental health support is often scarce.[34] Prediction models could help prioritise the appropriate utilisation of limited healthcare resources.

Only one study considered predictor variables other than clinical or sociodemographic factors. In this study, the additional predictors did not add significant value.[22] In recent years substantial progress has been made in elucidating the pathophysiological mechanisms underpinning the development of psychosis. We now recognise important roles for genetic factors, neurodevelopmental factors, dopamine and glutamate.[35] Prediction model performance may be improved by the incorporation of these biological relevant disease markers as predictor variables. However, the cost-benefit of adding more expensive and less accessible disease markers must be carefully considered, especially if models are to be utilised in settings where resources are more limited.

Machine learning can be operationally defined as “models that directly and automatically learn from data”. This is to be contrasted with regression models which “are based on theory and assumptions, and benefit from human intervention and subject knowledge for model specification.”[36] Just two studies employed machine learning techniques for their modelling.[22,26] The rest of the studies employed logistic regression. We were unable to make any comparison between the discrimination and calibration ability of the two studies employing machine learning and the other studies because these metrics were not provided. However, a recent systematic review found no evidence of

superior performance of clinical prediction models using machine learning methods over logistic regression.[36] In any case, the distinction between regression models and machine learning has been viewed to be artificial. Instead, algorithms may exist "along a continuum between fully human-guided to fully machine-guided data analysis".[37] An alternative comparison may be between linear and non-linear classifiers. Only one study employed a non-linear classifier,[26] but again we were unable to gain meaningful insights into its relative performance because appropriate metrics were not provided.

A principal finding from our systematic review is the presence of methodological limitations across the majority of studies. Steyerberg *et al* outline four key measures of predictive performance that should be assessed in any prediction modelling study – two measures of calibration (the model intercept (A) and the calibration slope (B)), discrimination via a concordance statistic (C), and clinical usefulness with decision-curve analysis (D).[6] Model calibration is the level of agreement between the observed outcomes and the predictions. For example, if a model predicts a 5% risk of cancer, then, according to such a prediction, the observed proportion should be five cancers per 100 people. Discrimination is the ability of a model to distinguish between a patient with the outcome and one without.[6] Our review found that only seven studies (54%) reported discrimination and just five (38%) reported any measure of calibration. The remaining studies reported only classification metrics, such as accuracy or balanced accuracy. The problem with solely reporting classification metrics is that they vary both across models and across different probability thresholds for the same model. This renders the comparison between models less meaningful. It is further argued that setting a classification threshold for a probability generating model is premature. Rather, a clinician may choose to set different probability thresholds for the same prediction model depending on the situation at hand in order to optimise the balance between false positives and false negatives. For example, in the case of a model predicting cancer, a clinician may choose a lower probability threshold to offer a non-invasive screening test and a higher probability threshold to suggest an invasive and potentially harmful biopsy. Further, without any measure of model calibration we are

unable to assess if the model can make unbiased estimates of outcome.[38] The final key step in assessing the performance of a prediction model is to determine its clinical usefulness – that is, can better decisions be made with the model than without? Decision-curve analysis considers the net-benefit (the treatment threshold weighted sum of true- minus false-positive classifications) for a prediction model in comparison the default strategy of treating all or no patients, across an entire range of treatment thresholds.[39] Only two studies (15%) included in our review considered whether the model was clinically useful. Without proper validation of prediction models, the reported performances are likely to be overly optimistic. Four studies (31%) report only apparent validity. Just four studies (31%) reported external validation, considered essential before applying a prediction model to clinical practice.[14]

Altogether, just two studies (15%) had an overall 'low' risk of bias according to PROBAST, reflecting these methodological limitations. Neither study considered real-world implementation. To progress with implementation, impact studies are required. These would involve a cluster randomised trial comparing patient outcomes between a group with treatment informed by a clinical prediction model and a control group.[40] We are not aware of any such study having been carried out within the field of psychiatry. However, Salazar de Pablo *et al* suggest that PROBAST thresholds for considering a study to be a 'low' risk of bias may be too strict.[9] Indeed, in the field of machine learning multiple imputation is frequently computationally infeasible and single imputation may be viewed as sufficient. This is especially true in larger datasets or in the presence of relatively few missing values.[41]

Strengths and limitations

Our review had a number of strengths. We provide the first systematic overview of prediction modelling studies for use in patients with first episode psychosis. We offer a detailed critique of the study characteristics, their methodologies and model performance metrics. Further, our review

adheres to gold standard guidance for extracting data from prediction models and for assessing bias, namely the CHARMS checklist and PROBAST.

There were several limitations. Our initial aim was to perform a meta-analysis of any prediction model which was validated across different settings and populations. However, no meta-analysis was possible because no single prediction model was validated more than once. In addition, as a consequence of poor reporting of discrimination and calibration performance across the studies, it was often difficult to make meaningful comparison between the prediction models. Also, the lack of consensus as to the most important outcome measure in FEP, with six different outcomes considered across only 13 included studies, further hindered efforts at drawing meaningful comparisons between the included studies and their respective prediction models. Likewise, if more studies had considered the same outcome measures, this may have afforded the opportunity to validate existing prediction models rather than necessitating the creation of additional new models. All published prediction modelling studies in FEP reported significant positive findings. It is possible that studies which had negative findings were held back from publication reflecting the possibility of publication bias. We originally intended to evaluate the overall certainty in the body of evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.[42] GRADE was originally designed for reviews of intervention studies but has not yet been adapted for use in systematic reviews of prediction models. Consequently, in its current form we did not find GRADE to be a suitable tool for our review and decided not to use it. Future research should consider how to adapt GRADE for use in systematic reviews of prediction models.

Implications for future research

It is clear that there is a growing trend for the development of prediction models in FEP.[9] FEP is an illness which responds best to an early intervention paradigm.[43] Prediction models have the potential to optimise the allocation of time-critical interventions, like clozapine for treatment resistance.[44] However, prior to meaningful implementation into real-world clinical practice several

steps are necessary. The field must prioritise external validation and replication of existing prediction models in larger sample sizes to increase the EPV. This is best accomplished by an emphasis on data-sharing and open collaboration. Prediction studies should include FEP cohorts from low-income countries where there is considerable potential for benefit by helping to prioritise limited resources to those most in need. Harmonisation of data collection across the field both in terms of predictors and outcomes measured would facilitate validation efforts. There should be a greater consideration of biologically relevant and cognitive predictors based on our growing understanding of disease mechanisms, which could optimise prediction model performance. Finally, our review highlights considerable methodological pitfalls in much of the current literature. Future prediction modelling studies should focus on methodological rigour with adherence to accepted best practice guidance.[6,14,38] Our goal in psychiatry should be to develop an innovative approach to care using prediction models. Application of these approaches into clinical practice would enable rapid and targeted intervention thereby limiting treatment associated risks and reducing patient suffering.

Declaration of Interest

GVG has received support from H2020-EINFRA, the NIHR Birmingham ECMC, NIHR Birmingham SRMRC, the NIHR Birmingham Biomedical Research Centre, and the MRC HDR UK, an initiative funded by UK Research and Innovation, Department of Health and Social Care (England), the devolved administrations, and leading medical research charities. JC has received grants from Wellcome Trust and Sackler Trust and honorariums from Johnson & Johnson. PKM has received honorariums from Sunovion and Sage and is a Director of Noux Technologies Limited. All other authors declare no competing interests.

Funding

RL is funded by an Institute for Mental Health Priestley Scholarship, University of Birmingham. SPL is funded by a Clinical Academic Fellowship from the Chief Scientist Office, Scotland (CAF/19/04). SJW is funded by the Medical Research Council, UK (MR/K013599).

Author Contribution

PKM and RL formulated the research question and designed the study. RL, SPL, LT and PKM collected the data. RL, SPL and PKM analysed the data and drafted the manuscript. LT, GVG, SJW,S-JHF, FD and JC critically evaluated and revised the manuscript.

Data Availability

Data is available on request from the corresponding author.

Author Details

Rebecca Lee*[1], Samuel P Leighton*†[2], Lucretia Thomas[3], Georgios V Gkoutos[4], Stephen J Wood[5-7], Sarah-Jane H Fenton[1], Fani Deligianni[8], Jonathan Cavanagh[9], Pavan K Mallikarjun[1]

*Joint first authors

†Corresponding author – samuel.leighton@glasgow.ac.uk

1. Institute for Mental Health, University of Birmingham, UK.
2. Institute of Health and Wellbeing, University of Glasgow, UK.
3. Birmingham Medical School, University of Birmingham, UK.
4. Institute of Cancer and Genomic Sciences, University of Birmingham, UK.
5. Orygen Youth Health Research Centre, National Centre of Excellence in Youth Mental Health, Parkville, Victoria, Australia
6. School of Psychological Sciences, University of Melbourne, Australia
7. School of Psychology, University of Birmingham, UK.
8. School of Computing Science, University of Glasgow, UK.
9. Institute of Infection, Immunity and Inflammation, University of Glasgow, UK.

References

1. Moreno-Küstner B, Martín C, Pastor L. Prevalence of psychotic disorders and its association with methodological issues. A systematic review and meta-analyses [Internet]. Vol. 13, PLoS ONE. Public Library of Science; 2018 [cited 2021 May 10]. Available from: /pmc/articles/PMC5896987/

2. IHME. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. [Internet]. Seattle, WA.: IHME, University of Washington; 2021. Available from: http://vizhub.healthdata.org/gbd-compare.

3. Lally J, Ajnakina O, Stubbs B, Cullinane M, Murphy KC, Gaughran F, et al. Remission and recovery from first-episode psychosis in adults: Systematic review and meta-analysis of long-term outcome studies [Internet]. Vol. 211, British Journal of Psychiatry. The Royal College of Psychiatrists; 2017 [cited 2018 Jan 18]. p. 350–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28982659

4. Darcy AM, Louie AK, Roberts LW. Machine Learning and the Profession of Medicine. JAMA [Internet]. 2016 Feb 9 [cited 2018 Jan 18];315(6):551. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.18421

5. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ [Internet]. 2017 May 23 [cited 2018 Jun 29];357:j2099. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28536104

6. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J [Internet]. 2014 Aug 1 [cited 2018 Jan 18];35(29):1925–31. Available from: https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehu207

7. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015 Jan;350:g7594.

8. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. BMJ [Internet]. 2020 Apr 7 [cited 2021 Jun 16];369:26. Available from: https://www.bmj.com/content/369/bmj.m1328

9. Salazar de Pablo G, Studerus E, Vaquerizo-Serrano J, Irving J, Catalan A, Oliver D, et al. Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice. Schizophr Bull [Internet]. 2021 Mar 1 [cited 2021 May 10];47(2):284–97. Available from: https://academic.oup.com/schizophreniabulletin/article/47/2/284/5903901

10. 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 [Internet]. Vol. 47, Psychological Medicine. Cambridge University Press; 2017 [cited 2021 Oct 20]. p. 1163–78. Available from: https://www.cambridge.org/core/journals/psychological-medicine/article/abs/prediction-of-transition-to-psychosis-in-patients-with-a-clinical-high-risk-for-psychosis-a-systematic-review-of-methodology-and-reporting/1CEA9147A2ED19BE6162ECABD57B129F

11. Rosen M, Betz LT, Schultze-Lutter F, Chisholm K, Haidl TK, Kambeitz-Ilankovic L, et al. Towards clinical application of prediction models for transition to psychosis: A systematic review and external validation study in the PRONIA sample [Internet]. Vol. 125, Neuroscience and Biobehavioral Reviews. Neurosci Biobehav Rev; 2021 [cited 2021 Oct 20]. p. 478–92. Available from: https://pubmed.ncbi.nlm.nih.gov/33636198/

12. Sullivan S, Northstone K, Gadd C, Walker J, Margelyte R, Richards A, et al. Models to predict

relapse in psychosis: A systematic review. PLoS One [Internet]. 2017 Sep 1 [cited 2021 Oct 20];12(9):e0183998. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0183998

13. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med [Internet]. 2014 [cited 2021 May 6];11(10):e1001744. Available from: www.plosmedicine.org

14. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol [Internet]. 2016 Jan [cited 2019 Mar 19];69:245–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25981519

15. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019 Jan 1;170(1):51–8.

16. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1–33.

17. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews [Internet]. Vol. 372, The BMJ. BMJ Publishing Group; 2021 [cited 2021 May 6]. Available from: http://dx.doi.org/10.1136/bmj.n160

18. Ajnakina O, Agbedjro D, Lally J, Forti M Di, Trotta A, Mondelli V, et al. Predicting onset of early- and late-treatment resistance in first-episode schizophrenia patients using advanced shrinkage statistical methods in a small sample. Psychiatry Res. 2020 Dec 1;294:113527.

19. Bhattacharyya S, Schoeler T, Patel R, di Forti M, Murray RM, McGuire P. Individualized

421 prediction of 2-year risk of relapse as indexed by psychiatric hospitalization following
422 psychosis onset: Model development in two first episode samples. Schizophr Res. 2021 Feb
423 1;228:483–92.

424 20. Chua YC, Abdin E, Tang C, Subramaniam M, Verma S. First-episode psychosis and vocational
425 outcomes: A predictive model. Schizophr Res. 2019 Sep 1;211:63–8.

426 21. Demjaha A, Lappin JM, Stahl D, Patel MX, MacCabe JH, Howes OD, et al. Antipsychotic
427 treatment resistance in first-episode psychosis: Prevalence, subtypes and predictors. Psychol
428 Med [Internet]. 2017 Aug 1 [cited 2021 Mar 25];47(11):1981–9. Available from:
429 https://doi.org/10.1017/S0033291717000435

430 22. de Nijs J. The outcome of psychosis [Internet]. Utrecht University; 2019. Available from:
431 https://dspace.library.uu.nl/bitstream/1874/376436/1/22_01_3_jessica_de_nijs_compleet_fi
432 nal.pdf

433 23. Derks EM, Fleischhacker WW, Boter H, Peuskens J, Kahn RS. Antipsychotic drug treatment in
434 first-episode psychosis should patients be switched to a different antipsychotic drug after 2,
435 4, or 6 weeks of nonresponse? J Clin Psychopharmacol. 2010 Apr;30(2):176–80.

436 24. Flyckt L, Mattsson M, Edman G, Carlsson R, Cullberg J. Predicting 5-Year Outcome in First-
437 Episode Psychosis: Construction of a Prognostic Rating Scale. J Clin Psychiatry [Internet]. 2006
438 Jun [cited 2021 Mar 25];67(6):916–24. Available from:
439 https://dx.doi.org/10.4088/jcp.v67n0608

440 25. González-Blanch C, Perez-Iglesias R, Pardo-García G, Rodríguez-Snchez JM, Martínez-García O,
441 Vázquez-Barquero JL, et al. Prognostic value of cognitive functioning for global functional
442 recovery in first-episode schizophrenia. Psychol Med [Internet]. 2010 Jun [cited 2021 Mar
443 25];40(6):935–44. Available from: https://doi.org/10.1017/S0033291709991267

444 26. Koutsouleris N, Kahn RS, Chekroud AM, Leucht S, Falkai P, Wobrock T, et al. Multisite

prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. The Lancet Psychiatry [Internet]. 2016 Oct 1 [cited 2021 Mar 25];3(10):935–46. Available from: http://dx.doi.org/10.1016/S2215-0366(16)30171-7

27. Leighton SP, Krishnadas R, Chung K, Blair A, Brown S, Clark S, et al. Predicting one-year outcome in first episode psychosis using machine learning. Acampora G, editor. PLoS One [Internet]. 2019 Mar 7 [cited 2021 Mar 25];14(3):e0212846. Available from: https://dx.plos.org/10.1371/journal.pone.0212846

28. Leighton SP, Upthegrove R, Krishnadas R, Benros ME, Broome MR, Gkoutos G V., et al. Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: a machine learning approach. Lancet Digit Heal [Internet]. 2019 Oct 1 [cited 2021 Jan 29];1(6):e261–70. Available from: http://dx.doi.org/10.1016/

29. Leighton SP, Krishnadas R, Upthegrove R, Marwaha S, Steyerberg EW, Broome MR, et al. Development and validation of a non-remission risk prediction model in First Episode Psychosis: An analysis of two longitudinal studies. Schizophr Bull Open. 2021;in press.

30. Puntis S, Whiting D, Pappa S, Lennox B. Development and external validation of an admission risk prediction model after treatment from early intervention in psychosis services. Transl Psychiatry [Internet]. 2021 Jun 1 [cited 2021 Mar 29];11(1):35. Available from: https://doi.org/10.1038/s41398-020-01172-y

31. Breitborde NJK, Srihari VH, Woods SW. Review of the operational definition for first-episode psychosis [Internet]. Vol. 3, Early Intervention in Psychiatry. Early Interv Psychiatry; 2009 [cited 2021 Jun 16]. p. 259–65. Available from: https://pubmed.ncbi.nlm.nih.gov/22642728/

32. National Institute for Health and Care Excellence (NICE). Implementing the Early Intervention

in Psychosis Access and Waiting Time Standard: Guidance. 2016.

33. Jääskeläinen E, Juola P, Hirvonen N, McGrath JJ, Saha S, Isohanni M, et al. A systematic review and meta-analysis of recovery in schizophrenia. Schizophr Bull. 2013 Nov;39(6):1296–306.

34. Singh SP, Javed A. Early intervention in psychosis in low- and middle-income countries: a WPA initiative [Internet]. Vol. 19, World Psychiatry. Blackwell Publishing Ltd; 2020 [cited 2021 May 13]. p. 122. Available from: /pmc/articles/PMC6953594/

35. Lieberman JA, First MB. Psychotic Disorders. Ropper AH, editor. N Engl J Med [Internet]. 2018 Jul 19 [cited 2021 May 13];379(3):270–80. Available from: http://www.nejm.org/doi/10.1056/NEJMra1801490

36. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Vol. 110, Journal of Clinical Epidemiology. Elsevier USA; 2019. p. 12–22.

37. Beam AL, Kohane IS. Big data and machine learning in health care [Internet]. Vol. 319, JAMA - Journal of the American Medical Association. American Medical Association; 2018 [cited 2021 Oct 20]. p. 1317–8. Available from: https://jamanetwork.com/journals/jama/fullarticle/2675024

38. Harrell , FE. Regression Modeling Strategies [Internet]. Cham: Springer International Publishing; 2015 [cited 2021 May 14]. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-3-319-19425-7

39. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic Progn Res [Internet]. 2019 Dec 4 [cited 2021 May 14];3(1):1–8. Available from: https://doi.org/10.1186/s41512-019-0064-7

40. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment [Internet].

493 Vol. 98, Heart. BMJ Publishing Group Ltd and British Cardiovascular Society; 2012 [cited 2021
494 May 14]. p. 691–8. Available from: http://heart.bmj.com/

495 41. Steyerberg EW. Clinical Prediction Models [Internet]. Cham: Springer International Publishing;
496 2019 [cited 2021 Oct 20]. (Statistics for Biology and Health). Available from:
497 http://link.springer.com/10.1007/978-3-030-16399-0

498 42. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of
499 evidence and strength of recommendations for diagnostic tests and strategies [Internet]. Vol.
500 9, Chinese Journal of Evidence-Based Medicine. BMJ; 2009 [cited 2021 Aug 27]. p. 503–8.
501 Available from: https://pubmed.ncbi.nlm.nih.gov/18483053/

502 43. Birchwood M, Todd P, Jackson C. Early intervention in psychosis: The critical period
503 hypothesis. Br J Psychiatry [Internet]. 1998;172(S33):53–9. Available from:
504 https://doi.org/10.1192/S0007125000297663

505 44. Farooq S, Choudry A, Cohen D, Naeem F, Ayub M. Barriers to using clozapine in treatment-
506 resistant schizophrenia: systematic review. BJPsych Bull [Internet]. 2019 Feb [cited 2021 Jan
507 29];43(1):8–16. Available from: https://www.cambridge.org/core.

508

Figure Legends

Figure 1 – PRISMA flow diagram.

Table 1 – Study characteristics

Study ID	Country	Multi-centre	Recruitment Dates	Type of Study	Setting	Participants included in modelling					Outcome		Predictor Categories	No. of Models
						Sex (% male)	Age (mean)	Ethnicity	DUP (mean weeks)	FEP Definition	Definition	Timing		
AJNAKINA 2020	UK	No	Dec 2005 to Oct 2010	Cohort	In-patients & out-patients	67.5%	27.2 (at baseline)	39.9% white, 60.1% black	34.3	Non-affective	Early treatment resistance from illness onset Later treatment resistance	f/u for 5 years	Socio-demographic, Clinical	4
BHATTACHARYYA 2021	UK	No	Sample 1 - 1st Apr 2006 to 31st Mar 2012 Sample 2 - 12th Apr 2002 to 26th Jul 2013	Sample 1 - Case Registry Sample 2 - Cohort	Sample 1 - out-patients Sample 2 - out-patients	Sample 1 - 63.9% Sample 2 - 60%	Sample 1 - 24.4 (at onset) Sample 2 - 28.1 (at onset)	Sample 1 - 31.1% white, 50.6% black Sample 2 - 34.2% white, 54.2% black	N.R.	Sample 1 - Non-affective & affective Sample 2 - Non-affective & affective	Psychiatric rehospitalisation	f/u for 2 years	Socio-demographic, Clinical	3
CHUA 2019	Singapore	No	2001 to 2012	Cohort	N.R.	49.2%	27.5 (at baseline)	76.7% Chinese	65.4	Non-affective	EET status	At 2 years	Socio-demographic, Clinical	2
DEMJAHA 2017	UK	Yes	Sep 1997 to Aug 1999	Cohort	N.R.	58.4%	28.9 (at onset)	48.2% white, 39.8% black	N.R.	Non-affective & affective	Early treatment resistance from illness onset	f/u for 10 years	Socio-demographic, Clinical	1
DENIJS 2019	Netherlands & Belgium	Yes	8th Jan 2004 to 6th Feb 2008	Cohort	In-patients & out-patients	76.9%	27.6 (at baseline)	85.9% white	N.R.	Non-affective	Andreasen symptom remission (6 months duration) GAF ≥65	At 3 years & at 6 years	Socio-demographic, Clinical, Genetic, Environmental	8
DERKS 2010	Austria, Belgium, Bulgaria, Czech Republic, Germany, France, Israel, Italy, Netherlands, Poland, Rumania, Spain, Sweden & Switzerland	Yes	23rd Dec 2002 to 14th Jan 2006	Randomised Controlled Trial	N.R.	56.5%	26.0 (at baseline)	N.R.	N.R.	Non-affective	Andreasen symptom remission (6 months duration)	f/u for 1 year	Socio-demographic, Clinical	1

FLYCKT 2006	Sweden	Yes	1st Jan 1996 to 31st Dec 1997	Cohort	N.R.	52.9%	28.8 (at baseline)	N.R.	62.4	Non-affective & affective (with mood-incongruent delusions)	Global functioning (independent living, EET status & GAF ≥60)	At mean of 5.4 years	Socio-demographic, Clinical	1
GONZALEZ-BLANCH 2010	Spain	No	Feb 2001 to Feb 2005	Cohort	N.R.	62%	26.6 (at baseline)	N.R.	66.6	Non-affective	Global functioning (EET status & DAS ≤1)	At 1 year	Socio-demographic, Clinical	1
KOUTSOULERIS 2016	Austria, Belgium, Bulgaria, Czech Republic, Germany, France, Israel, Italy, Netherlands, Poland, Rumania, Spain, Sweden & Switzerland	Yes	23rd Dec 2002 to 14th Jan 2006	Randomised Controlled Trial	N.R.	56%	26.1 (at baseline)	N.R.	N.R.	Non-affective	GAF ≥65	At 1 year	Socio-demographic, Clinical	1
LEIGHTON 2019 (1)	UK	Yes	Dev. - 2011 to 2014 Val. - 1st Sep 2006 to 31st Aug 2009	Dev. – Cohort Val. - Cohort	Dev. - In-patients & out-patients Val. - In-patients & out-patients	Dev. - 66% Val. - 68%	Dev. - 25.2 (at baseline) Val. - 24.6 (at baseline)	Dev. - 81% white Val. - 96% white	N.R.	Dev. - Non-affective & affective Val. - Non-affective & affective	EET Status Andreasen symptom remission (no duration criteria) Andreasen symptom remission (6 months duration)	At 1 year	Socio-demographic, Clinical	3
LEIGHTON 2019 (2)	UK & Denmark	Yes	Dev. - Aug 2005 to Apr 2009 Val. UK - 1st Sep 2006 to 31st Aug 2009 & 2011 to 2014 Val Denmark - Jan 1998 to Dec 2000	Dev. - Cohort Val. UK - 2 Cohort studies Val. Denmark - Randomised Controlled Trial	Dev. - N.R. Val. UK - In-patients & out-patients Val. Denmark - In-patients & out-patients	Dev. - 69% Val. UK - 67% Val. Denmark - 59%	Dev. - 21.3 (at baseline) Val. UK - 24.9 (at baseline) Val. Denmark - 26.6 (at baseline)	Dev. - 73% white Val. UK - 88% white Val. Denmark - 94% white	Dev. - 44 Val. UK - 44.4 Val. Denmark - 106	Dev. - Non-affective, affective & drug induced Val. UK - Non-affective & affective Val. Denmark - Non-affective	EET Status GAF ≥65 Andreasen Symptom Remission (6 months duration) Quality of Life	At 1 year	Socio-demographic, Clinical	4

LEIGHTON 2021	UK	Yes	Dev. - Aug 2005 to Apr 2009 Val. - Apr 2006 to Feb 2009	Dev – Cohort Val - Cohort	N.R.	Dev. - 68.8% Val. - 61.8%	Dev - 22.6 (at baseline) Val. - 25.0 (at baseline)	N.R.	Dev. - 41.3 Val. - 48.9	Dev. - Non-affective, affective & drug induced Val. - Non-affective, affective & drug induced	Andreasen Symptom Remission (6 months duration)	At 1 year	Socio-demographic, Clinical	1
PUNTIS 2021	UK	Yes	Dev. - 1st Jan 2011 to 8th Oct 2019 Val. - 31st Jan 2006 to 18th Jun 2019	Dev. - Case Registry Val. - Case Registry	Dev. - out-patients Val. - out-patients	Dev. - 63% Val. - 63%	Dev. - 25.6 (at baseline) Val. - 26.7 (at baseline)	Dev. - 74.8% white Val. - 35.4% white	N.R.	N.R.	Psychiatric hospitalisation after discharge from early intervention	f/u for 1 year	Socio-demographic, Clinical	1

FEP – first episode psychosis; N.R. – not reported; DUP – duration of untreated psychosis; Dev. – development sample; Val. – validation sample; EET – employment, education or training; f/u – follow-up; GAF – Global Assessment of Functioning; DAS – Disability Assessment Schedule

Table 2 – Study Methodology

Study ID	Sample Size	EPV	No. Events in Validation Dataset	No. Candidate Predictors	No. Retained Predictors	Variable Selection	Missing Data Per Predictor	Handling of Missing Data	Modelling Method	Shrinkage	Validation Method Reported	Re-calibration Performed	Full Model Presented	Clinical Usefulness Assessed
AJNAKINA 2020	Recruited – 283; Included in modelling - 190 to 222	2 to 4	No external validation	13	12 to 13	Full model approach or LASSO	up to 59.9%	Single imputation	Logistic regression via ridge & LASSO	Penalised estimation & then uniform	Internal	Yes	Yes	No
BHATTACHARYYA 2021	Sample 1 - Recruited - 1738; Included in modelling - 1663 Sample 2 - Recruited - 240; Included in modelling - 240	4 to 62	No external validation	10 to 21	10 to 21	Full model approach	Sample 1 - up to 4.3% Sample 2 - none	Complete case analysis	Logistic regression via MLE	None	Apparent & internal	No	Yes	No
CHUA 2019	Recruited - 1724; Included in modelling - 1177	16	No external validation	22	22	Full model approach	Yes but N.R.	Complete case analysis	Logistic regression via MLE	None	Apparent	No	No	No
DEMJAHA 2017	Recruited - 557; Included in modelling - 286	8	No external validation	8	6	LASSO	Yes but N.R.	Complete case analysis	Logistic regression via LASSO	Penalised estimation	Internal	No	Yes	No
DENIJS 2019	Recruited - 1100; Included in modelling - 442 to 523	2	No external validation	258	119 to 152	Recursive feature elimination	up to 20%	Single imputation	Linear Support Vector Machine	None	Internal & internal-external	No	No	No
DERKS 2010	Recruited - 498; Included in modelling - 297	9 to 18	No external validation	10 to 20	10 to 20	Full model approach	Yes but N.R.	Complete case analysis	Logistic regression via MLE	None	Apparent	No	No	No
FLYCKT2006	Recruited 175; Included in modelling - 111	2	No external validation	32	5	Forward selection	Yes but N.R.	Complete case analysis	Logistic regression via MLE	None	Apparent	No	Yes	No
GONZALEZ-BLANCH 2010	Recruited - 174; Included in modelling – 92	4	No external validation	23	2	Univariate significance testing (p<0.1) then forward selection	Yes but N.R.	Complete case analysis	Logistic regression via MLE	None	Apparent	No	Yes	No
KOUTSOULERIS 2016	Recruited - 498; Included in modelling - 334	<1	No external validation	189	N.R.	Forward selection	up to 20%	Single imputation	Nonlinear Support Vector Machine	None	Internal & internal-external	No	No	No

LEIGHTON 2019 (1)	Dev. - Recruited - 83; Included in modelling - 67 to 75 Val. - Recruited - 79; Included - 64 to 67	<1	27 to 46	56	5 to 13	Elastic net	Dev. - up to 13% Val. - up to 37%	Single imputation	Logistic regression via elastic net	Penalised estimation	External	No	No	No
LEIGHTON 2019 (2)	Dev. - Recruited - 1027; Included in modelling - 673 to 829 Val. UK - Recruited - 162; Included - 47 to 142 Val. Denmark - Recruited - 578; Included - 226 to 553	1 to 2	23 to 173	163	17 to 26	Elastic net	Dev. - up to 20% Val. - Yes but N.R.	Single imputation	Internal Validation - Logistic regression via elastic net External Validation - Logistic regression via MLE	Internal-external validation - penalised estimation External validation - none	Internal-external & external	No	No	No
LEIGHTON 2021	Dev. - Recruited - 1027; Included in modelling – 673 Val. - Recruited - 399; Included - 191	25	103	14	14	Full model approach	Dev.- up to 14.9% Val. - up to 56.5%	Multiple imputation	Logistic regression via MLE	Uniform	Internal & external	Yes	Yes	Yes
PUNTIS 2021	Dev. - Recruited - N.R.; Included in modelling - 831 Val. - Recruited - N.R.; Included - 1393	10	162	8	8	Full model approach	Dev. - up to 15.4% Val. - up to 5.5%	Multiple imputation	Logistic regression via MLE	Uniform	Internal & external	Yes	Yes	Yes

N.R. – not reported; Dev. – development sample; Val. – validation sample; EPV – events per variable; LASSO – least absolute shrinkage and selection operator; MLE – maximum likelihood estimation

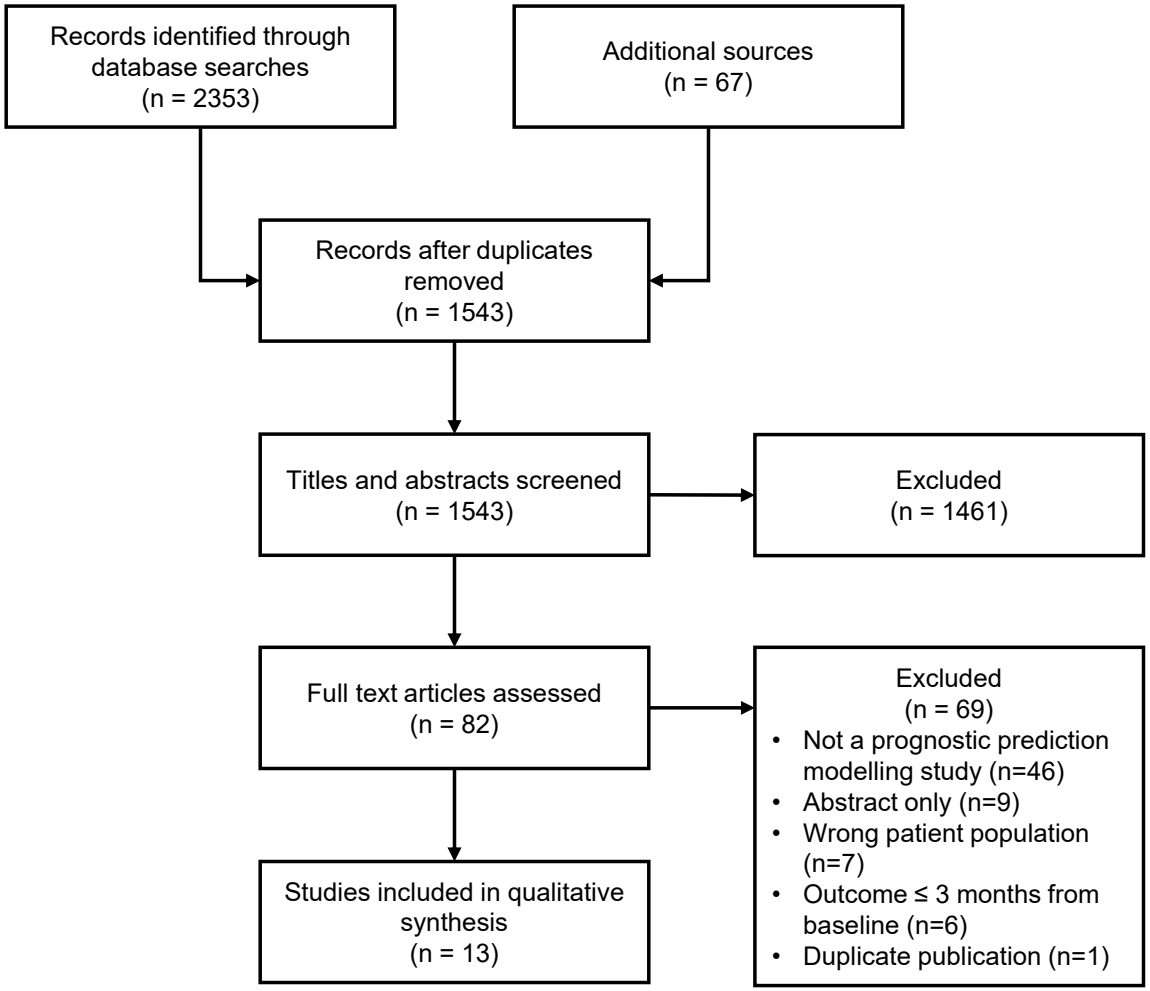
Table 3 – Performance metrics for best model per outcome in each study

Study ID	Outcome	Discrimination C-Statistic	Calibration	Other Global Performance Metrics	Classification Metrics
Studies Reporting Apparent Validity					
BHATTACHARYYA 2021	Psychiatric rehospitalisation	0.749	Calibration plot only; No α or β	Brier score - 0.192	N.R.
CHUA 2019	EET Status at 2 years	0.759 (95%CI: 0.728, 0.790)	N.R.	N.R.	Classification Accuracy - 0.759; PPV - 0.64; NPV - 0.78; PSI - 0.42
DERKS 2010	Andreasen symptom remission (6 months duration) with 1 year f/u	N.R.	N.R.	N.R.	Classification Accuracy - 0.63; Balanced Accuracy - 0.665; Sensitivity - 0.73; Specificity - 0.60; PPV - 0.73; NPV - 0.61; PSI - 0.34
FLYCKT 2006	Global functioning (Independent living, EET status, GAF ≥60) at mean 5.4 years	N.R.	N.R.	N.R.	Classification Accuracy - 0.81; Balanced Accuracy - 0.805; Sensitivity - 0.84; Specificity - 0.77
GONZALEZ-BLANCH 2010	Global functioning (EET status, DAS ≤1) at 1 year	N.R.	Hosmer–Lemeshow test - p = >0.05	N.R.	Classification Accuracy - 0.750; Balanced Accuracy - 0.587; Sensitivity - 0.261; Specificity - 0.913; PPV - 0.500; NPV - 0.788; PSI - 0.288
Studies Reporting Internal Validity					
	Early treatment resistance from illness onset with 5 years f/u	0.77	α - 0.028; β - 1.264; No calibration plot	N.R.	Balanced Accuracy - 0.5; Sensitivity - 0; Specificity - 1.00; PPV - 0.48, NPV - 0.84; PSI - 0.32
AJNAKINA 2020	Later treatment resistance with 5 years f/u	0.77	α - 0.504; β - 1.838; No calibration plot	N.R.	Balanced Accuracy - 0.81; Sensitivity - 0.62; Specificity - 1.00; PPV - 0.42; NPV - 1.00; PSI - 0.42
BHATTACHARYYA 2021	Psychiatric rehospitalisation	0.66	Calibration plot only; No α or β	Brier score - 0.232	N.R.
DEMJAHA 2017	Early Treatment Resistance from Illness Onset with 10 years f/u	N.R.	N.R.	Brier score - 0.146; McFadden pseudo R^2 - 0.1	N.R.
	Andreasen Symptom Remission (6 months duration) at 3 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.644; Sensitivity - 0.76; Specificity - 0.50; PPV - 0.722; NPV - 0.548; PSI - 0.27
	GAF ≥65 at 3 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.676; Sensitivity - 0.749; Specificity - 0.584; PPV - 0.701; NPV - 0.642; PSI - 0.343
	Andreasen symptom remission (6 months duration) at 6 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.647; Sensitivity - 0.787; Specificity - 0.465; PPV - 0.690; NPV - 0.590; PSI - 0.28
DENIJS 2019	GAF ≥65 at 6 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.676; Sensitivity - 0.818; Specificity - 0.477; PPV - 0.718; NPV - 0.616; PSI - 0.334
KOUTSOULERIS 2016	GAF ≥65 at 1 year	N.R.	N.R.	N.R.	Balanced Accuracy - 0.738; Sensitivity - 0.667; Specificity - 0.809; PPV - 0.515; NPV - 0.888; PSI - 0.403
LEIGHTON 2021	Andreasen symptom remission (6 months duration) at 1 year	0.74 (0.73, 0.75)	β - 0.84 (95%CI: 0.81, 0.86); No calibration plot	N.R.	N.R.
PUNTIS 2021	Psychiatric hospitalisation after discharge from early intervention	0.76 (0.75, 0.77)	α - 0.01 (95%CI: -0.25, 0.24); β - 0.89 (95%CI: 0.88, 0.89); Calibration plot	Brier score - 0.078	N.R.
Studies Reporting Internal-External Validity					

	Andreasen symptom remission (6 months duration) at 3 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.638; Sensitivity - 0.629; Specificity - 0.647; PPV - 0.758; NPV - 0.485; PSI - 0.243
	GAF ≥65 at 3 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.648; Sensitivity - 0.658; Specificity - 0.638; PPV - 0.727; NPV - 0.565; PSI - 0.292
	Andreasen symptom remission (6 months duration) at 6 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.625; Sensitivity - 0.685; Specificity - 0.565; PPV - 0.743; NPV - 0.493; PSI - 0.236
DENIJS 2019	GAF ≥65 at 6 years	N.R.	N.R.	N.R.	Balanced Accuracy - 0.640; Sensitivity - 0.718; Specificity - 0.561; PPV - 0.732; NPV - 0.553; PSI - 0.285
KOUTSOULERIS 2016	GAF ≥65 at 1 year	N.R.	N.R.	N.R.	Balanced Accuracy - 0.711; Sensitivity - 0.641; Specificity - 0.781; PPV - 0.472; NPV - 0.877; PSI - 0.349
	EET Status at 1 year	0.736 (95%CI: 0.702 - 0.771)	N.R.	N.R.	Classification Accuracy - 0.693 (95%CI: 0.660, 0.725); Balanced Accuracy - 0.694 (95%CI: 0.562, 0.812); Sensitivity - 0.722 (95%CI: 0.573, 0.821); Specificity - 0.666 (95%CI: 0.550, 0.803); PPV - 0.719 (95%CI: 0.673, 0.785); NPV - 0.668 (95%CI: 0.606, 0.736); PSI - 0.387 (95%CI: 0.279, 0.521)
	GAF ≥65 at 1 year	0.731 (95%CI: 0.697, 0.765)	N.R.	N.R.	Classification Accuracy - 0.687 (95%CI: 0.657, 0.718); Balanced Accuracy - 0.691 (95%CI: 0.541, 0.825); Sensitivity - 0.722 (95%CI: 0.487, 0.778); Specificity - 0.660 (95%CI: 0.594, 0.871); PPV - 0.650 (95%CI: 0.616, 0.769); NPV - 0.726 (95%CI: 0.655, 0.766); PSI - 0.376 (95%CI: 0.271 - 0.535)
	Andreasen symptom remission (6 months duration) at 1 year	0.703 (95%CI: 0.664, 0.742)	N.R.	N.R.	Classification Accuracy - 0.670 (95%CI: 0.636, 0.703); Balanced Accuracy - 0.668 (95%CI: 0.518, 0.827); Sensitivity - 0.584 (95%CI: 0.491, 0.827); Specificity - 0.751 (95%CI: 0.544, 0.827); PPV - 0.679 (95%CI: 0.601, 0.739); NPV - 0.667 (95%CI: 0.631, 0.734); PSI - 0.346 (95%CI: 0.232, 0.473)
LEIGHTON 2019 (2)	Quality of life at 1 year	0.704 (95%CI: 0.667, 0.742)	N.R.	N.R.	Classification Accuracy - 0.668 (95%CI: 0.632, 0.704); Balanced Accuracy - 0.667 (95%CI: 0.532, 0.789); Sensitivity - 0.623 (95%CI: 0.512, 0.774); Specificity - 0.711 (95%CI: 0.551, 0.803); PPV - 0.633 (95%CI: 0.575, 0.701); NPV 0.700 (95%CI: 0.659, 0.759); PSI - 0.333 (95%CI: 0.234, 0.460)
Studies Reporting External Validity					
	EET status at 1 year	0.876 (95%CI: 0.864, 0.887)	N.R.	N.R.	Classification Accuracy - 0.851; Balanced Accuracy - 0.845; Sensitivity - 0.815; Specificity - 0.875; PPV - 0.815; NPV - 0.875; PSI - 0.690
	Andreasen symptom remission (no duration criteria) at 1 year	0.652 (95%CI: 0.635, 0.670)	N.R.	N.R.	Classification Accuracy - 0.612; Balanced Accuracy - 0.623; Sensitivity - 0.578; Specificity - 0.667; PPV - 0.794; NPV - 0.424; PSI - 0.218
LEIGHTON 2019 (1)	Andreasen symptom remission (6 months duration) at 1 year	0.630 (95%CI: 0.612, 0.647)	N.R.	N.R.	Classification Accuracy - 0.625; Balanced Accuracy - 0.626; Sensitivity - 0.606; Specificity - 0.645; PPV - 0.645; NPV - 0.606; PSI - 0.251
	EET Status at 1 year	0.867 (95%CI: 0.805, 0.930)	N.R.	N.R.	Classification Accuracy - 0.838 (95%CI: 0.775, 0.894); Balanced Accuracy - 0.853 (95%CI: 0.740, 0.935); Sensitivity - 0.898 (95%CI: 0.780, 0.966); Specificity - 0.807 (95%CI: 0.699, 0.904); PPV - 0.766 (95%CI: 0.679, 0.867); NPV - 0.911 (95%CI: 0.840, 0.971); PSI - 0.677 (95%CI: 0.519, 0.838)
	Andreasen symptom remission (6 months duration) at 1 year	0.680 (95%CI: 0.587, 0.773)	N.R.	N.R.	Classification Accuracy - 0.695 (95%CI: 0.618, 0.771); Balanced Accuracy - 0.695 (95%CI: 0.535, 0.841); Sensitivity - 0.621 (95%CI: 0.455, 0.773); Specificity - 0.769 (95%CI: 0.615, 0.908); PPV - 0.729 (95%CI: 0.636, 0.854); NPV - 0.667 (95%CI: 0.593, 0.759); PSI - 0.396 (95%CI: 0.229, 0.613)
LEIGHTON 2019 (2) - Validated in UK	Quality of life at 1 year	0.679 (95%CI: 0.522, 0.836)	N.R.	N.R.	Classification Accuracy - 0.702 (95%CI: 0.596, 0.809); Balanced Accuracy - 0.729 (95%CI: 0.407, 0.917); Sensitivity - 0.957 (95%CI: 0.564, 1.000); Specificity - 0.500 (95%CI: 0.250, 0.833); PPV - 0.640 (95%CI: 0.561, 0.800); NPV - 0.900 (95%CI: 0.643, 1.000); PSI - 0.540 (95%CI: 0.204, 0.800)

LEIGHTON 2019 (2) - Validated in Denmark	EET Status at 1 year	0.660 (95%CI: 0.610, 0.710)	N.R.	N.R.	Classification Accuracy - 0.680 (95%CI: 0.609, 0.725); Balanced Accuracy - 0.655 (95%CI: 0.516, 0.774); Sensitivity - 0.584 (95%CI: 0.457, 0.723); Specificity - 0.726 (95%CI: 0.574, 0.824); PPV - 0.490 (95%CI: 0.421, 0.563); NPV - 0.793 (95%CI: 0.760, 0.831); PSI - 0.283 (95%CI: 0.181, 0.394)
	GAF ≥65 at 1 year	0.573 (95%CI: 0.504, 0.643)	N.R.	N.R.	Classification Accuracy - 0.456 (95%CI: 0.328, 0.817); Balanced Accuracy - 0.589 (95%CI: 0.234, 0.926); Sensitivity - 0.781 (95%CI: 0.233, 0.945); Specificity - 0.396 (95%CI: 0.234, 0.906); PPV - 0.179 (95%CI: 0.158, 0.333); NPV - 0.914 (95%CI: 0.876, 0.967); PSI - 0.093 (95%CI: 0.034, 0.300)
	Andreasen symptom remission (6 months duration) at 1 year	0.616 (95%CI: 0.553, 0.679)	N.R.	N.R.	Classification Accuracy - 0.618 (95%CI: 0.524, 0.704); Balanced Accuracy - 0.621 (95%CI: 0.342, 0.864); Sensitivity - 0.612 (95%CI: 0.306, 0.843); Specificity - 0.629 (95%CI: 0.378, 0.885); PPV - 0.476 (95%CI: 0.412, 0.636); NPV - 0.742 (95%CI: 0.687, 0.829); PSI - 0.217 (95%CI: 0.099, 0.465)
	Quality of life at 1 year	0.556 (95%CI: 0.481, 0.631)	N.R.	N.R.	Classification Accuracy - 0.589 (95%CI: 0.540, 0.637); Balanced Accuracy - 0.589 (95%CI: 0.312, 0.845); Sensitivity - 0.876 (95%CI: 0.419, 0.947); Specificity - 0.301 (95%CI: 0.204, 0.743); PPV - 0.559 (95%CI: 0.527, 0.642); NPV - 0.706 (95%CI: 0.555, 0.841); PSI - 0.265 (95%CI: 0.081, 0.483)
LEIGHTON 2021	Andreasen symptom remission (6 months duration)	0.73 (95%CI: 0.71, 0.75)	α - 0.12 (95%CI: 0.02, 0.22); β - 0.98 (95%CI: 0.85, 1.11); Calibration plot	N.R.	N.R.
PUNTIS 2021	Psychiatric hospitalisation after discharge from early intervention	0.70 (95%CI: 0.66, 0.75)	α - -0.01 (95%CI: -0.17, 0.167); β - 1.00 (95%CI: 0.78, 1.22); Calibration plot	Brier score - 0.094	N.R.

N.R. – not reported; EET – employment, education or training; GAF – Global Assessment of Functioning; DAS – Disability Assessment Schedule; PPV – positive predictive value; NPV – negative predictive value; PSI – prognostic summary index; f/u – follow-up


Records identified through database searches (n = 2353)
Additional sources (n = 67)
Records after duplicates removed (n = 1543)
Titles and abstracts screened (n = 1543)
Excluded (n = 1461)
Full text articles assessed (n = 82)
Excluded (n = 69)
• Not a prognostic prediction modelling study (n=46)
• Abstract only (n=9)
• Wrong patient population (n=7)
• Outcome ≤ 3 months from baseline (n=6)
• Duplicate publication (n=1)
Studies included in qualitative synthesis (n = 13)

Supplementary Table 1 – PROBAST risk of bias for each study

Study ID	Participants	Predictors	Outcome	Analysis	Overall
AJNAKINA 2020	Low	Low	Low	High	High
BHATTACHARYYA 2021	Low	Low	Low	High	High
CHUA 2019	Low	High	Low	High	High
DEMJAHA 2017	Low	Low	Low	High	High
DENIJS 2019	Low	Low	Low	High	High
DERKS 2010	Low	Low	Low	High	High
FLYCKT 2006	Low	Low	Low	High	High
GONZALEZ-BLANCH 2010	Low	Low	Low	High	High
KOUTSOULERIS 2016	Low	Low	Low	High	High
LEIGHTON 2019 (1)	Low	Low	Low	High	High
LEIGHTON 2019 (2)	Low	Low	Low	High	High
LEIGHTON 2021	Low	Low	Low	Low	Low
PUNTIS 2021	Low	Low	Low	Low	Low

Population	Patients with a first episode of psychosis
Intervention (model)	Any prognostic prediction model
Comparator	N/A
Outcome(s)	Any outcome
Timing	Greater than three months from baseline
Setting	Any setting

PsycINFO Search:

Psychosis Terms:

1.Acute Psychosis/ or Psychosis/
2.first episode psychosis.m_titl.
3.psychosis.m_titl.

4. 1 or 2 or 3

Outcomes Terms:

5.Treatment Outcomes/ or Health Outcomes/ or Psychotherapeutic Outcomes/ or Psychosocial Outcomes/ or Symptom Remission/
6.“recovery (disorders)”/ or relapse prevention/
7.treatment resistant disorders/
8.“quality of life”/ or “health related quality of life”/ or “quality of work life”/
9.vocational rehabilitation/
10.relapse prevention.m_titl.
11.(outcome* or remission or recovery).m_titl.
12.“treatment resis*.”.m_titl.
13.quality of life.m_titl.
14.social recovery.m_titl.
15.vocational recovery.m_titl.

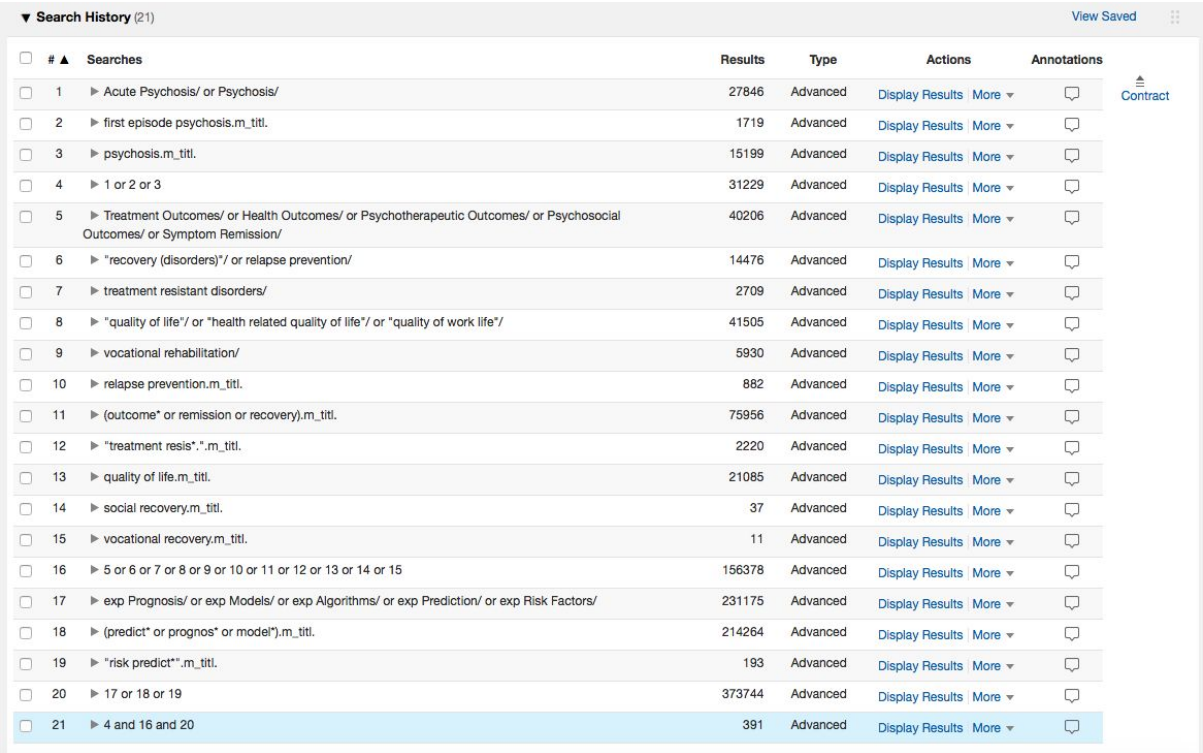
16. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15

Prediction Terms:

17.exp Prognosis/ or exp Models/ or exp Algorithms/ or exp Prediction/ or exp Risk Factors/
18.(predict* or prognos* or model*).m_titl.
19.“risk predict*”.m_titl.

20. 17 or 18 or 19

21. 4 and 16 and 20



▼ Search History (21) View Saved

	# ▲	Searches	Results	Type	Actions	Annotations
☐	1	▶ Acute Psychosis/ or Psychosis/	27846	Advanced	Display Results | More	
☐	2	▶ first episode psychosis.m_titl.	1719	Advanced	Display Results | More	
☐	3	▶ psychosis.m_titl.	15199	Advanced	Display Results | More	
☐	4	▶ 1 or 2 or 3	31229	Advanced	Display Results | More	
☐	5	▶ Treatment Outcomes/ or Health Outcomes/ or Psychotherapeutic Outcomes/ or Psychosocial Outcomes/ or Symptom Remission/	40206	Advanced	Display Results | More	
☐	6	▶ "recovery (disorders)"/ or relapse prevention/	14476	Advanced	Display Results | More	
☐	7	▶ treatment resistant disorders/	2709	Advanced	Display Results | More	
☐	8	▶ "quality of life"/ or "health related quality of life"/ or "quality of work life"/	41505	Advanced	Display Results | More	
☐	9	▶ vocational rehabilitation/	5930	Advanced	Display Results | More	
☐	10	▶ relapse prevention.m_titl.	882	Advanced	Display Results | More	
☐	11	▶ (outcome* or remission or recovery).m_titl.	75956	Advanced	Display Results | More	
☐	12	▶ "treatment resis*.".m_titl.	2220	Advanced	Display Results | More	
☐	13	▶ quality of life.m_titl.	21085	Advanced	Display Results | More	
☐	14	▶ social recovery.m_titl.	37	Advanced	Display Results | More	
☐	15	▶ vocational recovery.m_titl.	11	Advanced	Display Results | More	
☐	16	▶ 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	156378	Advanced	Display Results | More	
☐	17	▶ exp Prognosis/ or exp Models/ or exp Algorithms/ or exp Prediction/ or exp Risk Factors/	231175	Advanced	Display Results | More	
☐	18	▶ (predict* or prognos* or model*).m_titl.	214264	Advanced	Display Results | More	
☐	19	▶ "risk predict*".m_titl.	193	Advanced	Display Results | More	
☐	20	▶ 17 or 18 or 19	373744	Advanced	Display Results | More	
☐	21	▶ 4 and 16 and 20	391	Advanced	Display Results | More	

Contract

EMBASE Search:

Psychosis Terms:

1.*acute psychosis/ or *psychosis/
2.psychosis.m_titl.

3. 1 or 2

Outcomes Terms:

4.*treatment outcome/
5.*outcomes research/
6.*remission/
7.*“quality of life”/
8.*relapse/
9.*vocational rehabilitation/
10.relapse prevention.m_titl.
11.(outcome* or remission or recovery).m_titl.
12.“treatment resis*”.m_titl.
13.quality of life.m_titl.
14.social recovery.m_titl.
15.vocational recovery.m_titl.

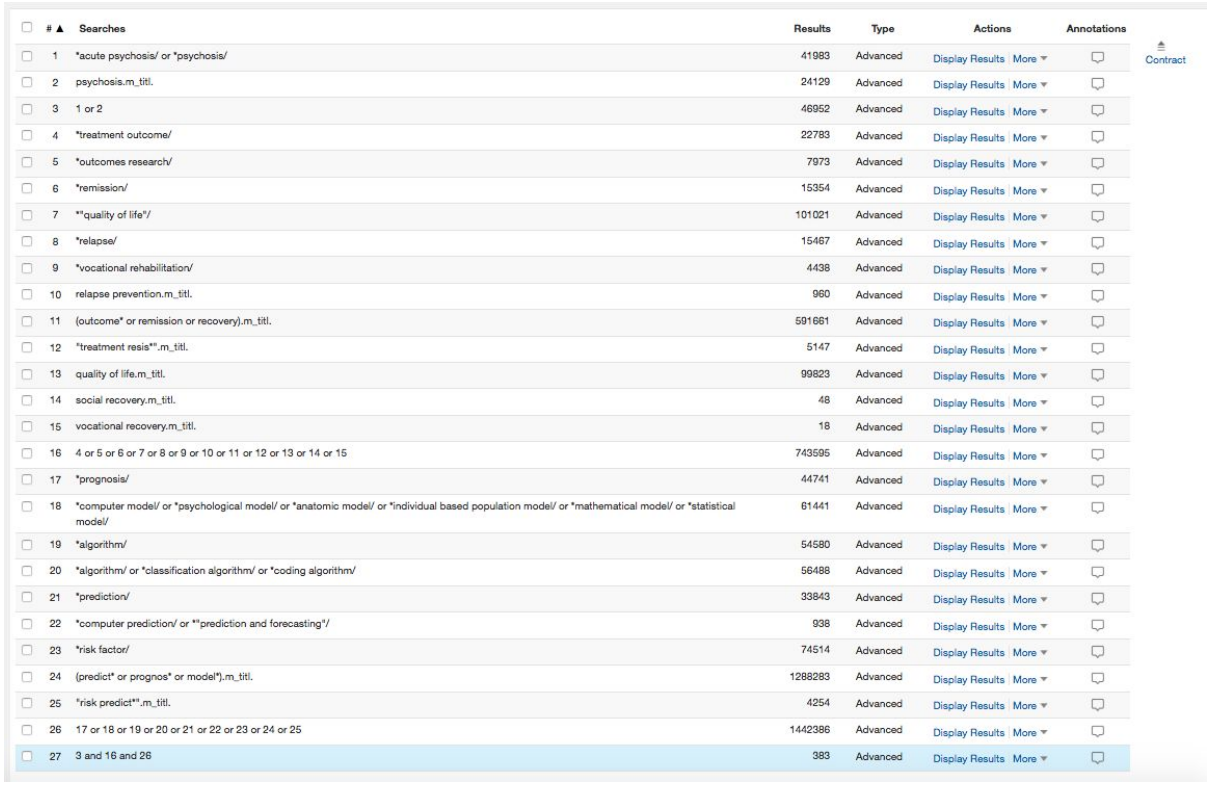
16. 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15

Prediction Terms:

17.*prognosis/
18.*computer model/ or *psychological model/ or *anatomic model/ or *individual based population model/ or *mathematical model/ or *statistical model/
19.*algorithm/
20.*algorithm/ or *classification algorithm/ or *coding algorithm/
21.*prediction/
22.*computer prediction/ or *“prediction and forecasting”/
23.*risk factor/
24.(predict* or prognos* or model*)m_titl.
25.“risk predict*”.m_titl.

26. 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25

27. 3 and 16 and 26



# ▲	Searches	Results	Type	Actions	Annotations
1	*acute psychosis/ or *psychosis/	41983	Advanced	Display Results | More	
2	psychosis.m_titl.	24129	Advanced	Display Results | More	
3	1 or 2	46952	Advanced	Display Results | More	
4	*treatment outcome/	22783	Advanced	Display Results | More	
5	*outcomes research/	7973	Advanced	Display Results | More	
6	*remission/	15354	Advanced	Display Results | More	
7	*"quality of life"/	101021	Advanced	Display Results | More	
8	*relapse/	15467	Advanced	Display Results | More	
9	*vocational rehabilitation/	4438	Advanced	Display Results | More	
10	relapse prevention.m_titl.	960	Advanced	Display Results | More	
11	(outcome* or remission or recovery).m_titl.	591661	Advanced	Display Results | More	
12	*treatment resis*".m_titl.	5147	Advanced	Display Results | More	
13	quality of life.m_titl.	99823	Advanced	Display Results | More	
14	social recovery.m_titl.	48	Advanced	Display Results | More	
15	vocational recovery.m_titl.	18	Advanced	Display Results | More	
16	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	743595	Advanced	Display Results | More	
17	*prognosis/	44741	Advanced	Display Results | More	
18	*computer model/ or *psychological model/ or *anatomic model/ or *individual based population model/ or *mathematical model/ or *statistical model/	61441	Advanced	Display Results | More	
19	*algorithm/	54580	Advanced	Display Results | More	
20	*algorithm/ or *classification algorithm/ or *coding algorithm/	56488	Advanced	Display Results | More	
21	*prediction/	33843	Advanced	Display Results | More	
22	*computer prediction/ or *"prediction and forecasting"/	938	Advanced	Display Results | More	
23	*risk factor/	74514	Advanced	Display Results | More	
24	(predict* or prognos* or model*).m_titl.	1288283	Advanced	Display Results | More	
25	*risk predict*".m_titl.	4254	Advanced	Display Results | More	
26	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25	1442386	Advanced	Display Results | More	
27	3 and 16 and 26	383	Advanced	Display Results | More	

Contract

CINAHL Plus Search:

Psychosis Terms:

S1 (MH "Psychotic Disorders")
S2 TI psychosis

S3 S1 or S2

Outcomes Terms:

S4 (MH "Outcomes (Health Care)") OR (MH "Treatment Outcomes") OR (MH "Outcomes Research")
S5 (MH "Recovery")
S6 (MH "Quality of Life") OR (MH "Psychological Well-Being")
S7 (MH "Rehabilitation, Vocational") OR (MH "Rehabilitation, Psychosocial")
S8 TI relapse prevention
S9 TI (outcome* OR remission OR recovery)
S10 TI treatment resis*
S11 TI quality of life
S12 TI social recovery
S13 TI vocational recovery

S14 S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13

Prediction Terms:

S15 (MH "Prognosis")
S16 (MH "Models, Psychological") OR (MH "Models, Anatomic") OR (MH "Models, Statistical")
S17 (MH "Algorithms")
S18 (MH "Predictive Research")
S19 (MH "Risk Factors")
S20 TI risk predict*
S21 TI (predict* OR prognos* OR model*)

S22 S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21

S23 S3 AND S14 AND S22

Search ID#	Search Terms	Search Options	Actions
S23	S3 AND S14 AND S22	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (246) | View Details | Edit
S22	S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (467,072) | View Details | Edit
S21	TI (predict* OR prognos* OR model*)	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (217,676) | View Details | Edit
S20	TI risk predict*	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (6,385) | View Details | Edit
S19	(MH "Risk Factors")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (162,669) | View Details | Edit
S18	(MH "Predictive Research")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (2,092) | View Details | Edit
S17	(MH "Algorithms")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (33,815) | View Details | Edit
S16	(MH "Models, Psychological") OR (MH "Models, Anatomic") OR (MH "Models, Statistical")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (43,628) | View Details | Edit
S15	(MH "Prognosis")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (73,492) | View Details | Edit
S14	S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (592,390) | View Details | Edit
S13	TI vocational recovery	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (13) | View Details | Edit
S12	TI social recovery	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (211) | View Details | Edit
S11	TI quality of life	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (36,786) | View Details | Edit
S10	TI treatment resis*	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (2,695) | View Details | Edit
S9	TI (outcome* OR remission OR recovery)	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (167,643) | View Details | Edit
S8	TI relapse prevention	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (417) | View Details | Edit
S7	(MH "Rehabilitation, Vocational") OR (MH "Rehabilitation, Psychosocial")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (8,270) | View Details | Edit
S6	(MH "Quality of Life") OR (MH "Psychological Well-Being")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (118,485) | View Details | Edit
S5	(MH "Recovery")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (29,255) | View Details | Edit
S4	(MH "Outcomes (Health Care)") OR (MH "Treatment Outcomes") OR (MH "Outcomes Research")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (361,369) | View Details | Edit
S3	S1 OR S2	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (12,635) | View Details | Edit
S2	TI psychosis	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (6,103) | View Details | Edit
S1	(MH "Psychotic Disorders")	**Expanders** - Apply equivalent subjects **Search modes** - Boolean/Phrase	View Results (10,866) | View Details | Edit

Web of Science – Core Collection Search:

Psychosis Terms:

#1 TS=Psychosis

Outcome Terms:

#2 TI=(outcome* OR recovery OR remission OR “quality of life” OR treatment resis*)

Prediction Terms:

#3 TI=(predict* OR prognos* OR model*)

#4 #3 AND #2 AND #1

# 4	493	#3 AND #2 AND #1 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years*	Edit	☐	☐
# 3	3,154,077	TI=(predict* OR prognos* OR model*) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years*	Edit	☐	☐
# 2	783,746	TI=(outcome* OR recovery OR remission OR "quality of life" OR treatment resis*) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years*	Edit	☐	☐
# 1	62,890	TS=Psychosis *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years*	Edit	☐	☐

PubMed Search:

Psychosis Terms

#1 psychosis[Title/Abstract]

Outcome Terms

#2 ((((((((((“outcome assessment, health care”[MeSH Major Topic]) OR “treatment outcome”[MeSH Major Topic]) OR “quality of life”[MeSH Major Topic]) OR “mental health recovery”[MeSH Major Topic]) OR “rehabilitation, vocational”[MeSH Major Topic]) OR relapse prevention[Title]) OR treatment resis*[Title]) OR outcome*[Title]) OR remission[Title]) OR recovery[Title]) OR “quality of life”[Title]) OR social recovery[Title]) OR vocational recovery[Title]

Prediction Terms

#3 (((((((((“prognosis”[MeSH Major Topic]) OR “forecasting”[MeSH Major Topic]) OR “algorithms”[MeSH Major Topic]) OR “models, psychological”[MeSH Major Topic]) OR “models, statistical”[MeSH Major Topic]) OR “risk factors”[MeSH Major Topic]) OR predict*[Title]) OR prognos*[Title]) OR model*[Title]) OR risk predict*[Title]

#4

((psychosis[Title/Abstract]) AND ((((((((((“outcome assessment, health care”[MeSH Major Topic]) OR “treatment outcome”[MeSH Major Topic]) OR “quality of life”[MeSH Major Topic]) OR “mental health recovery”[MeSH Major Topic]) OR “rehabilitation, vocational”[MeSH Major Topic]) OR relapse prevention[Title]) OR treatment resis*[Title]) OR outcome*[Title]) OR remission[Title]) OR recovery[Title]) OR “quality of life”[Title]) OR social recovery[Title]) OR vocational recovery[Title])) AND ((((((((((“prognosis”[MeSH Major Topic]) OR “forecasting”[MeSH Major Topic]) OR “algorithms”[MeSH Major Topic]) OR “models, psychological”[MeSH Major Topic]) OR “models, statistical”[MeSH Major Topic]) OR “risk factors”[MeSH Major Topic]) OR predict*[Title]) OR prognos*[Title]) OR model*[Title]) OR risk predict*[Title])

History — Download history Clear history

Search	Add to builder	Query	Items found	Time
#4	Add	Search **((psychosis[Title/Abstract]) AND ((((((((((“outcome assessment, health care”[MeSH Major Topic]) OR “treatment outcome”[MeSH Major Topic]) OR “quality of life”[MeSH Major Topic]) OR “mental health recovery”[MeSH Major Topic]) OR “rehabilitation, vocational”[MeSH Major Topic]) OR relapse prevention[Title]) OR treatment resis*[Title]) OR outcome*[Title]) OR remission[Title]) OR recovery[Title]) OR “quality of life”[Title]) OR social recovery[Title]) OR vocational recovery[Title])) AND ((((((((((“prognosis”[MeSH Major Topic]) OR “forecasting”[MeSH Major Topic]) OR “algorithms”[MeSH Major Topic]) OR “models, psychological”[MeSH Major Topic]) OR “models, statistical”[MeSH Major Topic]) OR “risk factors”[MeSH Major Topic]) OR predict*[Title]) OR prognos*[Title]) OR model*[Title]) OR risk predict*[Title])**	328	10:20:08
#3	Add	Search **((((((((“prognosis”[MeSH Major Topic]) OR “forecasting”[MeSH Major Topic]) OR “algorithms”[MeSH Major Topic]) OR “models, psychological”[MeSH Major Topic]) OR “models, statistical”[MeSH Major Topic]) OR “risk factors”[MeSH Major Topic]) OR predict*[Title]) OR prognos*[Title]) OR model*[Title]) OR risk predict*[Title]**	1165269	10:19:42
#2	Add	Search **((((((((((“outcome assessment, health care”[MeSH Major Topic]) OR “treatment outcome”[MeSH Major Topic]) OR “quality of life”[MeSH Major Topic]) OR “mental health recovery”[MeSH Major Topic]) OR “rehabilitation, vocational”[MeSH Major Topic]) OR relapse prevention[Title]) OR treatment resis*[Title]) OR outcome*[Title]) OR remission[Title]) OR recovery[Title]) OR “quality of life”[Title]) OR social recovery[Title]) OR vocational recovery[Title]**	526390	10:11:26
#1	Add	Search **psychosis[Title/Abstract]**	36962	10:05:33

Google Scholar Search:

Allintitle: psychosis AND (predict OR prognos OR model)