

# Predicting suicide risk in 137 112 people with severe mental illness in Finland: external validation of the Oxford Mental Illness and Suicide tool (OxMIS)

Amir Sariaslan (✉ [amir.sariaslan@psych.ox.ac.uk](mailto:amir.sariaslan@psych.ox.ac.uk))

University of Oxford <https://orcid.org/0000-0002-6151-3278>

Thomas Fanshawe

Joonas Pitkänen

Andrea Cipriani

University of Oxford <https://orcid.org/0000-0001-5179-8321>

Pekka Martikainen

Seena Fazel

University of Oxford

---

## Article

### Keywords:

**Posted Date:** September 2nd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1959949/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Oxford Mental Illness and Suicide tool (OxMIS) is a standardised, scalable, and transparent instrument for suicide risk assessment in people with severe mental illness (SMI) based on 17 sociodemographic, criminal history, familial, and clinical risk factors. However, alongside most prediction models in psychiatry, external validations are currently lacking. We utilized a Finnish population sample of all persons diagnosed by mental health services with SMI (schizophrenia-spectrum and bipolar disorders) between 1996 and 2017 ( $n = 137\,112$ ). To evaluate the performance of OxMIS, we initially calculated the predicted 12-month suicide risk for each individual by weighting risk factors by effect sizes reported in the original OxMIS prediction model and converted to a probability. This probability was then used to assess the discrimination and calibration of the OxMIS model in this external sample. Within a year of assessment, 1.1% of people with SMI ( $n = 1475$ ) had died by suicide. The overall discrimination of the tool was good, with an area under the curve of 0.70 (95% confidence interval: 0.69–0.71). The model initially overestimated suicide risks in those with elevated predicted risks of > 5% over 12 months (Harrell's  $E_{\max} = 0.114$ ), which applied to 1.3% ( $n = 1780$ ) of the cohort. However, when we used a 5% maximum predicted suicide risk threshold as is recommended clinically, the calibration was excellent (ICI = 0.002;  $E_{\max} = 0.005$ ). Validating clinical prediction tools using routinely collected data can address research gaps in prediction psychiatry and is a necessary step to translating such models into clinical practice.

## Introduction

Prognostic models and tools have been increasingly implemented in routine clinical practice and support clinical decision-making in other areas of medicine, including cardiovascular<sup>1</sup> and cancer medicine.<sup>2–4</sup> However, their use is inconsistent and highly variable in mental health. Part of the reason for this is that prediction models in mental health are very rarely externally validated.<sup>5</sup> Such validation is a necessary step on the path to implementation alongside work on feasibility, acceptability and clinical impact. In the suicide field, this is not different - few models have been externally validated despite their clinical use in some settings.<sup>6</sup> Clinical surveys have found that despite routine use of risk tools in most emergency departments, around two-thirds relied on unvalidated and locally-developed instruments.<sup>7</sup> Where external validations have been reported for suicide risk assessment tools, they have commonly omitted key performance measures,<sup>8</sup> particularly calibration, which tests the agreement between observed and predicted outcomes, that is necessary to appraise them.

Oxford Mental Illness and Suicide tool (OxMIS) is a brief, scalable, structured risk assessment tool designed to assess the risk of suicide in patients with severe mental illness (schizophrenia-spectrum disorders or bipolar disorder). The tool was developed and validated in a cohort of 75,158 individuals diagnosed with severe mental illness in Sweden, is available as a free web-based calculator with its coefficients published (<https://oxrisk.com/oxmis>) and protocol available.<sup>9</sup> The model provides probability scores for a 12-month suicide risk for individuals with SMI (with a ceiling of 5%) with good measures of

discrimination (AUC = 0.71), and calibration. As OxMIS has not been validated in a new independent sample, we aimed to test its performance using population-based data from Finland over a 22-year period, which allowed us to examine the predictive performance of the tool in over 137,000 persons with SMI.

## Subjects And Methods

### Data

A unique civic registration number is assigned to every resident of Finland, allowing for the precise linking of multiple administrative population registers.<sup>10</sup> Following approvals by the Ethics Board of Statistics Finland (TK-53-1490-18) and the Finnish Institute for Health and Welfare (THL/2180/14.02.00/2020), we were granted access to pseudonymized data from governmental agencies in a secure remote access platform maintained by Statistics Finland. In Finland, informed consent is not required for population-based register studies of this nature.

We initially used the Care Register for Health Care,<sup>11</sup> which is maintained by the Finnish Institute for Health and Welfare, to identify all individuals between the ages of 15 and 65 years who were diagnosed with a severe mental illness between 1 January 1996 and 31 December 2017. The Care Register for Health Care includes all inpatient hospital episodes in Finland since 1 January 1970, classified according to the International Classification of Diseases (ICD; eighth revision [ICD-8]: 1970–1986; ninth revision [ICD-9]: 1987–1995; tenth revision [ICD-10]: 1996–2017). The register also includes outpatient specialist visits in secondary care using the same classification system, albeit during a more limited period (1998–2017) compared to the inpatient care data. Data on treatments provided during care were not available. Severe mental illness (SMI) was defined as an episode of either schizophrenia-spectrum disorder (ICD-8: 295, 297–298; ICD-9: 295, 297–299; ICD-10: F20-F29) or bipolar disorder (ICD-8: 296 [excluding 296.2]; ICD-9: 296 [excluding 296.1A-D, 296.1F-G], ICD-10: F30-F31). We used a hierarchical definition, where individuals who had been diagnosed with both a schizophrenia-spectrum disorder and bipolar disorder were considered to have the former. External validation studies of single-episode diagnoses of schizophrenia-spectrum disorders (positive predictive value, PPV  $\geq$  84%),<sup>12</sup> bipolar disorder (PPV range between 87–93%),<sup>13</sup> have demonstrated excellent validity.

Using this approach, we were able to identify a cohort of 137,112 patients with 5,261,732 recorded episodes. Similar to the original study, we randomly selected one episode per patient, with equal probability, to be the index episode, as the intention of the tool is its use at a single time point. Each individual was followed up from the date of their index inpatient or outpatient episode until they either emigrated, died or reached the end of the follow-up period (12 months post-episode). Consistent with the original OxMIS study, we did not exclude individuals who did not reach the end of the follow-up period.

### Outcome definition

By linking the cohort of patients diagnosed with a SMI to the Causes of Death Register, which includes the main and contributory causes of nearly all deaths (> 99.7%)<sup>14</sup> recorded in Finland throughout the entire follow-up period, we were able to identify those who had died of suicide, either as the main or contributory causes of death, within 12 months of the index episode. Consistent with the original OxMIS study<sup>15</sup> and research literature,<sup>16</sup> we included both certain (ICD-10: X60-X84) and uncertain (ICD-10: Y10-Y34) deaths by suicide.

## Predictors

The OxMIS prediction rule uses 17 predictors representing sociodemographic (i.e., sex, age, educational attainment, benefit receipt), antisocial and suicidal (i.e., previous violent crime, drug use disorder, alcohol use disorder, self-harm), familial (i.e., parental substance use disorders, psychiatric hospitalization, and suicide), and clinical (i.e., recent treatments for antidepressants or antipsychotics, treatment in inpatient care at the time of assessment, length of stay exceeding 7 days, more than 7 previous episodes, and comorbid depression [in schizophrenia-spectrum disorders]) risk factors for suicide in Sweden. Using the Finnish registries, we created an equivalent set of predictors whose definitions we sought to match as closely as possible (**eText**). Due to the close geographical proximity between the countries and similarity of their healthcare provision, including the organization of health care and administrative data, nearly all the predictor definitions were similar. There were, however, two differences between the cohort predictors. First, we defined 'previous violent crime' as any conviction for an offence 'against life and health' (e.g., any assault, manslaughter, and murder) that had occurred prior to the admission. Although this definition is commonly used in Finnish criminological studies,<sup>17</sup> it is narrower than the equivalent measure used in the original OxMIS study, which additionally included violent property and sexual crimes, and unlawful threats. Second, the original OxMIS study lacked prescription drug data prior to July 1, 2005, thus necessitating the use of multiple imputation to estimate values for the predictors measuring recent antipsychotic and antidepressant treatment for a large proportion of the participants. In contrast, the Finnish prescription drug data were comprehensive for the duration of the entire follow-up period. There were no missing data in the predictors.

## Analytical approach

Prior to performing any data analyses, we outlined the outcomes, predictors, risk categories, and analytic strategy in a protocol (**eText**). We then conducted compare the distributions of all predictors across the present Finnish external validation sample with the OxMIS derivation sample. To quantify the associations between the predictors and death by suicide within 12 months of the index episode, we fitted a single logistic regression model to the validation sample. The estimates derived from the model were expressed as adjusted odds ratios. As equivalent estimates were provided in the original OxMIS study, we then compared the magnitude of associations for each predictor across both samples.

Following this, we evaluated the OxMIS prediction model in the Finnish validation sample by calculating the predicted suicide risk one year after the index episode. We did this by calculating the linear predictor (LP; the sum of the products between the untransformed regression coefficients from the original OxMIS

study and each of the predictors in the current study) and converting it to the probability scale [ $P; P = 1/(1 + \exp(-LP))$ ]. This continuous probability score was subsequently used to determine the discriminatory accuracy of the tool, or its ability to differentiate between those who died from suicide and those who did not. We did this by examining the receiver operating characteristic (ROC) curve, which depicts the sensitivity and specificity values of the tool across multiple risk thresholds, and estimating its area under the curve (AUC), where a value of 1 indicated perfect discrimination.<sup>18</sup>

In addition to evaluating it as a continuous probability score, we also assessed the performance based on the pre-specified 1% predicted suicide risk cut-off to calculate the number of true and false positives and negatives in our sample and to estimate measures of classification accuracy (e.g., sensitivity, specificity, positive and negative predictive values).

Calibration, or the degree to which predicted risks differ from observed ones, was primarily assessed using a calibration plot. We graphically evaluated the calibration by plotting the observed risks against the predicted risks and visualized their relationship in the form of a smoothed calibration curve, which we estimated using the 'loess' algorithm (locally weighted least squares regression smoother).<sup>19</sup>

## Deviations from the protocol

We complemented the graphical evaluation of the model calibration by estimating numeric calibration measures, including the Integrated Calibration Index (ICI), which measures the average absolute difference between the smoothed calibration curve and the observed risks.<sup>20</sup> We also considered the 90th percentile of the absolute difference between observed and predicted risks (E90), as well as the maximum absolute difference between observed and predicted risks ( $E_{\max}$ ). For the ICI, E90 and  $E_{\max}$  measures, a value of 0 indicates that the prediction model is perfectly calibrated.<sup>20</sup> Additionally, we evaluated the performance of OxMIS in accordance with how it has been applied clinically (i.e., in its online calculator), where people with a predicted suicide risk of > 5% are assumed to have a predicted suicide risk of 5%.

Stata MP 17<sup>21</sup> and R 3.5<sup>22</sup> were used for all analyses. We followed the TRIPOD reporting guidelines for prediction models (**eTable 1**).<sup>23</sup>

## Results

We identified 137 112 people with an SMI diagnosis in Finland between 1996 and 2017, of whom 101 655 (74%) were diagnosed with a schizophrenia-spectrum disorder and 35 457 (26%) with bipolar disorder. In the 12 months following their index episode, 1475 individuals within this cohort died by suicide, which represents a suicide prevalence rate of 1.1%. In comparison to the Swedish derivation cohort, the Finnish external validation cohort was younger (mean age at discharge: 41 vs. 44 years), had lower rates of previous violent crime (10% vs. 16%), drug use (8% vs. 12%), and self-harm (13% vs. 20%), and recent medication treatments for antipsychotics (45% vs. 54%) and antidepressants (25% vs. 39%), but higher rates of previous alcohol use (19% vs. 15%), benefit receipt (74% vs. 64%) and comorbid

depression in schizophrenia-spectrum disorders (43% vs. 32%) (**eTable 2**). However, there were no clear differences in the association of individual predictors on suicide (**eTable 3**).

The discriminatory accuracy of the tool was good with an area under the curve of 0.70 (95% CI: 0.69–0.71; Fig. 1). The calibration plot (Fig. 2) indicated that the calibration was very good for those with a predicted probability of 5% or less, which is the maximum risk level used in the OxMIS web calculator at present. However, the plot also showed some overestimation of the tool's prediction in those with risks > 5%, which applied to 1.3% of the cohort (n = 1780; **eTable 4**). Quantitative measures of prediction accuracy confirmed these findings, indicating that the differences between the observed and predicted suicide risks amounted to 0.2% on average (ICI: 0.002) and up to the 90th percentile (E90: 0.002), but the maximum difference was 11.4% ( $E_{\max}$ : 0.114). In complementary sensitivity analyses, we recoded all individuals with a predicted suicide risk of > 5% as having a 5% predicted suicide risk, consistent with the way OxMIS online calculator classifies. Using this updated definition, we found that the maximum difference between the observed and predicted suicide risks was reduced from 11.4–0.5%, demonstrating very good calibration (ICI: 0.002; E90: 0.002;  $E_{\max}$ : 0.005).

The number of true and false positives and negatives using a pre-specified 1% suicide risk cut-off are presented in **eTable 5**. Using this cut-off (although cut-offs are not recommended clinically but for research reporting purposes), the tool had a sensitivity of 58.8% (95% CI: 56.3%-61.4%), specificity of 70.3% (95% CI: 70.1%-70.6%), positive predictive value of 2.1% (95% CI: 2.0%-2.3%) and negative predictive value of 99.4% (95% CI: 99.3%-99.4%).

## Discussion

We evaluated the performance of a novel suicide risk assessment tool (OxMIS) in 137 112 people in Finland diagnosed with a schizophrenia-spectrum disorder or bipolar disorder. Good discrimination was indicated by an area under the curve of 0.70. This means that in 70% of the instances when we randomly select two people from our sample, one of whom had died by suicide and the other had not, the tool would give the person who had died by suicide a higher predicted suicide risk score.

Consistent with the original OxMIS study, we discovered that model overestimated risk for extremely high-risk patients (i.e., those with a predicted suicide risk of > 5%) and the calibration was poor. However, this only affected a small percentage (1.3%) of our sample who had predicted probabilities of this magnitude. In our complementary sensitivity analysis, we observed improved calibration in these patients when we assigned them a suicide risk prediction of no more than 5%. Given our findings, it appears that specific suicide risk predictions above 5 percent are unlikely to be accurate and should therefore be reported as > 5%, consistent with how the OxMIS online calculator currently reports risks.

Suicide risk prediction models are commonly limited in the reported performance metrics. Furthermore, a recent systematic review of clinical prediction models in psychiatry found that only 16% of the studies had been validated in wholly independent samples and reported measures of discrimination in

development and validation samples.<sup>5</sup> Importantly, the same review found that nearly four out of five (78%) of these studies reported poorer out-of-sample discrimination. In contrast, we found that OxMIS maintained a similar performance from its Swedish validation, with no clear differences between the reported AUCs (range of 0.70–0.71).

The strengths of our study included the use of national registers that allowed us to study suicide risk in all individuals with SMI diagnosed in the entire country of Finland over a 22-year period. In contrast to the many validation studies,<sup>24</sup> this approach further enabled us to test OxMIS on an external validation sample that was more than twice as large as the derivation sample ( $n = 137\,112$  vs.  $n = 58\,777$ ), thereby providing excellent statistical power. In Finland, the definitions of the predictors were comparable to those in the original study with no missing data. There are two methodological limitations to consider. First, our crime data consisted of criminal conviction records in which assaults, manslaughter, and murder were predefined as violent crimes. In light of this, our definition of ‘previous violent crime’, one of the 17 predictors, was narrower than that of the original OxMIS study, which included violent property and sexual crimes in addition to unlawful threats. Nonetheless, the influence of this misclassification bias remained minimal, as the tool performed similarly in both countries. Second, although we had complete coverage of inpatient care episodes throughout the entire follow-up period, outpatient care data did not begin until 1998, indicating that the prevalence rates for the predictors derived from the patient data were higher for younger cohorts due to left truncation bias. However, despite variations in the birth cohorts included and coverage of outpatient care data, which began in 2001 in Sweden, we obtained similar results to the original OxMIS study.

External validation is only one, albeit key, component of a comprehensive evaluation of any risk prediction model and tool. Other important considerations include the extent to which it is feasible to implement the tool in clinical practice, e.g. the availability of data to calculate risks, transparency in its development and reporting, and for it to gain acceptability among clinicians.<sup>25</sup> Furthermore, any such tool needs linkage to interventions for outcomes to improve. One such intervention in relation to OxMIS is that it underscores safety planning as everyone receives a risk score. At the same time, potential harms need to be considered and OxMIS should be used to support rather than replace clinical decision-making, the latter which will necessarily consider individual, more proximal and contextual factors.

In conclusion, we have conducted a large external validation of one prediction model for suicide (OxMIS). In contrast to the majority of external validation studies in psychiatry, we have pre-specified predictors and outcomes, ensured adequate statistical power (with 1475 outcomes), published a research protocol, and transparently reported our findings, including presenting measures of both discrimination and calibration. We further reduced risks of potential biases by using nationwide registry data with no missing data for the predictors included in the model. With further research on feasibility and work considering how to link risk scores to interventions, OxMIS could assist mental health services in reducing suicide rates in people with SMI.

## Declarations

## Acknowledgements

The study was supported by the Academy of Finland (#308247, #294861). AS, SF, and AC are supported by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005). SF is further funded by the Wellcome Trust as part of a Senior Research Fellowship in Clinical Science (#202836/Z/16/Z). AC is supported by the NIHR Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006) and by the NIHR Oxford and Thames Valley Applied Research Collaboration. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health. PM has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101019329). TF receives funding from the National Institute for Health Research (NIHR) Community Healthcare MedTech and In Vitro Diagnostics Co-operative at Oxford Health NHS Foundation Trust (MIC-2016-018) and the NIHR Applied Research Collaboration Oxford and Thames Valley at Oxford Health NHS Foundation Trust. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

## Conflicts of interest

AC has received research, educational and consultancy fees from INCiPiT (Italian Network for Paediatric Trials), CARIPLO Foundation and Angelini Pharma; he is the CI/PI of two trials about seltorexant in depression, sponsored by Janssen. The remaining authors declare that they have no conflicts of interest. The funders were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

## References

1. Damen, J. A. A. G. *et al.* Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* **353**, i2416 (2016).
2. Meads, C., Ahmed, I. & Riley, R. D. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res. Treat.* **132**, 365–377 (2012).
3. Gray, E. P., Teare, M. D., Stevens, J. & Archer, R. Risk Prediction Models for Lung Cancer: A Systematic Review. *Clin. Lung Cancer* **17**, 95–106 (2016).
4. Usher-Smith, J. A., Walter, F. M., Emery, J. D., Win, A. K. & Griffin, S. J. Risk Prediction Models for Colorectal Cancer: A Systematic Review. *Cancer Prev. Res. (Phila. Pa.)* **9**, 13–26 (2016).
5. Meehan, A. J. *et al.* Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. *Mol. Psychiatry* **27**, 2700–2708 (2022).
6. Fazel, S. & Runeson, B. Suicide. *N. Engl. J. Med.* **382**, 266–274 (2020).
7. Quinlivan, L. *et al.* Scales for predicting risk following self-harm: an observational study in 32 hospitals in England. *BMJ Open* **4**, e004732 (2014).



8. Bjureberg, J. *et al.* Columbia-Suicide Severity Rating Scale Screen Version: initial screening for suicide risk in a psychiatric emergency department. *Psychol. Med.* 1–9 (2021)  
doi:10.1017/S0033291721000751.
9. Fazel, S., Wolf, A., Larsson, H., Mallet, S., & Fanshawe, T.R. Clinical prediction rules for persons diagnosed with an episode of severe mental illness for risk of violent crime and suicide – statistical analysis plan. (2019).
10. Gissler, M. & Haukka, J. Finnish health and social welfare registers in epidemiological research. *Nor. Epidemiol.* **14**, 113–120 (2004).
11. The Finnish Institute for Health and Welfare (THL). Care Register for Health Care.  
<https://thl.fi/en/web/thlfi-en/statistics-and-data/data-and-services/register-descriptions/care-register-for-health-care>.
12. Sund, R. Quality of the Finnish Hospital Discharge Register: A systematic review. *Scand. J. Public Health* **40**, 505–515 (2012).
13. Kieseppä, T., Partonen, T., Kaprio, J. & Lönqvist, J. Accuracy of register- and record-based bipolar I disorder diagnoses in Finland; a study of twins. *Acta Neuropsychiatr.* **12**, 106–109 (2000).
14. Documentation of statistics Causes of death - Statistics Finland.  
<https://www.stat.fi/en/statistics/documentation/ksyyt#Methods>.
15. Fazel, S., Wolf, A., Larsson, H., Mallett, S. & Fanshawe, T. R. The prediction of suicide in severe mental illness: development and validation of a clinical prediction rule (OxMIS). *Transl. Psychiatry* **9**, 98 (2019).
16. Neeleman, J. & Wessely, S. Changes in classification of suicide in England and Wales: time trends and associations with coroners' professional backgrounds. *Psychol. Med.* **27**, 467–72 (1997).
17. Ramakers, A., Aaltonen, M. & Martikainen, P. A closer look at labour market status and crime among a general population sample of young men and women. *Adv. Life Course Res.* **43**, 100322 (2020).
18. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
19. Austin, P. C. & Steyerberg, E. W. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat. Med.* **33**, 517–535 (2014).
20. Austin, P. C. & Steyerberg, E. W. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat. Med.* **38**, 4051–4065 (2019).
21. StataCorp. Stata Statistical Software: Release 17. (2021).
22. R Core Team. R: A language and environment for statistical computing. (2021).
23. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med.* **13**, 1–10 (2015).
24. Collins, G. S., Ogundimu, E. O. & Altman, D. G. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat. Med.* **35**, 214–226 (2016).

25. Senior, M. *et al.* Identifying Predictors of Suicide in Severe Mental Illness: A Feasibility Study of a Clinical Prediction Rule (Oxford Mental Illness and Suicide Tool or OxMIS). *Front. Psychiatry* **11**, 268 (2020).

Figures

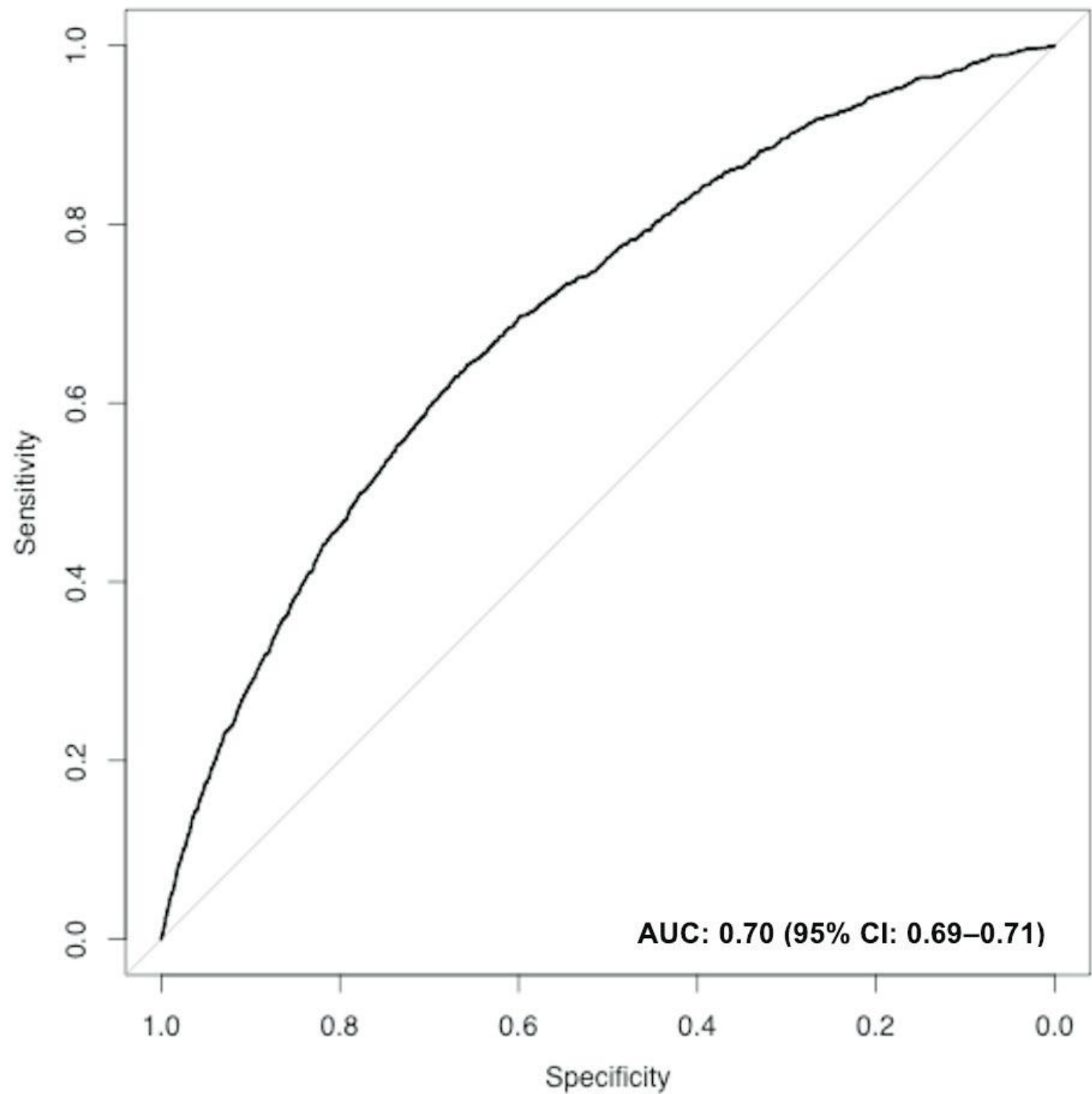
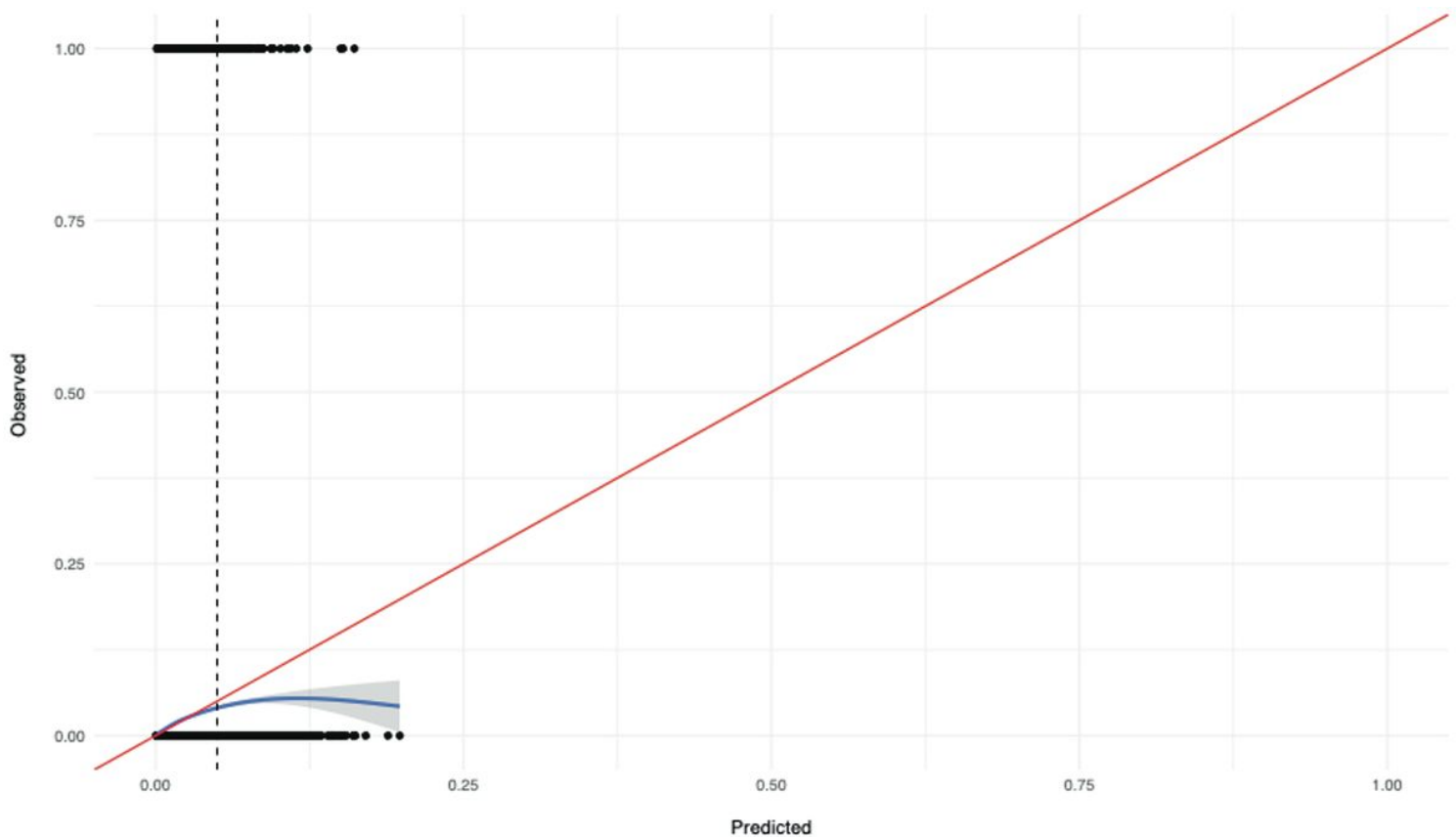


Figure 1

Suicide prediction model discrimination as demonstrated by a receiver operating characteristics curve in the Finnish external validation sample of individuals with severe mental illness (n=137 112)



**Figure 2**

Calibration plot comparing predicted and observed risks of suicide in the Finnish external validation sample of individuals with severe mental illness (n=137 112)

Note: The dashed vertical line denotes 5 percent predicted risk, which is the maximum cut-off value presently used in the OxMIS online calculator.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [eTables.docx](#)
- [oxmisprotocol.pdf](#)