What predicts risky alcohol use in a large cohort of Aus adolescents?

Lucinda Grummitt, Rachel Visontay, **Philip Clare**, Tim Slade, Louise Birrell

THE UNIVERSITY OF SYDNEY

Acknowledgement of country

I acknowledge the traditional owners of this land, the Gadigal people of the Eora Nation.

Declaration of Conflicts of Interest

Funding:

- I am supported by NSW Health via the Prevention Research Support Program
- Co-authors received funding from National Health and Medical Research Council (LB)

Conflicts:

- We have no relevant conflicts of interest to declare.

About this study

- We used data from 4,983 participants in the Longitudinal Study of Australian Children.
- We ran ensemble machine learning (SuperLearner) to predict risky alcohol use at age 18.
- Analysis included 220 predictors across childhood and adolescence.
- We had good predictive accuracy (AUC = 0.792)

Top Predictors

Predictors of harmful drinking at 18 years:

1. Weekly drinking at age 16
2. Lifetime cannabis use
3. Parent financial stress
4. Gender identity
5. ADHD diagnosis
6. Prenatal alcohol exposure
7. Housing insecurity

TABLE 4 Feature importance, weighted by SuperLearner coefficients and aggregated across folds for the top 20 predictors.

Feature (age in years)	Mean weighted importance	SD weighted importance
Weekly drinking (16)	0.99925097	0.002901
Lifetime cannabis use	0.44625379	0.05452562
Financial stress	0.4203031	0.03705029
Female	0.36512711	0.08061375
Male	0.3442192	0.05093263
ADHD	0.24809721	0.05555495
Pre-natal alcohol exposure	0.24807747	0.04023639
Housing insecurity	0.24328065	0.05270023
Religious involvement	0.23832599	0.04286569
Parent 1 alcohol use problem	0.21496378	0.04816134
Bully victimisation (16)	0.19304537	0.04394482
Anxiety (16)	0.1927689	0.04903178
Moral peers (16)	0.18138694	0.05219186
Neuroticism (16)	0.17994661	0.03000665
Family SES (4)	0.17742442	0.04458989
Lifetime other drug use	0.174376	0.0355995
Lower parental monitoring (14)	0.16979163	0.05638454
Parent 1 self-efficacy (12)	0.16660787	0.04208602
Moral peers (14)	0.16515154	0.04821889

Next steps

- Exploration of predictions in other data sets, particularly clinical data
- Casual inference analysis to test whether key predictors have a causal relationship with the outcomes

To find out more

Received: 17 December 2024 | Accepted: 15 June 2025

DOI: 10.1111/add.70145

RESEARCH REPORT

ADDICTION SSA

Multivariable machine learning prediction of risky alcohol use in contemporary youth

Lucinda Grummitt[1] | Rachel Visontay[1] | Philip Clare[2,3,4] | Tim Slade[1] | Louise Birrell[1]

[1]The Matilda Centre for Research in Mental Health and Substance Use, The University of Sydney, Sydney, Australia

[2]Prevention Research Collaboration, School of Public Health, The University of Sydney, Sydney, Australia

[3]National Drug and Alcohol Research Centre, UNSW Sydney, Sydney, Australia

[4]Charles Perkins Centre, The University of Sydney, Sydney, Australia

Correspondence
Lucinda Grummitt, The Matilda Centre for Research in Mental Health and Substance Use, Level 6, Jane Foss Russell Building, The University of Sydney, Sydney, NSW 2006, Australia.
Email: lucinda.grummitt@sydney.edu.au

Funding information
L.B. is supported by National Health and Medical Research Council Emerging Leader Investigator Grant 2016301. P.C. is supported by a New South Wales Health Prevention Research Support Program grant.

Abstract

Background and aims: Risky alcohol use in young adulthood is a significant public health concern. Understanding the predictors of risky drinking during this period is essential for prevention. This study aimed to measure the predictive accuracy of ensemble machine learning and identify the most important predictors of risky alcohol use in early adulthood.

Design and setting: Secondary analysis of the Longitudinal Study of Australian Children, an Australian national longitudinal cohort study.

Participants: A total of 4983 children, aged 4–5 years in 2004 (Wave 1), followed up for eight waves (to age 18/19 in 2018).

Measurements: Risky alcohol use was measured at age 18 and defined as more than 10 standard drinks per week, as per Australian National guidelines. Predictors from multiple domains—sociodemographic, adolescent substance use, adolescent mental health and behaviours, parental mental health and substance use, school factors, peer influences, parenting practices and parental stress—were included, measured from Wave 1 to 7. The SuperLearner package in R was used to test a series of models [regularised regression (LASSO, ridge and elastic net), random forest and kernel support vector machine (SVM)] using nested 10-fold cross-validation to identify the overall predictive ability of

DOI: 10.1111/add.70145