Risks and benefits of probabilistic suicide risk tools

Seena Fazel,[1,2] Bo Runeson,[3] Thomas Fanshawe,[4] Michael Sharpe[1]

1 Department of Psychiatry, University of Oxford, Oxford, UK;

2 NIHR Oxford Health Biomedical Research Centre, Oxford, UK;

3 Department of Clinical Neurosciences, Karolinska Institutet, Stockholm, Sweden;

4 Nuffield Department of Primary Health Care Sciences, University of Oxford, Oxford, UK;

Corresponding author: Professor Seena Fazel (seena.fazel@psych.ox.ac.uk)

Department of Psychiatry, University of Oxford, Warneford Lane, Headington, Oxford, OX3 7JX, United Kingdom

Precision psychiatry will require the development and application of evidence-based tools for diagnosis, prognosis, and treatment matching. In prognosis, assessment of suicide risk remains central to psychiatric practice, and can assist with informing decisions about further assessment and management. Such prognostic assessments are important considerations in deciding who is referred to more intensive services, such as crisis or assertive outreach teams and hospitalization (where in England and Wales, grounds for admission under the Mental Health Act include 'in the interests of his own health or safety').

The OxSATS tool is an externally validated and scalable prognostic model for suicide risk in people that present to hospital services with self-harm (1). It was developed in a cohort of 53,172 people with self-harm presentations and followed up for 2-year suicide mortality. Survival analysis methods were used to model the association between routinely collected prognostic factors and suicide. The final model contains 11 factors including age, sex, mental health and treatment, substance misuse, and history of self-harm. In a geographical external validation, OxSATS showed good discrimination (c-index 0.77, 95% CI: 0.75-0.78), and calibration. The online version outputs a probability score with no fixed cut-offs, which can be adapted to suit the clinical setting, linked interventions and available resources. This is consistent with prognostic models used in the rest of medicine, including cardiovascular (e.g. Framingham) and pulmonary (BODE) medicine, diabetes (SCORE), and cancer.

The Feature by Cockburn and Large (2) prompts a discussion of potential risks and benefits of suicide risk assessment tools. However, we argue that its conclusions are misleading as it both misrepresents the risks and benefits of OxSATS and does not consider the wider evidence on feasibility and acceptability of probabilistic tools for suicide assessment. We seek to address these shortcomings.

First, clinical override is necessary when using suicide risk assessment tools. Although tools can potentially inform and provide consistency to clinical decision-making, there are additional individual factors that should be considered. In this respect, the Feature's use of vignettes overlooks this important part of how tools should be used. Furthermore, it underscores the importance of clinical override for AI-based approaches to decision-making.

Second, evaluation of risk tools needs to consider model performance and feasibility separately. We think the Feature includes two errors: (a) conflating these two ways of

evaluating tools, and (b) not considering the most important measure of predictive performance for probabilistic tools, calibration (how well expected probabilities compare with actual outcomes). Instead, the Feature starts with presenting sensitivity and positive predictive value (PPV). These classification measures are dependent on choice of cut-off, which will necessarily vary by setting and context (and because of this we do not recommend using fixed cut-off scores). They inadvertently make this point by failing to mention the sensitivity (82%) of the primary cut-off of the OxSATS external validation (1), which undermines Cockburn and Large's statement that the tool will miss more than half suicide cases. The authors also refer to the area under the curve (AUC) statistic, a measure of discrimination, and state that OxSATS is similar to the 4-item Manchester Self-Harm Rule. However, this statement is incorrect – AUCs are insensitive to changes in model performance, cannot distinguish between well-calibrated and poorly-calibrated models, and differences between OxSATS (AUC 0.76) and the other tools is in fact large. Comparison with the alternative, unstructured clinical judgement, is relevant – the only study to our knowledge to investigate clinically-based suicide risk assessment after self-harm found an AUC of 0.54 over 12 months (3). In the US, the field has progressed beyond simplistic tool-vs-clinician debates, with research now focused on collaborative approaches in which suicide tools augment—rather than compete with—clinical judgement to improve outcomes (4).

Third, if used properly, vignettes can be a useful way to consider feasibility and acceptability. The use of vignettes requires rigorous qualitative methodology to examine the tool's influence on clinical decision-making, rather than as in the Feature using hypothetical short vignettes constructed and interpreted by the same two authors.

Fourth, whilst there is increasing evidence on the feasibility and acceptability of suicide tools, the Feature does not engage with this. A recent Harvard study interviewed 56 clinicians in 10 focus groups across psychiatry, emergency medicine, and other specialities (5). They concluded that the 'predominant attitude' towards structured risk tools was positive, particularly for informing clinical decision-making and treatment planning. Clinicians welcomed the potential of tools to identify higher risk patients that could be missed using only subjective clinical judgement. Evidence from other areas of medicine shows how AI tools can lead to more time with patients, improve care by incorporating more data to personalise prognosis and treatment, and reduce medical errors (6). A recent synthesis of

RCTs of AI models, including prognostic ones, found statistically significant improvements in clinical decision-making in 6 of 7 trials, and in patient care in 15 of 18 trials (7).

Fifth, the Feature makes unsupported generalizations. It states that tools risk ‘disempowering’ patients, who ‘tend to dislike’ them, with no evidence presented. But if presented in an appropriate way, tools can provide a basis for collaborative management, underscore safety planning and enable fairer resource allocation. The authors state that hospitalization ‘does not prevent suicide’, which it may do in acute crises. They assert that QRISK allows patients to explore how risk of cardiovascular events alters if they were to make changes to QRISK predictors. This is not correct - neither QRISK nor OxSATS are causal prediction models and should not be interpreted as such. Further, OxSATS should not be used to justify withdrawal of treatment as the authors suggest in discussion of vignettes 5 and 6.

Sixth, the Feature does not address the fact that suicide risk assessment is now an outlier in psychiatry. Prognostic models are increasingly advocated for personalising prognosis in schizophrenia and other psychiatric conditions. In forensic psychiatry, risk assessment tools for violence are routinely used (8), and one of the Feature’s authors recently recommended using one of these tools to stratify risk into low/medium/high for inpatient violence (9).

Finally, we accept that lack of validation in a new setting can limit generalizability. This should occur alongside robust qualitative research exploring feasibility and acceptability in clinical practice. Taken together, these lines of evidence underscore the need for a balanced, empirically grounded, and scientifically rigorous approach to investigating probabilistic suicide risk tools as aids to clinical decision-making. Such decision-making should be based on an empathetic and clinically relevant interview that includes comprehensive assessment of history, mental state, risk and needs. The possibility that structured evidence-based tools can augment, supplement and complement decision-making in one of the most challenging areas of psychiatry is an empirical question that should not be prematurely shut down by clinical overconfidence and inappropriate comparisons. It requires an open debate that considers whether and how tools can work alongside clinical judgement to improve patient outcomes.

Acknowledgements

All authors are part of the original team that developed the OxSATS model. We report no other conflicts of interest.

Funding

This study received no specific grant from any funding agency, commercial or not-for-profit sectors. SF is supported by the NIHR Oxford Health Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Author Contributions

SF drafted the commentary. All other authors critically revised and approved it for submission.

References

1. Fazel S, Vazquez-Montes M, Molero Y, Runeson B, D'Onofrio BM, Larsson H, et al. Risk of death by suicide following self-harm presentations to healthcare: development and validation of a multivariable clinical prediction rule (OxSATS). BMJ Ment Health. 2023;26(1).
2. Cockburn A, Large M. A clinician's guide to probabilistic suicide risk prediction tools: cautions and pitfalls. Br J Psychiatry. 2025:1–5.
3. Randall JR, Sareen J, Chateau D, Bolton JM. Predicting Future Suicide: Clinician Opinion versus a Standardized Assessment Tool. Suicide Life Threat Behav. 2019;49(4):941–51.
4. Simon GE, Matarazzo BB, Walsh CG, Smoller JW, Boudreaux ED, Yarborough BJH, et al. Reconciling Statistical and Clinicians' Predictions of Suicide Risk. Psychiatr Serv. 2021;72(5):555–62.
5. Bentley KH, Zuromski KL, Fortgang RG, Madsen EM, Kessler D, Lee H, et al. Implementing Machine Learning Models for Suicide Risk Prediction in Clinical Practice: Focus Group Study With Hospital Providers. JMIR Form Res. 2022;6(3):e30946.
6. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nature Medicine. 2022;28(1):31–8.
7. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digital Health. 2024;6(5):e367–e73.
8. Ogonah MGT, Seyedsalehi A, Whiting D, Fazel S. Violence risk assessment instruments in forensic psychiatric populations: a systematic review and meta-analysis. Lancet Psychiatry. 2023;10(10):780–9.
9. Picchioni M, Ruiz R, de Girolamo G, Iozzino L, Zamparini M, Wancata J, et al. The predictive validity and temporal characteristics of the HCR-20v3 for inpatient violence in forensic inpatient settings. An international study. Psychiatry Research. 2024;339:116079.