External validation of the Oxford Suicide Assessment Tool after Self-harm (OxSATS)

11. Risk assessment and screening

Tyra Lagerberg[1, 2] , *Maria Vazquez-Montes*[3], *Thomas Fanshawe*[3], *Seena Fazel*[1, 4]

[1] Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK

[2] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[3] Nuffield Department of Primary Health Care Sciences, University of Oxford, Oxford, UK

[4] Oxford Health NHS Foundation Trust, Oxford, UK

Abstract text

Background – OxSATS is a well-performing tool to predict suicide after self-harm. It was originally developed in Swedish national register data – the performance in other national contexts is unknown.

Aim – To externally validate the OxSATS tool in data from Oxford, United Kingdom.

Method – We selected individuals aged ≥10 years who presented with self-harm to an Oxford emergency department between 1 January 2000 and 31 December 2016. We calculated individuals' estimated risk of suicide within 12 months using the OxSATS model. Model performance was assessed using discrimination and calibration metrics.

Results – We identified 14,321 individuals with self-harm, of which 61 (0.4%) died by suicide within 12 months. OxSATS showed good discrimination (Harrell's c-index=0.75, 95%CI=0.72-0.77). At the 1% risk cut-off, sensitivity was 82% and specificity 60%. The calibration was poor (Observed:Expected events=0.41, 95% CI=0.32, 0.53) due to low outcome prevalence in the Oxford sample (0.4% versus 1% in the OxSATS development sample). We were not able to recalibrate the model due to the low number of outcome events.

Conclusion – OxSATS performs well in terms of discrimination in Oxford data, meaning it is good at sorting individuals by their risk of suicide. This suggests potential for clinical application. However, suicide risk prediction models require recalibration when the outcome prevalence differs from that in the development sample. Recalibrating the OxSATS model requires a larger sample – we plan to test this in future work on updated data.