Bharat Chrianna I (Orcid ID: 0000-0002-9778-0386)
McGrath John J (Orcid ID: 0000-0002-4792-6068)
Navarro-Mateu Fernando (Orcid ID: 0000-0002-3228-623X)
Degenhardt Louisa (Orcid ID: 0000-0002-8513-2218)

Development and evaluation of a risk algorithm predicting alcohol dependence after early onset of regular alcohol use

Chrianna Bharat, Meyer D. Glantz, Sergio Aguilar-Gaxiola, Jordi Alonso, Ronny Bruffaerts, Brendan Bunting, José Miguel Caldas-de-Almeida, Graça Cardoso, Stephanie Chardoul, Peter de Jonge, Oye Gureje, Josep Maria Haro, Meredith G. Harris, Elie G. Karam, Norito Kawakami, Andrzej Kiejna, Viviane Kovess-Masfety, Sing Lee, John J. McGrath, Jacek Moskalewicz, Fernando Navarro-Mateu, Charlene Rapsey, Nancy A. Sampson, Kate M. Scott, Hisateru Tachimori, Margreet ten Have, Gemma Vilagut, Bogdan Wojtyniak, Miguel Xavier, Ronald C. Kessler, Louisa Degenhardt.

Affiliations and addresses: see p2-3

Running head: Risk algorithm development for alcohol dependence onset

Word count: 4,223 (text); 296 (abstract)

Declaration of Interests: In the past three years, Dr Degenhardt has received investigator-initiated untied educational grants for studies of opioid medications in Australia from Indivior, Mundipharma and Seqirus. Dr. Harris reports consulting fees from RAND Corporation outside the submitted work. Dr. Kawakami reports grants and consulting fees outside the submitted work. He received grants from Fujitsu JAPAN, Ltd. and SBAtWork Corporation and consulting fees from Occupational Health Foundation, Japan Dental Association, Sekisui Chemicals, Junpukai Health Care Center and Osaka Chamber of Commerce and Industry. Dr. Kessler and Ms. Sampson report research grants from National Institute of Mental Health -U.S. (Grants: R01 MH070884; U01 MH60220); John D. and Catherine T. MacArthur Foundation; Pfizer Foundation; U.S. Public Health Service (Grants: R13-MH066849, R01-MH069864, and R01 DA016558); Fogarty International Center (Grant: R03-TW006481); Pan American Health Organization; Eli Lilly and Company; Ortho-McNeil Pharmaceutical; GlaxoSmithKline; Bristol-Myers Squibb; National Institute of Drug Abuse; Substance Abuse and Mental Health Services Administration -U.S.; Robert Wood Johnson Foundation (Grant 044708); and John W. Alden Trust. In the past 3 years, Dr. Kessler was a consultant for Datastat, Inc., Holmusk, RallyPoint Networks, Inc., and Sage Therapeutics. He has stock options in Mirah, PYM, and Roga Sciences. Dr. Tachimori reports research grants from the Ministry of Health, Labour and Welfare, Japan. All other authors report no conflicts of interest.

***Corresponding author**:
Chrianna Bharat
c.bharat@student.unsw.edu.au
National Drug and Alcohol Research Centre
UNSW Sydney, Sydney, Australia

Affiliations

National Drug and Alcohol Research Centre (NDARC), University of New South Wales Australia, Sydney, NSW, Australia. (Chrianna Bharat, BSc; Louisa Degenhardt, PhD).

Department of Epidemiology, Services, and Prevention Research (DESPR), National Institute on Drug Abuse (NIDA), National Institute of Health (NIH), Bethesda, Maryland, USA (Meyer D. Glantz, PhD).

Center for Reducing Health Disparities, UC Davis Health System, Sacramento, California, USA. (Sergio Aguilar-Gaxiola, MD, PhD).

Health Services Research Unit, IMIM-Hospital del Mar Medical Research Institute, Barcelona, Spain. (Jordi Alonso,MD, PhD; Gemma Vilagut, PhD).

Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. (Jordi Alonso; Gemma Vilagut; Fernando Navarro-Mateu, MD, PhD).

Department of Life and Health Sciences, Pompeu Fabra University (UPF), Barcelona, Spain. (Jordi Alonso).

Universitair Psychiatrisch Centrum - Katholieke Universiteit Leuven (UPC-KUL), Campus Gasthuisberg, Leuven, Belgium. (Ronny Bruffaerts, PhD).

School of Psychology, Ulster University, Londonderry, United Kingdom. (Brendan Bunting, PhD).

Lisbon Institute of Global Mental Health and Chronic Diseases Research Center (CEDOC), NOVA Medical School | Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisbon, Portugal. (José Miguel Caldas-de-Almeida, MD, PhD; Graça Cardoso, MD, PhD; Miguel Xavier, MD, PhD).

Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA. (Stephanie Chardoul, BA).

Department of Developmental Psychology, University of Groningen, Groningen, The Netherlands. (Peter de Jonge, PhD).

Department of Psychiatry, University College Hospital, Ibadan, Nigeria. (Oye Gureje, MD, DSc).

Research, Teaching and Innovation Unit, Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Barcelona, Spain. Centre for Biomedical Research on Mental Health (CIBERSAM), Madrid, Spain. (Josep Maria Haro, MD, PhD).

School of Public Health, The University of Queensland, Herston, QLD, Australia. (Meredith G. Harris, PhD).
Queensland Centre for Mental Health Research, The Park Centre for Mental Health, QLD 4072, Australia. (Meredith G. Harris; John J. McGrath, MD, PhD).

Department of Psychiatry and Clinical Psychology, St George Hospital University Medical Center, Balamand University, Faculty of Medicine, Beirut, Lebanon; Institute for Development, Research, Advocacy and Applied Care (IDRAAC), Beirut, Lebanon. (Elie G. Karam, MD).

Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. (Norito Kawakami, MD, DMSc).

Institute of Psychology, University of Lower Silesia, Wroclaw, Poland. (Andrzej Kiejna, MD, PhD).

Ecole des Hautes Etudes en Santé Publique (EHESP), EA 4057, Paris Descartes University, Paris, France. (Viviane Kovess-Masfety, MD, PhD).

Department of Psychiatry, Chinese University of Hong Kong, Tai Po, Hong Kong. (Sing Lee, MB,BS).

Queensland Brain Institute, The University of Queensland, St Lucia, QLD, Australia. National Centre for Register-based Research, Aarhus University, Aarhus V 8000 Denmark. (John J. McGrath).

Institute of Psychiatry and Neurology, Warsaw, Poland. (Jacek Moskalewicz, PhD).

Department of Basic Psychology and Methodology, University of Murcia, Murcia, Spain; Murcia Biomedical Research Institute (IMIB-Arrixaca), Murcia, Spain; Unidad de Docencia, Investigación y Formación en Salud Mental, Servicio Murciano de Salud, Murcia, Spain. (Fernando Navarro-Mateu).

Department of Psychological Medicine, University of Otago, Dunedin, Otago, New Zealand. (Charlene Rapsey, PhD; Kate M. Scott, PhD).

Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA. (Nancy A. Sampson, BA; Ronald C. Kessler, PhD).

Endowed Course for Health System Innovation, Keio University School of Medicine, Tokyo, Japan. Department of Clinical Data Science, Clinical Research & Education Promotion Division, National Center of Neurology and Psychiatry, Tokyo, Japan. (Hisateru Tachimori, PhD).

Trimbos-Instituut, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands. (Margreet ten Have, PhD).

Centre of Monitoring and Analyses of Population Health, National Institute of Public Health-National Research Institute, Warsaw, Poland. (Bogdan Wojtyniak, ScD).

Abstract

Aim: Likelihood of alcohol dependence (AD) is increased among people who transition to greater levels of alcohol involvement at a younger age. Indicated interventions delivered early may be effective in reducing risk but could be costly. One way to increase cost-effectiveness would be to develop a prediction model that targeted interventions to the subset of youth with early alcohol use who are at highest risk of subsequent AD.

Design: A prediction model was developed for DSM-IV AD onset by age 25 using an ensemble machine learning algorithm known as super learner. Shapley additive explanations (SHAP) assessed variable importance.

Setting and participants: Respondents reporting early onset of regular alcohol use (i.e., by 17 years of age) who were aged 25 years or older at interview from 14 representative community surveys conducted in 13 countries as part of WHO's World Mental Health Surveys.

Measurements: The primary outcome to be predicted was onset of lifetime DSM-IV AD by age 25 as measured using the Composite International Diagnostic Interview, a fully structured diagnostic interview

Findings: AD prevalence by age 25 was 5.1% across the 10,687 individuals who reported drinking alcohol regularly by age 17. The prediction model achieved an external area under the curve (0.78; 95% confidence interval [CI] 0.74-0.81) higher than any individual candidate risk model (0.73-0.77) and an area under the precision-recall curve of 0.22. Overall calibration was good (ICI, 1.05%), however, miscalibration was observed at the extreme ends of the distribution of predicted probabilities. Interventions provided to the 20% of people with highest risk would identify 49% of AD cases and require treating four people without AD to reach one with AD. Important predictors of increased risk included younger onset of alcohol use, males, higher cohort alcohol use and more mental disorders.

Conclusion: A risk algorithm can be created using data collected at the onset of regular alcohol use to target youth at highest risk of alcohol dependence by early adulthood. Important considerations remain for advancing the development and practical implementation of such models.

Keywords: alcohol dependence; childhood; predictive modelling; machine learning;

INTRODUCTION

Alcohol use is a leading global risk factor for population health, estimated to be responsible for 3.7% of the total burden of disease and injuries worldwide. The interval of risk for starting alcohol use begins typically during adolescence and often progresses from experimentation to regular use[1-3]. Early use is known to be a strong predictor of later progression into alcohol use disorder[4,5]. Therefore, the World Health Organisation recommends that screening and brief interventions be part of routine health care for adolescents[6].

Screening involves the use of brief, easy-to-administer tools that can help to determine an individual's level of involvement with alcohol, detect risky and harmful drinking patterns and assist clinicians to effectively monitor an individual's progress over time[7]. Item responses are scored and compared to a scale which conveys a certain risk level to guide the clinician's response[7,8]. Good evidence is available regarding the validity of certain instruments when used with young people[9,10], however, these risk evaluations are often indicative of current or short-term, rather than long-term, outcomes.

There has been recent interest in the development of risk algorithms to predict, among youth, substance use disorder onset over time periods extending into adulthood[11,12]. Results indicate that substance use disorder onset by age 30 can be predicted with fair to good accuracy using sociodemographic, clinical, and environmental information obtained prior to the age of 18. While promising, the generalisability of these findings is unclear, given models to date have been developed/applied in unrepresentative samples or their performance evaluated using the same data used in model development[13]. Further, as these models predict the outcome of any substance use disorder diagnosis rather than for a specific substance, the indicative risk for specific substance(s) are unknown, and this may limit the usefulness of the results in guiding interventions. For these reasons, it remains unclear the extent to which alcohol dependence (AD) prognoses over the long-term can be evaluated among youth.

Using data collected retrospectively from cross-sectional, general population surveys in 13 high-income countries, this study aimed to develop and internally validate preliminary prognostic models to predict AD onset by early adulthood among youth with an early onset of regular alcohol use using sociodemographic, mental health, and contextual/environmental variables. If good performance measures are obtained, the algorithm could subsequently be validated in other settings, and the results used to inform data collection in future prospective trials. In addition, the use of the risk-evaluation strategies could be used to support the recruitment of target groups for studies evaluating brief interventions and treatment.

METHOD

Study design and setting

Data from 14 surveys (13 countries) in the World Mental Health (WMH) Survey Initiative which were defined as high-income at the time of data collection were used in developing and internally validating the prognostic models. The WMH Survey Initiative is a series of community epidemiological household surveys which include retrospective assessments of alcohol use and use disorders among other mental and substance use disorders. All surveys were based on national household samples, the exceptions being a survey of urban areas in Argentina, a specific region in Spain (Murcia region) and a series of metropolitan areas in Japan. Surveys were carried out between 2001-2015 (depending on survey, see **Table 1**) in respondents' homes face-to-face with trained lay interviewers. Informed consent was obtained prior to all interviews using procedures approved by local ethics committees.

Table 1 around here

Outcome measure

The primary outcome to be predicted was onset of lifetime DSM-IV AD by age 25 as measured using the Composite International Diagnostic Interview[14], a fully structured diagnostic interview. The cut-off of 25 years was selected because it captured the interval of time with the highest acceleration of onset of AD relative to other periods. AD onset was evaluated as the age the respondent reported first experiencing at least three disorder-related symptoms in the same year. The outcome measure was a binary variable representing whether AD criteria were met between the onset of regular alcohol use and 25 years of age.

Sample

The master sample was defined as all individuals who, at the time of interview, were 25 years of age or older and retrospectively reported onset of regular use by 17 years of age. The cut-off of 17 years was selected because it indicated an early and accelerated involvement with alcohol, was associated with an increased risk of AD relative to a later onset of regular use and fell below the legal drinking age in most included countries (see **Appendix 1**). The analysis sample was drawn from this master sample in a variation of the case-control approach.

The cases were individuals in this master sample who reported an AD onset by age 25. Cases were separated based on serpentine sampling[15,16] of person-level weights into a training set of 70% and a holdout set with the remaining 30%. Cases in the training sample were further subdivided into ten training folds based on serpentine sampling of person-level weights[15,16].

The controls were a stratified probability sample of other individuals in the master sample. Controls were sampled into a training set with 70% of controls and a holdout set with the

remaining 30% based on a serpentine sort of person-level weights with implicit stratification by sex, age, and survey. Controls in the training sample were further divided into ten training folds based on serpentine sampling of person-level weights. Sampling fractions within folds were set to generate a sample of controls approximately five times the number of cases with probability of selection equal to person-level weight.

Cases and controls were weighted by the inverse of their probabilities of selection for purposes of analysis and population projection. The 70% of cases combined with the subsampled controls created the training sample from which the prediction model was developed. The remaining 30% of cases and controls were held out to validate the model. For model development, weights for cases in the training sample were multiplied by five times the average weight of cases in the same fold. Controls in the training folds were assigned a weight of one. This created a balanced weighted case-control ratio within each fold for model development.

Predictor Classes

All survey questions that could be used to operationalise risk factors identified in the literature review and which were collected in all surveys were identified. A total of fifty-nine predictors (including 14 survey indicators) were evaluated as at the age the respondent reported onset of regular alcohol use across four categories of predictor classes.

Sociodemographics

There were two variables, including a dichotomous indicator for sex and one continuous variable representing age at prediction (i.e., age of onset of regular alcohol use (range: 6-17)).

Pattern of alcohol use

There were four variables relating to history of alcohol use. Continuous variables were defined to represent the age of onset of first alcohol use (range: 6-17) and the number of years from onset of alcohol use to regular use (range: 0-12). The cumulative lifetime prevalence of regular alcohol use among 25–35-year-olds in the respondent's country as of the calendar year when the respondent reported their own onset of regular alcohol use was evaluated, and categorised into units of width five (range: 15-20% to 90-95%, or 16 levels).

History of prior mental disorders

There were six history of prior mental disorder variables. These included dichotomous indicators of generalised anxiety disorders, panic disorder, major depression, and broad bipolar disorder. An ordered categorical variable representing the highest level of drug use (no use, use with no abuse, abuse with no dependence, dependence) was created and

standardised. A continuous measure for the total number of prior mental disorders was also created (range: 0-5).

Traumatic experiences

There were 34 traumatic experience variables. These included dichotomous indicators of 26 different traumatic experiences[17]. Continuous measures of the total number of any traumatic experience types (range: 0-14; entered as both standardised and stabilised variables), as well as the number from each trauma domain were evaluated. Trauma domains included exposure to organised violence (range: 0-3), participated in organised violence (range: 0-5), interpersonal violence (range: 0-4), sexual relationship violence (range: 0-6), other life-threatening trauma (range: 0-6), and secondary trauma (range: 0-2) (see **Appendix 2**).

Missing age of onset values were imputed using regression-based imputation controlling for key sociodemographic and clinical variables at the smallest level possible (within-country otherwise within the high-income group). A skip error in a subset of earlier surveys meant symptoms of AD were only assessed among respondents with a history of alcohol abuse. Missing values were imputed using regression-based imputation models developed and validated on data from later surveys without the skip error (among which, 13.5% of AD cases had not met criteria for alcohol abuse). Validation results suggest the imputation models produce consistent prevalence estimates, strong individual-level classification accuracy metrics and similar distributions of important correlates; full details are described elsewhere[18].

Statistical analysis

Model Selection

To predict AD, we used the Super Learner (SL) ensemble method to combine predicted probabilities of AD across a large number of different machine learning algorithms (the "library")[19]. The library included a generalised linear model with a logistic link function, a series of penalised generalized linear models with different mixing models parameters, random forests, neural networks, a series of gradient boosted decision trees, Bayesian adaptive regression trees, and a series of adaptive regression splines (see **Appendix 3** for details).

A number of the algorithms in our SL library require hyperparameter tuning for optimal performance. We addressed this in simple cases by including a series of models for a single algorithm with different hyperparameter values in the SL ensemble (e.g., ten penalised regression classifiers that differed in values of the mixing parameter, several adaptive regression splines that differed in maximum degrees). In more complex cases (random

forests, gradient boosted decision trees), we used cross-validated penalized generalized linear models to select optimal combinations of hyperparameters.

Ten-fold cross-validation was used for internal SL cross-validation both to build optimal models with each classifier and to determine the optimal cross-validated weight for each classifier in the ensemble. Model performance was assessed in terms of discrimination (ability to distinguish between cases and controls) and calibration (agreement between observed and predicted probabilities) in the holdout sample.

Discrimination

Discrimination was estimated by means of generating the receiver operating characteristic (ROC) and precision-recall (PR) curves. Both ROC and PR curves visualise trade-offs in different aspects of performance as the threshold applied to a model's prediction varies. A ROC curve shows the sensitivity (the proportion of cases that were above a given prediction threshold) as a function of the false positive rate (the proportion of controls that were above the same prediction threshold), with the area under the ROC curve (AUC) providing a measure of discriminative ability. In comparison, the PR curve shows the positive predictive value (PPV; the probability of AD among respondents above a given prediction threshold) as a function of the sensitivity, with the area under the PR curve called the average precision. The PR curve is of interest because it does not incorporate correctly predicted controls and is therefore less prone to exaggerate model performance for imbalanced datasets[20]. While the baseline for ROC curves is fixed at 0.5, the baseline of a PR curve is determined by the ratio of the minority class (case) and majority class (controls)[20]. Because the PR curve is prevalence dependent, the ROC curve should be used when comparing model performance between datasets with different event prevalence.

Calibration

Calibration was estimated by means of calibration plots and the median (E_{50}), 90th percentile (E_{90}), maximum (E_{max})[21], and weighted mean (integrated calibration index; ICI) of the absolute difference between observed and predicted probabilities[22]. The ICI provides a measure of central tendency for summarising the absolute differences, with weights determined by the empirical distribution of predicted probabilities.

Operating characteristics, including sensitivity, specificity, and PPV, were calculated for a variety of thresholds. The 5%, 10%, 15%, and 20% of observations with highest predicted probabilities of AD were defined by rank ordering predicted probabilities from the SL in the training sample and using the cut points to create risk tiers of predicted probabilities in the holdout sample.

Variable Importance

Predictor variables of greatest overall importance were investigated using the model-agnostic shapley additive explanations (SHAP) method[23], providing a summary measure of contribution to the model prediction from each variable.

Fairness

Predictive models can be subject to biases stemming from the original data or through model development which, when applied, may not provide equal benefit to every population subgroup[24]. To evaluate fairness, Poisson regression analyses were used to determine whether the association of observed AD with predicted probabilities were independent of key covariates. Main effects included the key covariate (evaluated at the time of regular alcohol use onset), the SL predicted probability and the interaction of these two variables. Key covariates were selected to identify groups at risk of marginalisation for which data was available, and included gender, age (pre-teen: <13 years of age v. teenager: 13-17 years) and predominant country culture (western v. non-western).

Sensitivity Analysis

To examine the impact of including other potential predictors of SUD onset on the findings, a sensitivity analysis was conducted which included all variables in the main analysis as well as variables capturing childhood adversities, social phobia, and specific phobia disorder. Surveys which did not assess these additional variables, including Australia, New Zealand, and Israel, were excluded from the sensitivity analysis. For comparison, the original model was rerun with the original set of predictors and excluding the subset of surveys which did not assess the additional predictors.

All analyses were conducted in R[25] using the SuperLearner[26], glmnet[27], xgboost[28], ranger[29], nnet[30], and earth[31] packages. The analysis was not pre-registered and results should be considered exploratory.

RESULTS

A total of 41,787 respondents participated in the two-part survey and completed the alcohol use sections. Among the 10,687 individuals who were 25 years or older at time of interview and reported drinking alcohol regularly by 17 years of age, 791 (5.1% [SE 0.2]) reported an onset of AD by age 25. Age at onset of regular alcohol use in the cohort ranged from 6 to 17 years, with a median of 15.5 (SE 0.1).

The training sample, containing 553 cases (i.e., individuals with AD onset by age 25) and 2,765 sampled controls (i.e., other individuals), was used to develop the prediction model. The remaining 30% of cases (n=238) and 2,970 other sampled controls were held out to validate the model.

Library Weighting

The cross-validated weight that defines the relative importance of each classifier in the SL ensemble is shown in **Table 2**. Extreme gradient boosting was the best classifier for predicting AD onset by age 25 (β=0.464). Other weighed classifiers included a generalised linear model (β=0.332), neural network (β=0.107), multivariate adaptive polynomial spline regression (β=0.062), and multivariate adaptive regression splines (1 degree, β=0.027; 4 degrees, β=0.008).

Discrimination

The AUC of the SL model when applied to the holdout sample was 0.78 (95% CI 0.74-0.81) (**see Table 2 & Figure 1**). This was higher than the AUCs of the six library algorithms weighted by the SL which were in the range of 0.73-0.77. However, a comparison of the performance of all candidate algorithms in the ensemble found four classifiers assigned weights of zero in the fitted SL model outperformed the SL model in the holdout sample. AUC was calculated as 0.78 (95% CI 0.74-0.82) for two random forest models, 0.78 (95% CI 0.74-0.82) for a differently parameterised multivariate adaptive regression spline fit (compared to the multivariate adaptive regression spline assigned a non-zero weight in the SL model), and 0.78 (95% CI 0.74-0.82) for a different extreme gradient boosted model (see **Appendix 4**). The average precision for the SL was 0.220 which, when compared to the baseline precision of 0.052 (determined by 5.2% prevalence in the original sample), indicates a significant improvement in classification over chance (see **Figure 1**).

Table 2 & Figure 1 around here

Calibration

Calibration was satisfactory for the SL model when applied to the holdout sample with an ICI of 1.05% and E_{50} of 0.55%. E_{90} and E_{max} values for the SL model were 3.02% and 18.62%,

respectively, which indicates some variation for calibration-in-the-large. The calibration plot reveals this miscalibration occurred at the extreme ends of the predicted probability distributions, with high risks underestimated and low risks overestimated (see **Figure 2**).

Figure 2 around here

Operating Characteristics

Inspection of the ROC curves for the SL model and the weighted classifiers in the holdout sample showed that the slope was steepest for 1-specificity in the range 0-0.05, which corresponds roughly to the 5% of respondents with highest predicted AD risk in the models (**see Table 3**). Visual inspection of the PR curve also shows precision of the model was highest over this interval. Another inflection point in the slope was at the 20th percentile. The sensitivity at these two thresholds show that these respondents accounted for 15.2% and 49.0% of all cases of probable AD, respectively. This means that an intervention delivered only to the 5% or 20% of respondents with highest AD risk would capture 15.2% or 49.0%, respectively, of people who would otherwise develop AD by 25 years of age.

The highest PPV for the SL model is 40.0% for the 5% threshold, indicating this is the proportion of respondents above that threshold who would be expected to have an AD onset by age 25 in the absence of any interventions. PPV decreases to 19.2% at the most liberal threshold considered (20%). That is, we would expect one in five respondents above the 20% threshold to have an AD onset by 25 years of age. Specificity in the top 5% and 20% were 98.7% and 88.6%, respectively.

Table 3 around here

Predictor Importance

The nine top predictors included four indicators of individual or cohort alcohol involvement, two relating to traumatic experiences, and one each for the number of mental disorders, level of drug use involvement, and gender (see **Figure 3**). Specifically, increased risk was associated with younger onset of both first and regular alcohol use, males, higher cohort alcohol use, fewer years from first to regular alcohol use, more secondary traumas, more mental disorders, higher levels of drug use and having experienced the unexpected death of a loved one.

Figure 3 around here

Fairness

Prevalence of AD by age 25 following early onset of regular alcohol use was higher for males (5.7% v. females, 4.2%), pre-teens (<13 years, 6.6% v. teenagers, 4.6%) and respondents in western countries (5.5% v. non-western countries 0.9%). However, there was no substantial difference in the association between predicted probabilities from the SL model and the observed event rates for these variables (rate ratios ranged from 0.44-1.2; see **Appendix 5**).

Sensitivity analysis

Results of the sensitivity analyses are summarised in **Appendix 6**. Compared to the main analysis including all high-income countries, both the subset analysis (**Appendix 6.1**) and the sensitivity analysis (**Appendix 6.2**) generated better measures of discrimination and sensitivity, similar measures of specificity and PPV, and worse calibration. Top ranked predictors similar to both the main and sensitivity analyses included age of onset of drinking, age of onset of regular drinking, prevalence of regular use among persons aged 25-35 years of age in the same country, gender, and drug involvement. For each variable added to the sensitivity analysis, including childhood adversities, social phobia and specific phobia, the relative contribution to the model was less than 1%.

DISCUSSION

Reflecting the need for innovative approaches to improve the delivery of alcohol prevention programmes[32], this study developed a risk model to identify, among youth who are drinking alcohol regularly, those at highest risk of developing alcohol dependence by early adulthood. In addition to evaluating the predictive performance of a preliminary prognostic algorithm for this purpose, our findings illustrate some of the standard processes and considerations in developing prognostic models as well as indicate potential approaches to advancing their development and practical implementation.

First, the prevalence of AD by age 25 among the subgroup with the highest quintile of risk assigned by the algorithm was 19.2%; this is approximately three times as high as in the sample of people with an early onset of regular alcohol use (5.1%), and eight times as high as in the full sample of survey respondents (2.4%). These results show promise that a fairly useful risk algorithm could be used to identify a more precise subgroup of individuals at highest risk, providing a tool which could assist health professionals in implementing timely, indicated interventions, especially when endeavouring to maximise cost-effectiveness and resource allocation. However, these findings must be considered alongside the observed miscalibration at the extreme ends of the risk distribution, and the relatively low measures of sensitivity and PPV. If the algorithm presented in this study was implemented in practice, these findings imply that the model would misclassify some AD cases and that most people identified as being at risk would not develop AD by age 25. The implications and consequences of inaccurate identification from a predictive algorithm will vary depending on the context; even where an indicated intervention is warranted, stigma of the risk-label may contribute to a range of negative psychosocial factors[33,34]. Procedures for averting such risks need to be carefully considered prior to model implementation and in the design of potential interventions.

Second, investigating each predictor in the algorithm and disaggregating its contribution to the risk profiles helps to understand which factors provide the greatest signal for the outcome of interest. The most important predictors for developing AD were related to the individuals' and cohort's use of alcohol, traumatic experiences, comorbid mental disorders, other drug use, and gender. Although these features are known risk factors of problematic drinking during adulthood, these results provide confidence in terms of the contextual validity of the model based on existing domain knowledge. For outcomes or settings where there is a less established understanding of the mechanisms impacting the outcome of interest, evaluations of variable importance may lead to the discovery of new predictors. More generally, variable importance measures may overcome barriers to transparency by providing "human-intelligible explanations" of decisions formed by otherwise uninterpretable "black-box" algorithms[35].

Third, while the algorithm in the main analysis showed fairly good internal and external validation performance, the sensitivity analyses incorporating additional information about childhood adversities and phobia-related diagnoses resulted in improved discrimination but worse calibration. However, this improvement in discrimination was shown to result primarily from the exclusion of select surveys, rather than from the inclusion of additional predictors. These findings suggest there may be considerable variability in the extent to which AD can be predicted across geographic sites, even when accounting for site and/or geographic variation in the predictive algorithm. The availability of an expansive list of potentially important variables might help to identify other important predictors across settings, and potentially improve model accuracy. Ideally, subsequent studies developing predictive models would focus on conducting external validation of algorithms and endeavour to identify, from a wide range of predictors, the smallest number from which optimal accuracy can be achieved. This would advance the development and potential practical implementation of future prognostic models globally.

The current findings should be interpreted within the context of several limitations which may provide opportunities for subsequent studies. First, AD and other mental disorders were evaluated based on retrospective reporting, which may be subject to recall bias. These measures, however, were collected using well-validated instruments and have shown high diagnostic concordance with clinical diagnoses[36]. Survival bias may also contribute to a downward trend in predictive performance. Second, as data come from 13 high-income countries there is not full representation of all regions, income levels, and other country characteristics. There was variation in response rates, the years in which the surveys were conducted, and cross-national differences in the legal age of drinking and potential willingness to disclose personal information about alcohol use and other mental health and life experiences, which may be found in external validation studies to either positively or negatively impact model generalisability. Third, our sensitivity test considered an expanded list of variables in model development but other variables that the literature suggests would be useful in predicting AD, such as behavioural disorders and family history of alcohol use disorder[37], were not collected in all surveys, so could not be include in the analyses.

Finally, this study aimed to determine the potential of a risk algorithm to identify a sufficiently high concentration of at-risk individuals to support the design and implementation of indicated interventions. This has been achieved by showing that nearly half of all AD cases occur among the 20% of patients classified by our model as having highest risk. Whether this threshold is optimal for implementing indicated interventions for AD, or if any intervention would be cost-effective at any threshold is beyond the scope of this report. Given a much more extensive set of predictor variables was available to build this model than exists in most routinely collected datasets, these findings provide evidence to justify a more focused effort

to collect expanded information during routine health screening. If youth were prospectively followed up to age 25 to observe which of them develop AD, longitudinal analysis could be used to see how well the predictors collected at onset of regular alcohol use help to predict which ones will be at risk of AD. This would advance the agenda of creating broadly useful AD risk algorithms to target interventions based on the presenting profiles of people with an early onset of regular alcohol drinking, as well as subsequent algorithms to predict other types of substance use disorders and mental disorders.

Acknowledgements

Collaborators

The WHO World Mental Health Survey collaborators are Ali Al-Hamzawi, MD; Yasmin A. Altwaijri, PhD; Laura Helena Andrade, MD, PhD; Lukoye Atwoli, MD, PhD; Corina Benjet, PhD; Guilherme Borges, ScD; Evelyn J. Bromet, PhD; Somnath Chatterji, MD; Alfredo H. Cia, MD; Koen Demyttenaere, MD, PhD; Silvia Florescu, MD, PhD; Giovanni de Girolamo, MD; Oye Gureje, MD, DSc, FRCPsych; Hristo Hinkov, MD, PhD; Chi-yi Hu, MD, PhD; Aimee Nasser Karam, PhD; Elie G. Karam, MD; Georges Karam, MD; Jean-Pierre Lepine, MD; Maria Elena Medina-Mora, PhD; Marina Piazza, MPH, ScD; Jose Posada-Villa, MD; Tim Slade, PhD; Juan Carlos Stagnaro, MD, PhD; Dan J. Stein, FRCPC, PhD; Yolanda Torres, MPH, Dra.HC; Maria Carmen Viana, MD, PhD; Daniel V. Vigo, MD, DrPH; Harvey Whiteford, MBBS, PhD; David R. Williams, MPH, PhD;

The authors and collaborators thank the staff of the WMH Data Collection and Data Analysis Coordination Centres for assistance with instrumentation, fieldwork, and consultation on data analysis.

Funding

This work was supported by an Australian National Health and Medical Research Council (NHMRC) project grant (no. 1081984). C.B. is supported by a UNSW Scientia PhD scholarship and a National Drug and Alcohol Research Centre (NDARC) scholarship. L.D. is supported by an NHMRC Senior Principal Research Fellowship (1135991) and a US a National Institute of Health (NIH) National Institute on Drug Abuse (NIDA) grant (R01DA1104470). NDARC, UNSW Sydney, is supported by funding from the Australian Government Department of Health under the Drug and Alcohol Program.

The World Health Organization World Mental Health (WMH) Survey Initiative is supported by the United States National Institute of Mental Health (NIMH; R01 MH070884), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, the United States Public Health Service (R13-MH066849, R01-MH069864, and R01 DA016558), the Fogarty International Center (FIRCA R03-TW006481), the Pan American Health Organization, Eli Lilly and Company, Ortho-McNeil Pharmaceutical Inc., GlaxoSmithKline, and Bristol-Myers Squibb.

The 2007 Australian National Survey of Mental Health and Wellbeing is funded by the Australian Government Department of Health and Ageing.

The Argentina survey − Estudio Argentino de Epidemiología en Salud Mental (EASM) − was supported by a grant from the Argentinian Ministry of Health (Ministerio de Salud de la Nación) − (Grant Number 2002-17270/13-5).

The ESEMeD project is funded by the European Commission (Contracts QLG5-1999-01042; SANCO 2004123, and EAHC 20081308), the Piedmont Region (Italy)), Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Spain (FIS 00/0028), Ministerio de Ciencia y Tecnología, Spain (SAF 2000-158-CE), Generalitat de Catalunya (2017 SGR 452; 2014 SGR 748), Instituto de Salud Carlos III (CIBER

CB06/02/0046, RETICS RD06/0011 REM-TAP), and other local agencies and by an unrestricted educational grant from GlaxoSmithKline.

The World Mental Health Japan (WMHJ) Survey is supported by the Grant for Research on Psychiatric and Neurological Diseases and Mental Health (H13-SHOGAI-023, H14-TOKUBETSU-026, H16-KOKORO-013, H25-SEISHIN-IPPAN-006) from the Japan Ministry of Health, Labour and Welfare.

Te Rau Hinengaro: The New Zealand Mental Health Survey (NZMHS) is supported by the New Zealand Ministry of Health, Alcohol Advisory Council, and the Health Research Council.

The Northern Ireland Study of Mental Health was funded by the Health & Social Care Research & Development Division of the Public Health Agency.

The Polish project Epidemiology of Mental Health and Access to Care – EZOP Project (PL 0256) was carried out by the Institute of Psychiatry and Neurology in Warsaw in consortium with Department of Psychiatry - Medical University in Wroclaw and National Institute of Public Health-National Institute of Hygiene in Warsaw and in partnership with Psykiatrist Institut Vinderen – Universitet, Oslo. The project was funded by the European Economic Area Financial Mechanism and the Norwegian Financial Mechanism. EZOP project was co-financed by the Polish Ministry of Health.

The Portuguese Mental Health Study was carried out by the Department of Mental Health, Faculty of Medical Sciences, NOVA University of Lisbon, with collaboration of the Portuguese Catholic University, and was funded by Champalimaud Foundation, Gulbenkian Foundation, Foundation for Science and Technology (FCT) and Ministry of Health.

The Psychiatric Enquiry to General Population in Southeast Spain – Murcia (PEGASUS-Murcia) Project has been financed by the Regional Health Authorities of Murcia (Servicio Murciano de Salud and Consejería de Sanidad y Política Social) and Fundación para la Formación e Investigación Sanitarias (FFIS) of Murcia.

The US National Comorbidity Survey Replication (NCS-R) is supported by the National Institute of Mental Health (NIMH; U01-MH60220) with supplemental support from the National Institute of Drug Abuse (NIDA), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Robert Wood Johnson Foundation (RWJF; Grant 044708), and the John W. Alden Trust.

A complete list of all within-country and cross-national WMH publications can be found at http://www.hcp.med.harvard.edu/wmh/.

Data Sharing Statement

Access to the cross-national World Mental Health (WMH) data is governed by the organizations funding and responsible for survey data collection in each country. These organizations made data available to the WMH consortium through restricted data sharing agreements that do not allow us to release the data to third parties. The exception is that the U.S. data are available for secondary analysis via the Inter-University Consortium for Political and Social Research (ICPSR), http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/00527.

Role of the Funding Source

None of the funders had any role in the design, analysis, interpretation of results, or preparation of this paper. The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of the World Health Organization, other sponsoring organizations, agencies, or governments.

REFERENCES

1. Degenhardt L, Chiu W-T, Sampson N, et al. Toward a Global View of Alcohol, Tobacco, Cannabis, and Cocaine Use: Findings from the WHO World Mental Health Surveys. *PLOS Medicine* 2008; **5**(7): e141.
2. Masten AS, Faden VB, Zucker RA, Spear LP. Underage drinking: A developmental framework. *Pediatrics* 2008; **121**(Supplement 4): S235-S51.
3. Glantz MD, Bharat C, Degenhardt L, et al. The epidemiology of alcohol use disorders cross-nationally: Findings from the World Mental Health Surveys. *Addictive behaviors* 2020; **102**: 106128.
4. Dawson DA, Goldstein RB, Chou SP, Ruan WJ, Grant BF. Age at first drink and the first incidence of adult-onset DSM-IV alcohol use disorders. *Alcoholism, Clinical and Experimental Research* 2008; **32**(12): 2149-60.
5. Warner LA, White HR, Johnson V. Alcohol initiation experiences and family history of alcoholism as predictors of problem-drinking trajectories. *Journal of studies on alcohol and drugs* 2007; **68**(1): 56-65.
6. World Health Organization. Orientation programme on adolescent health for health care providers. 2006.
7. Harris SK, Louis-Jacques J, Knight JR. Screening and brief intervention for alcohol and other abuse. *Adolescent Medicine: State of the Art Reviews* 2014; **25**(1): 126-56.
8. Pilowsky DJ, Wu L-T. Screening instruments for substance use and brief interventions targeting adolescents in primary care: A literature review. *Addictive Behaviors* 2013; **38**(5): 2146-53.
9. Toner P, Böhnke JR, Andersen P, Mccambridge J. Alcohol screening and assessment measures for young people: A systematic review and meta-analysis of validation studies. *Drug and Alcohol Dependence* 2019; **202**: 39-49.
10. Clark DB, Gordon AJ, Ettaro LR, Owens JM, Moss HB. Screening and Brief Intervention for Underage Drinkers. *Mayo Clinic Proceedings* 2010; **85**(4): 380-91.
11. Meier MH, Hall W, Caspi A, et al. Which adolescents develop persistent substance dependence in adulthood? Using population-representative longitudinal data to inform universal risk assessment. *Psychological medicine* 2016; **46**(4): 877-89.
12. Jing Y, Hu Z, Fan P, et al. Analysis of substance use and its outcomes by machine learning I. Childhood evaluation of liability to substance use disorder. *Drug and alcohol dependence* 2020; **206**: 107605.
13. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European Heart Journal* 2014; **35**(29): 1925-31.
14. Kessler RC, Üstün TB. The world mental health (WMH) survey initiative version of the world health organization (WHO) composite international diagnostic interview (CIDI). *International journal of methods in psychiatric research* 2004; **13**(2): 93-121.
15. Chromy JR. Sequential sample selection methods. Proceedings of the Survey Research Methods Section of the American Statistical Association; 1979; p. 401-6.
16. Williams RL, Chromy JR. SAS Sample Selection Macros. Proceedings of the Fifth Annual SAS Users Group International Conference. Cary, NC: SAS Institute Inc; 1980.
17. Kessler RC, Rose S, Koenen KC, et al. How well can post-traumatic stress disorder be predicted from pre-trauma risk factors? An exploratory study in the WHO World Mental Health Surveys. *World Psychiatry* 2014; **13**(3): 265-74.

18. Lago L, Glantz MD, Kessler RC, et al. Substance dependence among those without symptoms of substance abuse in the World Mental Health Survey. *International journal of methods in psychiatric research* 2017; **26**(3).
19. Polley EC. Super Learner. Berkeley: University of California; 2010.
20. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one* 2015; **10**(3): e0118432.
21. Harrell FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer; 2015.
22. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Statistics in Medicine* 2019; **38**(21): 4051-65.
23. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st international conference on neural information processing systems; 2017; p. 4768-77.
24. Bharat C, Hickman M, Barbieri S, Degenhardt L. Big data and predictive modelling for the opioid crisis: existing research and future potential. *The Lancet Digital Health* 2021; **3**(6): e397-e407.
25. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. . 2019.
26. Polley E, LeDell E, Kennedy C, van der Laan M. SuperLearner: Super Learner Prediction. R package version 2.0-26; 2019.
27. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 2010; **33**(1): 1-22.
28. Chen T, He T, Benesty M, et al. xgboost: Extreme Gradient Boosting. R package version 1.3.2.1; 2021.
29. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* 2017; **77**(1): 1-17.
30. Venables WN, Ripley BD. Modern Applied Statistics with S. Springer, New York; 2002.
31. Milborrow S, Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's Fortran utilities with Thomas Lumley's leaps wrapper. earth: Multivariate Adaptive Regression Splines. R package version 5.3.0; 2020.
32. Demant J, Schierff LM. Five typologies of alcohol and drug prevention programmes. A qualitative review of the content of alcohol and drug prevention programmes targeting adolescents. *Drugs: Education, Prevention and Policy* 2019; **26**(1): 32-9.
33. Yang LH, Link BG, Ben-David S, et al. Stigma related to labels and symptoms in individuals at clinical high-risk for psychosis. *Schizophrenia Research* 2015; **168**(1): 9-15.
34. Livingston JD, Boyd JE. Correlates and consequences of internalized stigma for people living with mental illness: A systematic review and meta-analysis. *Social Science & Medicine* 2010; **71**(12): 2150-61.
35. Goodman B, Flaxman S. European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Mag* 2017; **38**: 50-7.
36. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, et al. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *International journal of methods in psychiatric research* 2006; **15**(4): 167-80.
37. Hawkins JD, Catalano RF, Miller JY. Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: implications for substance abuse prevention. *Psychol Bull* 1992; **112**(1): 64-105.

Results for “Development and evaluation of a risk algorithm predicting alcohol dependence after early onset of regular alcohol use”

Table 1. World Mental Health sample characteristics of high Income countries[a]

Country	Sampling	Field dates	Age range at time of interview[b]	Sample size		Response rate[c]	Mean (SD) age of regular use onset among study cohort[d]
				Part 1	Part 2		
Argentina	Eight largest urban areas of the country (approx. 50% of the total national population)	2015	18-98	3,927	2,116	77.3%	15.2 (2.0)
Australia	Nationally representative	2007	18-85	8,463	8,463	60.0%	15.7 (1.5)
Belgium	Nationally representative	2001-2	18-95	2,419	1,043	50.6%	15.4 (1.7)
France	Nationally representative	2001-2	18-97	2,894	1,436	45.9%	15.1 (2.0)
Germany	Nationally representative	2002-3	19-95	3,555	1,323	57.8%	15.7 (1.4)
Japan	Eleven metropolitan areas	2002-6	20-98	4,129	1,682	55.1%	15.7 (1.7)
The Netherlands	Nationally representative	2002-3	18-95	2,372	1,094	56.4%	15.4 (1.3)
New Zealand	Nationally representative	2004-5	18-98	12,790	7,312	73.3%	15.0 (1.9)
Northern Ireland	Nationally representative	2005-8	18-97	4,340	1,986	68.4%	15.6 (1.5)
Poland	Nationally representative	2010-11	18-65	10,081	4,000	50.4%	15.8 (2.0)
Spain	Nationally representative	2001-2	18-98	5,473	2,121	78.6%	15.2 (1.9)
Spain - Murcia	Murcia region	2010-12	18-96	2,621	1,459	67.4%	15.5 (1.6)
United States	Nationally representative	2001-3	18-99	9,282	5,692	70.9%	15.1 (2.0)
Total				**76,195**	**41,787**	**62.2%**	**15.3 (1.8)**

[a] The World Bank (2018) Data at: http://data.worldbank.org/country. Some of the WMH countries have moved into new income categories since the surveys were conducted. The income groupings above reflect the status of each country at the time of data collection. The current income category of each country is available at the preceding URL.

[b] For the purposes of cross-national comparisons we limit the sample to those 18+.

[c] The response rate is calculated as the ratio of the number of households in which an interview was completed to the number of households originally sampled, excluding from the denominator households known not to be eligible either because of being vacant at the time of initial contact or because the residents were unable to speak the designated languages of the survey.

[d] Unweighted mean (standard deviation, SD) among the 10,687 individuals who were 25 years or older at time of interview and reported drinking alcohol regularly by 17 years of age.

Figure 1. Receiver operating characteristic curve (left) and precision-recall curve (right) for the super learner model predicting alcohol dependence (AD) among individuals with an onset of regular alcohol use by 17 years of age in the holdout sample (n=3,208)

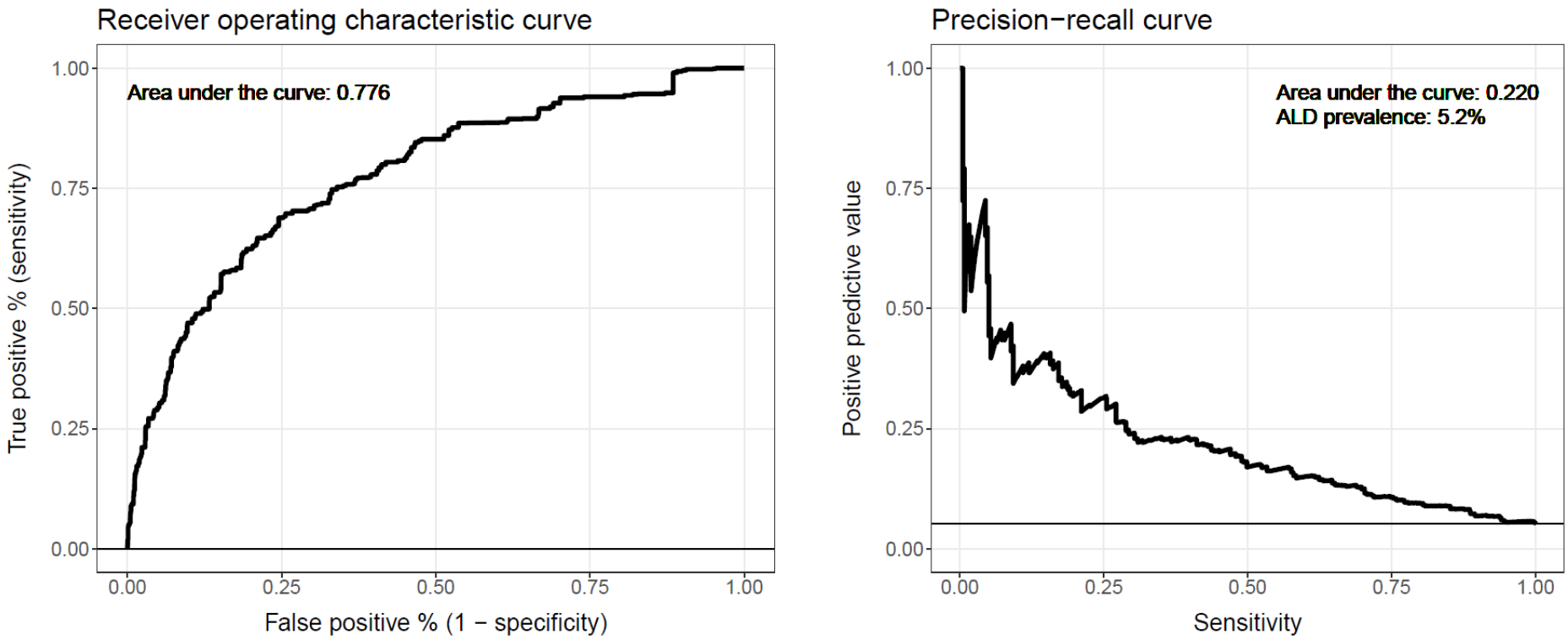


Figure 2. Calibration plot for super learner model predicting alcohol dependence (AD) among individuals with an onset of regular alcohol use by 17 years of age in the holdout sample (n=3,208)

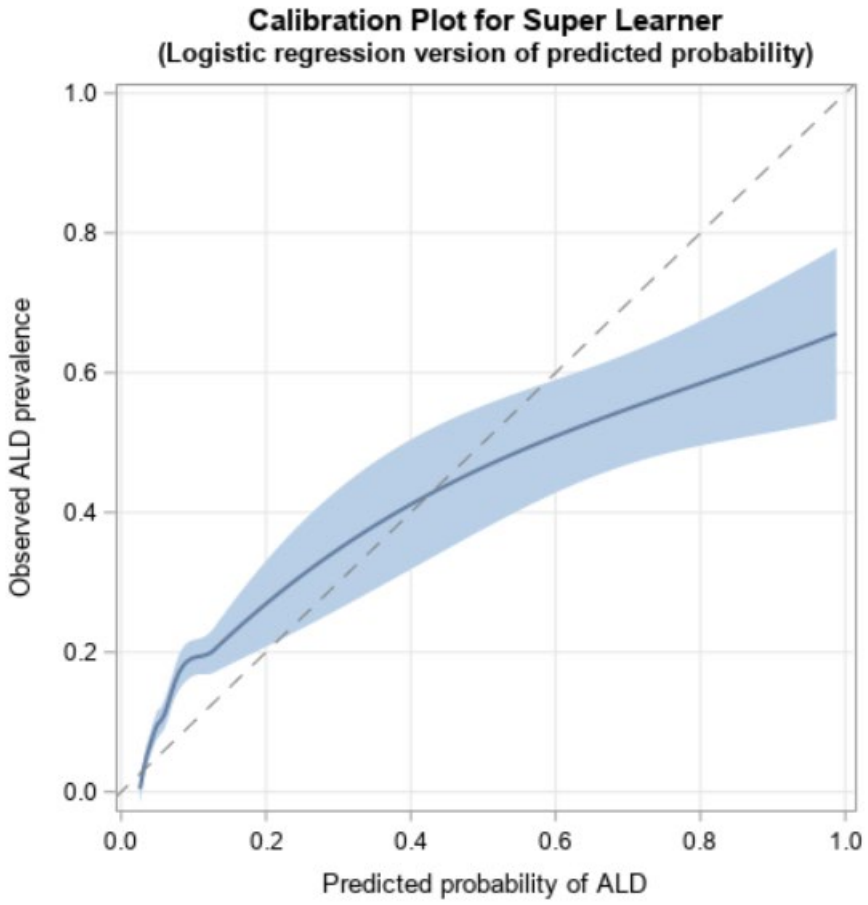


Table 2. Super learner coefficients and area under the receiver operating characteristic curve (AUC) for all candidate learners assigned non-zero coefficients in the super learner model as evaluated in the training sample[1]

Model	Super learner coefficient (β)	AUC (95% CI)	
		Training AUC	Holdout AUC[2]
Cross-validated SuperLearner	--	0.81 (0.80, 0.82)	--
SuperLearner	--	0.86 (0.85, 0.87)	0.78 (0.74, 0.81)
Extreme gradient boosting	0.48	0.85 (0.84, 0.86)	0.77 (0.73, 0.81)
Generalised linear model	0.33	0.84 (0.83, 0.85)	0.76 (0.72, 0.80)
Neural network	0.12	0.84 (0.83, 0.85)	0.73 (0.69, 0.77)
Adaptive polynomial spline regression	0.04	0.83 (0.84, 0.86)	0.77 (0.72, 0.80)
Adaptive regression splines (d=1)	0.03	0.82 (0.81, 0.83)	0.77 (0.73, 0.81)
Adaptive regression splines (d=4)	0.01	0.86 (0.85, 0.87)	0.74 (0.69, 0.78)

AUC, area under the curve; d, maximum degree of interaction; CI, confidence interval;

[1] The super learner model was developed using the training sample (n=3,318)

[2] Performance of the super learner model was evaluated using the holdout sample (n=3,208).

Table 3. Cross-validated performance metric estimates for predicted probability of alcohol dependence 10-years after early onset of regular alcohol use in the holdout sample

% Highest Risk[a]	Holdout sample in risk threshold group		Within risk threshold group			Cumulative[b]		
			Sensitivity	Specificity	PPV[d]	Sensitivity	Specificity	PPV
	N[c]	% of holdout sample (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)
5	81	2.0 (1.2, 2.7)	15.2 (9.6, 20.7)	1.3 (0.6, 2.0)	40.0 (23.7, 56.2)	15.2 (9.6, 20.7)	98.7 (98.0, 99.4)	40.0 (23.7, 56.2)
10	127	3.4 (2.7, 4.2)	12.2 (6.6, 17.9)	2.9 (2.2, 3.6)	18.9 (10.0, 27.8)	27.4 (20.5, 34.3)	95.8 (94.8, 96.8)	26.6 (19.1, 34.2)
15	144	3.4 (2.6, 4.1)	9.7 (6.7, 12.6)	3.0 (2.3, 3.7)	15.1 (9.0, 21.3)	37.0 (30.0, 44.1)	92.8 (91.6, 94.0)	22.2 (17.2, 27.3)
20	166	4.7 (3.6, 5.7)	12.0 (8.8, 15.1)	4.2 (3.2, 5.3)	13.5 (8.1, 18.9)	49.0 (42.0, 56.0)	88.6 (87.1, 90.0)	19.2 (15.4, 23.0)

PPV, positive predictive value; SE, standard error

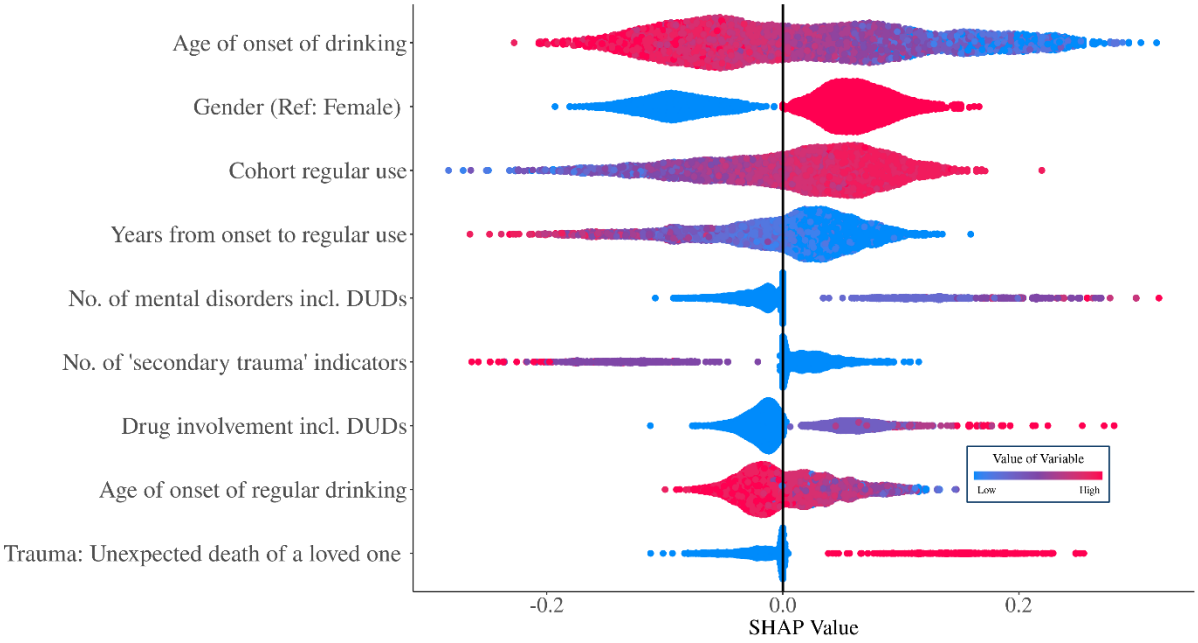
[a] Risk thresholds were created by rank ordering the super learner predicted probabilities in the training sample and creating 20 equally sized groups.

[b] Lower cutpoint of risk used as clinical threshold for defining someone as being at risk of developing alcohol dependence.

[c] Unweighted number of individuals from the holdout sample in the specific risk strata.

[d] Equal to the AD prevalence *within* the given risk threshold group.

Figure 3. SHAP summary plot for the nine top predictors in the Super Learner model as evaluated in the holdout sample



The y-axis shows the variable name, in order from importance from top to bottom. The x-axis shows the SHAP value. Measures of variable importance were evaluated as the change in the expected model prediction when conditioning on that feature across all possible combinations of variables. This marginal contribution was then weighted by the reciprocal of the total number of possible marginal contributions to all models of the same size and these predictor-specific weighted marginal contributions aggregated, providing a "local" summary measure of contribution to the model prediction from each variable. Gradient colour indicates the original value for that variable. For binary variables (gender and exposure to 'unexpected death of a loved one'), it will take two colours. For all other variables, the gradient colour can contain the whole spectrum.