Comment

Predictive tools in psychosis: what is 'good enough'?

Aida Seyedsalehi & Belinda Lennox

 Check for updates

Prediction tools offer great promise for clinicians in the prevention and treatment of psychosis, but none has been routinely implemented. Greater methodological rigour in the development and evaluation of these tools, along with consideration of a range of performance criteria, is necessary to maximize their potential for improving clinical decision making.

The diagnosis and treatment of early psychosis is an imprecise science. We have a strong evidence base for the treatments that we give; we know that they are effective for the majority of people. However, the same package of treatment is offered to everyone, and we are unable to say to any young person and their family what the future holds in terms of outcome or response to treatment. We do not undertake detailed investigation, whether that be an MRI scan, genetic testing or cognitive testing, to inform our prediction of outcomes.

To be able to do this would be a major advance for psychiatry, with the potential for tailored treatments for individuals and the next generation of research-informed clinical practice. In a Review in this issue, Fiona Coutts and colleagues provide a comprehensive overview of the research that is underway to develop such tools for psychosis, aiming to predict features ranging from the onset of psychosis through to response to treatment in those with established illness and longer-term outcomes. Here, we build on this overview to discuss the important issue of how to gauge whether predictive tools are good enough to be implemented in clinical care, and provide our opinions on next steps for the field

The most developed of the tools discussed by Coutts and colleagues are those to predict the onset of psychosis. Here there are some promising potential prediction models that combine biological and clinical variables in those at clinical high risk (CHR) of psychosis to predict conversion to psychosis. However, there are some important barriers to these tools being useful for patients and clinicians. One crucial issue is in the clinical utility of using CHR of psychosis diagnosis as the starting point. The diagnosis of CHR of psychosis was constructed with the aim of identifying the very early stage of psychosis to provide an opportunity for prevention. Diagnosis focuses largely on the recognition of attenuated psychotic symptoms. However, recent studies have shown that only a small minority of people with a first episode of psychosis will have initially had this diagnosis, even in places with very well established CHR clinical services (for example, 14% in Melbourne, Australia[1]). The lack of precision of the CHR diagnosis in identifying prodromal psychosis was highlighted by a study from Finland that found the same risk of future psychosis in all individuals attending emergency departments with self-harm as in individuals with a diagnosis of CHR of psychosis[2]. This Finnish study also found that a much more common precedent to a later psychotic diagnosis was being seen by child and adolescent mental health services (CAMHS), with 50% of all those with a psychosis diagnosis initially presenting to CAMHS with a range of non-psychotic presentations[3]. Future models for predicting onset of psychosis will therefore need to consider a much broader population as a starting point.

Aside from the specifics of predicting the onset of psychosis, we think there are important methodological issues for psychiatry to consider in the future development of prediction tools for clinical use. Coutts and colleagues report that the discrimination performance of prediction models for psychotic disorders (as quantified by the area under the receiver operating characteristic curve) is generally lower than those of models that are currently used clinically in other areas of medicine (for example, cardiovascular medicine). However, we do not think any specific cut-off indicates that the models are not useful — what is considered 'good' discrimination performance depends on the clinical area and available alternatives for the same purpose, and is also a matter of judgement[4].

"we do not think any specific cut-off indicates that the models are not useful"

The literature is characterized by an overwhelming focus on model discrimination; however, other criteria for assessing model performance are equally important. In particular, assessment of calibration — which refers to the agreement between predicted and observed risks — is essential but is often ignored. A model can have perfect discriminative ability (that is, separating out patients with and without the outcome) but produce risk estimates that are unreliable[5]. For example, a model might correctly assign a higher predicted probability to an individual with the outcome than one without, but these predicted probabilities might be imprecise (for example, 30% and 50%, when the true probabilities are 15% and 30%, respectively). If such a model is used to inform patients and clinicians — for example, about the risk of developing psychosis, experiencing relapse or showing treatment resistance — its predictions will be misleading and might result in false expectations[5]. There is also potential for patient harm if treatment decisions are made on the basis of poorly calibrated models, as over-estimation or under-estimation of risk can lead to the administration of unnecessary interventions or the withholding of necessary ones[5].

To answer the question of whether a model's performance is 'good enough' to be useful in clinical practice, we need to look beyond

discrimination and calibration. These are useful statistical measures of a model's predictive performance, but to claim that a prediction model can improve decisions about patient care – such as the decision to offer clozapine to individuals at high risk of treatment resistance – we need measures of clinical utility. One such measure is net benefit, which can capture the clinical consequences of using a model for decision making[6]. Net benefit weights the benefits associated with using the model (such as improved prognosis) against the harms (such as the costs and adverse effects of unnecessary treatment) by putting them on the same scale[6]. We agree with the recommendation by Coutts and colleagues that assessment of net benefit should be an important requirement for any model intended to support clinical decision making.

Ultimately, however, to determine whether a prediction model should be implemented in clinical practice, we need evidence of its effects on clinicians' behaviour, patient outcomes and the cost-effectiveness of care[4]. Such evidence can be obtained through cost-effectiveness modelling and prospective impact studies, ideally using a cluster randomized design[4]. Although these studies are time-consuming and resource-intensive, they are an important step for promising models that have been adequately validated and show preliminary evidence of clinical usefulness in decision curve analysis.

"we must encourage adherence to rigorous methodological and reporting standards"

To facilitate the implementation of prediction models in routine clinical practice, Coutts and colleagues highlight several important future directions for the field. In addition to these, we must encourage adherence to rigorous methodological and reporting standards. As highlighted in a recently published systematic review of prediction models in first-episode psychosis[7], much of the published literature contains numerous methodological shortcomings. Most prediction models are based on insufficient sample sizes and are developed using unsuitable statistical methods, including biased selection of predictor variables and inappropriate handling of missing data. Most lack appropriate internal validation, and many fail to assess key measures of model performance. Therefore, future studies clearly need to follow best-practice guidelines – an excellent collection of methods guidance for model development, validation and reporting is available at https://www.prognosisresearch.com/. Promising examples exist of high-quality studies that have developed and validated prediction models for other adverse outcomes, such as violence perpetration in people with psychosis, but they still need to be further assessed for clinical utility[8].

Finally, more studies should focus on external validation of existing models in independent data, including head-to-head comparisons of competing models designed for the same purpose, as was recently done for models predicting short-term mortality following hospital admission for COVID-19 (ref. [9]). The most promising opportunities for such comparisons lie in 'big' data, obtained from large electronic health records databases or through collaborative efforts aimed at meta-analysing individual participant data across multiple studies[10].

Aida Seyedsalehi ⓘ ⓘ **& Belinda Lennox** ⓘ ⓘ ✉
Department of Psychiatry, University of Oxford, Oxford, UK.
✉e-mail: belinda.lennox@psych.ox.ac.uk

References

1. Burke, T. et al. Proportion and characteristics of young people in a first-episode psychosis clinic who first attended an at-risk mental state service or other specialist youth mental health service. *Schizophr. Res.* **241**, 94–101 (2022).
2. Bolhuis, K. et al. Hospital presentation for self-harm in youth as a risk marker for later psychotic and bipolar disorders: a cohort study of 59 476 Finns. *Schizophr. Bull.* **47**, 1685–1694 (2021).
3. Lång, U. et al. Potential for prediction of psychosis and bipolar disorder in Child and Adolescent Mental Health Services: a longitudinal register study of all people born in Finland in 1987. *World Psychiatry* **21**, 436–443 (2022).
4. Moons, K. G. M., Altman, D. G., Vergouwe, Y. & Royston, P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* **338**, b606 (2009).
5. Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L. & Steyerberg, E. W. Calibration: the Achilles heel of predictive analytics. *BMC Med.* **17**, 230 (2019).
6. Vickers, A. J., Van Calster, B. & Steyerberg, E. W. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* **352**, i6 (2016).
7. Lee, R. et al. Prediction models in first-episode psychosis: systematic review and critical appraisal. *Br. J. Psychiatry* **220**, 1–13 (2022).
8. Whiting, D., Mallett, S., Lennox, B. & Fazel, S. Assessing violence risk in first episode psychosis: external validation, updating and net benefit of a prediction tool (OxMIV). *BMJ Mental Health* (in the press).
9. de Jong, V. M. T. et al. Clinical prediction models for mortality in patients with COVID-19: external validation and individual participant data meta-analysis. *BMJ* **378**, e069881 (2022).
10. Riley, R. D. et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* **353**, i3140 (2016).