Statistical Analysis Plan

for

Koori Growing Old Well Study (KGOWS)

Alcohol use, associated risk factors and longitudinal outcomes for older urban and regional Aboriginal Australians.

SAP Version Number: 1.0

Date: 14 October 2021

Contents

1 ADMINISTRATIVE INFORMATION

Statistical Analysis Protocol

Version history

Version	Date	Description
SAP Version 1	14 October 2021	Final version

List of contributors

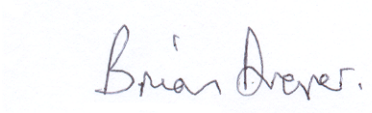
- Dr Louise Lavrencic – Senior Postdoctoral Fellow, Neuroscience Research Australia, University of New South Wales
- Dr Kylie Radford – Senior Research Scientist, Neuroscience Research Australia, University of New South Wales
- Prof Brian Draper – Conjoint Professor of Psychiatry, School of Psychiatry, University of New South Wales

2 SIGNATURE PAGE

We, the undersigned, have reviewed and approve this Statistical Analysis Plan.

The signatories confirm that:

a) They believe the procedures for the statistical analysis of the data from the *Koori Growing Old Well* study described in this document are appropriate;
b) Their intention is to analyse the data from the *Koori Growing Old Well* study using the statistical procedures described in this document; and
c) If, subsequently, the statistical analysis of the data is conducted in a way that differs importantly from the procedures described in this document, those differences will be explicitly outlined in reports of those analyses.

Signature	Date
	15/10/2021
Dr Louise Lavrencic Senior Postdoctoral Fellow, NeuRA	
	18/10/2021
Dr Kylie Radford Senior Research Scientist, NeuRA	
	15/10/2021
Prof Brian Draper Conjoint Professor, UNSW	

STUDY AIMS AND METHODS

The aims of this study are to:

1. Establish the prevalence of current and past alcohol use for older urban and regional Aboriginal Australians;

2. Determine the biomedical, psychological, cognitive and social lifecourse factors that are associated with alcohol use in this cohort; and

3. Examine the longitudinal outcomes of late-life alcohol use in this cohort.

Participants were part of the longitudinal Koori Growing Old Well Study (KGOWS), a population-based study spanning five Aboriginal communities in New South Wales (NSW), Australia with two metropolitan Sydney sites and three regional mid-North Coast sites. The baseline study has been described elsewhere (Radford et al., 2018; Radford & Mack, 2012; Radford et al., 2015; Radford et al., 2014). Baseline data collection occurred from March 2010 to September 2012 (n=336). Baseline participants who were contactable and willing to participate were recruited for a follow-up study, with follow-up data collection occurring from July 2016 to April 2018 (n=165).

The study was approved by the Aboriginal Health and Medical Research Council (AHMRC; 615/07), the University of New South Wales Human Research Ethics Committee (HREC 08003), and NSW Population & Health Services Research Ethics Committee (AU RED Ref: HREC/09/CIPHS/65; Cancer Institute NSW Ref: 2009/10/187).

STUDY POPULATION

Baseline inclusion criteria were identifying as an Aboriginal and/or Torres Strait Islander person, aged 60 years or older, and having resided in one of the five study catchment areas for at least 6 months. All baseline participants were eligible to take part in the follow-up study.

STATISTICAL ANALYSIS

The primary analyses will be conducted using binary logistic regression analyses. All tests will be 2-tailed and the nominal level of α will be 5%. Where data are missing for outcome measures, the number of observations will be reported; no imputation of missing values for the primary outcome will be carried out. P-values will not be adjusted for multiple comparisons; however, outcomes are clearly categorised by importance (i.e., primary and secondary). P-values will be rounded to three decimal places and values less than 0.001 will be reported as <.001.

Primary Outcomes

The primary outcome is baseline high-risk alcohol use. This will be assessed using the Alcohol Use Disorders Identification Test (AUDIT-C) (Bush, Kivlahan, McDonnell, Fihn, & Bradley, 1998).

Secondary Outcomes

Secondary outcomes include factors measured at follow-up that may be associated with baseline high-risk alcohol use.

Measures

Demographic data and comprehensive lifecourse and medical history were collected at baseline (see Table 1 for list of measures). These will be examined for statistical associations with baseline alcohol use.

Continuous predictor variables will be standardised (z-scored) prior to inclusion in models.

Table 1. *List of baseline measures to be used as predictor variables in statistical analyses.*

Measure	**Details**
Demographic	
Age	Continuous
Sex	1=female, 2=male
Urban/regional	0=regional, 1=urban
Education/cognitive	
Educational attainment	Continuous
Retrospective Indigenous Childhood Enrichment (RICE) scale	Continuous
Unskilled work history	0=partly skilled/skilled, 1=unskilled
Cognitive functioning (MMSE)	Continuous
Told had a learning disorder	0=no, 1=yes
Told had attention disorder or hyperactivity as a child	0=no, 1=yes
Adversity/trauma	
Premature baby	0=no, 1=yes
Death of a parent in childhood	0=no, 1=yes
Childhood trauma questionnaire (CTQ)	Continuous
Racism (MIRE)	Continuous
Taken away or family member taken away	0=no, 1-yes
Resilience (CD-RISC)	Continuous
Police custody	0=no, 1=yes
Childhood socioeconomic disadvantage (IRSD scores)	Continuous
Health/medical	
Cancer/leukaemia	0=no, 1=yes
Diabetes	0=no, 1=yes
Hypercholesterolemia	0=no, 1=yes
Hypertension	0=no, 1=yes
Hepatitis	0=no, 1=yes
Renal disease	0=no, 1=yes
Body mass index	Continuous
Heart disease	0=no, 1=yes
Stroke/TIA	0=no, 1=yes
Head injury	0=no, 1=yes
Epilepsy	0=no, 1=yes
Hospitalisation (past year)	0=no, 1=yes
Falls (past year)	0=no, 1=yes
Mean arterial pressure	Continuous
Polypharmacy	0=no, 1=yes
Physical activity (moderately energetic)	0=no, 1=yes

Impairment of activities of daily living	0=no, 1=yes
Past high-risk alcohol use (AUDIT-C past)	0=no, 1=yes
Smoking status	0=never, 1=ex-smoker, 2=current smoker
Illicit drug use	0=no, 1=yes
Multimorbidity (two or more conditions, based on presence/absence of high blood pressure, diabetes, heart problems, stroke, lung problems, current hepatitis, history of cancer, epilepsy, arthritis, AUDIT-C score ≥ 6, mPHQ9 score ≥ 9, and dementia diagnosis)	0=no, 1=yes
Psychological	
History of depression (in lifetime)	0=no, 1=yes
History of anxiety/PTSD (in lifetime)	0=no, 1=yes
Depressive symptoms (mPHQ9)	Continuous
History of psychotic illness ("other mood disorders like bipolar or manic depression" and "schizophrenia or similar disorders")	0=no, 1=yes
History of self-harm	0=no, 1=yes
History of suicidal ideation	0=no, 1=yes
Current suicidal/self-harm ideation (mPHQ9 item 9)	0=never, 1=a little/a lot/all the time
Social/cultural	
How many people in the house where grew up	Continuous
Suspension from school	0=never, 1=rarely, 2=monthly, 3=weekly or more
Expelled from school	0=no, 1=yes
Any relatives been in prison	0=no, 1=yes
Culture is a source of strength	0=none/a little, 1=somewhat/a lot
Lives alone	0=no, 1=yes
Feels lonely	0=almost never, 1=sometimes/quite often
Social activity participation	Continuous

Baseline alcohol use will also be examined in relation to factors collected at follow-up (see Table 2).

Table 2. *List of follow-up measures to be used as outcome variables in statistical analyses.*

Measure	**Details**
Cognitive	
Cognitive functioning (MMSE)	Continuous
Physical	
Cancer (incident cases)	0=no, 1=yes
Hospitalisation (past year)	0=no, 1=yes
Falls (past year)	0=no, 1=yes
Mean arterial pressure	Continuous
Impairment of activities of daily living	0=no, 1=yes
Psychological	

Depressive symptoms (mPHQ9)	Continuous
Resilience (CD-RISC)	Continuous

Missing Data and Outliers
To maximise inclusion of participants in analyses, for cases where psychological or cognitive tests have less than 25% missing items, total scores will be prorated according to the suggested formulas (from relevant test manual) or based on previous literature.

Both multiple imputation and inverse probability weighting (IPW) will be implemented prior to conducting analyses. The following process will be followed:

1. Create imputed datasets using multiple imputation to impute item missing values;
2. Use IPW on the imputed datasets to account for participants with baseline but not follow-up data (where reason for non-participation was not due to death);
3. Combine the imputed/weighted results via Rubin's method.

If outliers are present in the data or the model assumptions are grossly violated, sensitivity analyses will be conducted for those specific models.

Analyses of Primary Outcome
Descriptive statistics for continuous (mean, standard deviation, median, interquartile range, range) and categorical/dichotomous variables (number and percentage in each group) will be reported for the original data, as appropriate. Descriptive statistics for imputed data will be examined to ensure imputed values are reasonable.

Univariable comparisons will be made between groups (abstinent/low vs high-risk alcohol use) using logistic regression analyses for all baseline variables (Table 1). Significant variables ($p<.05$) from univariable analyses will then be entered into multivariable logistic regression models in blocks, predicting abstinent/low or high-risk alcohol use. The first block will include demographic variables; the second block will add to the first, variables that are potential antecedents of high-risk alcohol use (i.e., occurring in early-life); the third block will add to the second, variables that are potentially co-morbid (i.e., occurring in late-life). For this complete model including all variables, predictors with $p>.1$ will subsequently be removed; as such, only predictors with $p<.1$ will be retained in the final model.

Analyses of Secondary Outcomes
Secondary outcome measures include factors at follow-up that may be associated with baseline high-risk alcohol use (Table 2). The appropriate regression method (e.g., logistic, linear) will be used depending on the nature of the outcome variable. The follow-up variables will be entered as outcome variables in univariable analyses. High-risk alcohol use (dichotomous) will be entered as a predictor, and the corresponding baseline variable will also be entered as a predictor (e.g., if follow-up MMSE is the outcome, baseline MMSE will be entered as a predictor to control for baseline level); age will be included in the models to adjust for potential effects. High-risk alcohol use at follow-up (dichotomous), and an interaction term (baseline high-risk alcohol use × follow-up high risk alcohol use) will then be added to the models, to determine whether potential relationships are with contemporaneous or baseline level of drinking, or change in drinking habits; when adding these variables, collinearity diagnostics (tolerance, variance inflation factor) will be examined for linear regression models and estimates (beta, standard error, confidence intervals for odds

ratios) will be examined for logistic regression models. Based on these statistics, if the additional variables are collinear with baseline high-risk alcohol use and/or each other, these models will not be used or reported.

Separate multinomial logistic regression analyses will be conducted with a 3-level categorical variable reflecting participation in follow-up as the outcome: participated in follow-up, did not participate in follow-up, died before follow-up. Age and high-risk alcohol use at baseline will be entered as predictors in the model. There is a group of participants who did not participate in follow-up for whom whether they were alive at follow-up is unknown. Therefore, the model will be run twice as a sensitivity analysis, once assuming this group of people were all alive at follow-up, and once assuming this group had all died before follow-up.

Additional Analyses

A Chi Square analysis will be reported to investigate the relationship between past and baseline high-risk alcohol use.

RATIONALE FOR ANY DEVIATION FROM PRE-SPECIFIED ANALYSIS PLAN

Logistic regression analyses will not be conducted if the following criteria is not met: minimum of 5 events per variable for logistic regression analyses (Vittinghoff & McCulloch, 2006). Instead, in the event of separation or quasi-separation, either Firth's penalised regression for rare events procedure (Firth, 1993) or Bayesian logistic regression will be conducted depending on appropriateness.

Collinearity assumptions will be checked. If variables are highly collinear, one of the variables of concern will be removed (e.g., if examining corresponding baseline and follow-up variables in the same model).

REFERENCES

Bush, K., Kivlahan, D. R., McDonnell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT Alcohol Consumption Questions (AUDIT-C): An effective brief screening test for problem drinking. *JAMA Internal Medicine, 158*, 1789-1795.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*(1), 27-38. doi:10.1093/biomet/80.1.27

Radford, K., Lavrencic, L. M., Delbaere, K., Draper, B., Cumming, R., Daylight, G., . . . Broe, G. A. (2018). Factors associated with the high prevalence of dementia in older Aboriginal Australians. *Journal of Alzheimer's Disease*(Preprint), 1-11. doi:10.3233/JAD-180573

Radford, K., & Mack, H. A. (2012). *Dementia screening for urban Aboriginal Australians: The Modified Kimberly Indigenous Cognitive Assessment (mKICA)*. Sydney: Dementia Collaborative Research Centres.

Radford, K., Mack, H. A., Draper, B., Chalkley, S., Daylight, G., Cumming, R., . . . Broe, G. A. (2015). Prevalence of dementia in urban and regional Aboriginal Australians. *Alzheimer's & Dementia, 11*, 271-279. doi:10.1016/j.jalz.2014.03.007

Radford, K., Mack, H. A., Robertson, H., Draper, B., Chalkley, S., Daylight, G., . . . Broe, G. A. (2014). The Koori Growing Old Well Study: Investigating aging and dementia in urban Aboriginal Australians. *International Psychogeriatrics, 26*, 1033-1043. doi:10.1017/S1041610213002561

Vittinghoff, E., & McCulloch, C. E. (2006). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology, 165*, 710-718. doi:10.1093/aje/kwk052