Statistical Analysis Plan (Addendum)

Increasing Knowledge of Alcohol as a Risk Factor for Breast Cancer Among Women Attending Breast Screening Services (Health4Her)

ClinicalTrials.gov Identifier: NCT04715516.

Pre-registered: Wednesday 20 January 2021

Version: 1

Date version finalised: Friday 18 February 2022

Study rationale and design

Alcohol is a major modifiable risk factor for breast cancer in women, yet this is not widely understood by health practitioners or policy makers, let alone the general population. We planned to undertake a pragmatic, single-site, double-blind, parallel group randomised controlled trial to test the effects of an alcohol brief intervention among women attending breast screening services, to improve knowledge of alcohol as a risk factor for breast cancer and reduce harmful alcohol use.

Women attending Maroondah BreastScreen for routine mammography (N = 558) completed a baseline assessment before being randomised to receive 4 minutes of alcohol brief intervention plus 3 minutes of lifestyle health promotion (active arm, n = 279), or 3 minutes of lifestyle health promotion, not inclusive of alcohol information (control arm, n = 279). For the active arm, participant responses to questions about current alcohol use branched to personalised feedback consistent with level of alcohol consumption (i.e. drinking within or above current Australian Alcohol Guidelines). Interventions were delivered by way of an animation on an iPad. Both groups also received take-home information pertaining to group assignment. Follow-up assessments were completed by telephone at 4- and 12-weeks post-randomisation.

Statistical Analysis Plan (SAP)

This plan, or SAP Addendum, provides additional detail to the statistical considerations that were documented in the statistical analysis plan in Health4Her Protocol_V7_08 08 21 (available in the trial record on ClinicalTrials.gov).

Data will be collated, cleaned and validated using programmed edit checks.

Outcomes with measurements at more than two time points

In general, outcome variables with repeated measurements will be analysed by fitting linear mixed models (for continuous outcomes or outcomes treated as continuous, such as 5-point Likert scales) or (for categorical / binary outcomes) generalised linear mixed models (GLMMs) or generalised additive mixed models (GAMMs), with fixed effects for treatment and time, and their interaction, and random effects for subjects and assessments within subjects, using restricted maximum likelihood (REML) or similar algorithms. These models allow missing values to be accommodated.

These methods will allow the most suitable variance-covariance model for the repeated measures to be selected, using Akaike's Information Criterion, and, if appropriate, commonality of nonlinear trends over time to be explored via splines. The F-test (or equivalent chi square test, t-test of coefficients in logistic regression) will be used to test for an overall group by time interaction in outcomes measured at all three follow-up times (0, 4 and 12 weeks).

Outcomes measured at two or fewer time points

For outcome variables measured at two or fewer time-points, different models will be used, depending on the nature of the outcome variable:

1. Where the outcome is binary and proportions are required, a test of change in proportions for categorical outcomes, e.g. the Cochran-Mantel-Haenszel test, will be applied or, if covariates are required in the analysis, logistic regression with random effects for subjects. Such logistic regression models are instances of GLMMs or GAMMs.
2. For continuous outcomes or outcomes treated as continuous (such as 5-point Likert scales) linear mixed models will be used (as for outcomes measured at three time-points).
3. In simple cases of the preceding points, the analysis may reduce to (in the sense of being equivalent to) a test of whether there is a difference between the two groups (treatment and control) in the changes in the mean (for continuous outcomes). If just one time point is involved, we may use a binomial test of proportions (for categorical outcomes).

Outcomes measured at 4 and 12 weeks but without a baseline measurement will be covered by this case.

Missing values

Where possible, missing values will be handled under the missing at random assumption; that is, unless there is good reason to believe this assumption is untenable. In cases of doubt the analysis will be re-done with (a) missing values imputed and (b) the corresponding subjects removed from the analysis.

Outcomes where there is no baseline measure possible (that is, by design) will not be treated as having the baseline value missing, but rather as belonging to the two time-point case.

Exploratory analyses

Exploratory analyses will be performed using the same methods as for the primary and secondary outcomes (as appropriate) plus, optionally, such other modelling strategies as may be useful (e.g. random forests or decision trees as exploratory tools, or other classification tools).

In exploratory analyses of outcomes measured at three time-points, mixed models with covariates as relevant (see list of potential covariates below), including their interactions with treatment group, will be fitted in order to identify moderating factors. Categorical, ordinal and binary outcomes will be analysed in a similar way using generalised linear mixed models (GLMMs) or generalised additive mixed models (GAMMs).

In exploratory analyses of outcomes measured at two time-points, logistic regression with random effects for subjects will be utilised to assess the effects of potential covariates (see list of potential covariates below).

Statistical software

Analyses will be conducted using the most appropriate procedures in GenStat, R, SAS or STATA and additional analyses not specified in the published protocol or this addendum will be regarded as exploratory.

Statistical Analysis of the Primary and Key Secondary Outcome Variables

Primary Outcome

The primary analysis will take place after all subjects, not known to have withdrawn or not deemed lost to follow-up, have had their 12-week assessments, based on the intention to treat (ITT) principle (i.e. subjects' data are analysed as randomised and as stratified).

The primary outcome of this repeated measures RCT is the proportion of participants accurately identifying alcohol as a clear risk factor for breast cancer at 4-weeks post-randomisation. Inferences about the effect of the Health4Her intervention on the primary outcome will be based on a comparison, between treatment groups, of change from baseline to 4-weeks post-randomisation using a Cochran-Mantel-Haenszel test. This analysis tests for homogeneity of the odds ratio across strata (i.e. intervention and control). A binomial test (that is, a test of proportions) will be used as an ancillary analysis of the primary outcome (i.e. the difference in proportions between the two groups at 4-weeks post-randomisation).

Relevant covariates (from the list of potential covariates below) may also be used in further analysis, using generalised linear models (logistic regression) to make comparisons. Additional analyses may use covariates for post-stratification.

Secondary Outcomes

Alcohol Consumption

Alcohol consumption will be assessed by the Alcohol 14-day Timeline Follow-back (TLFB). Heavy drinking days are measured as >40 grams of alcohol (>4 Australian standard drinks). Alcohol consumption will also assessed by the AIHW alcohol frequency quantity items. Assessments are made at 0, 4 and 12 weeks and linear mixed models (continuous variables) or GLMMs or GAMMs (categorical / binary variables) will be used. The following outcomes will be analysed (transformed if necessary for analysis):

Secondary outcome 1 – Proportion of participants drinking less than or equal to 10 standard drinks per week at 4-weeks and 12-weeks post-randomisation. Derived from (operationalised in two ways):

- Has the participant consumed >10 standard drinks in week one AND/OR >10 standard drinks in week two? (tlfb_nhmrc_both_comprh)
- Has the participant consumed >10 standard drinks in both week one AND week two? (derived variable: tlfb_nhmrc_bothwks)

Secondary outcome 2 – Proportion of participants drinking less than or equal to 10 standard drinks per week at 4-weeks and 12-weeks post-randomisation, among only participants who drink more than 10 standard drinks per week at baseline.

Subgroup defined as either i. women who consumed >10 standard drinks in week one AND/OR >10 standard drinks in week two (tlfb_nhmrc_both_comprh) or ii. women who consumed >10 standard drinks in both week one AND week two (tlfb_nhmrc_bothwks).

Outcome derived from:

- Has the participant consumed >10 standard drinks in week one AND/OR >10 standard drinks in week two? (tlfb_nhmrc_both_comprh)
- Has the participant consumed >10 standard drinks in both week one AND week two? (tlfb_nhmrc_bothwks)

Secondary outcome 3 – Change in alcohol consumption at 4-weeks and 12-weeks post-randomisation. Derived from:

- TLFB: Drinking Days (tlfb_drinkdays)
- TLFB: Drinking Days with > 2 Standard Drinks (tlfb_days_2stand)
- TLFB: Drinking Days with > 4 Standard Drinks (tlfb_days_4stand)
- TLFB: Total Standard Drinks in the Past 14 days (tlfb_totalmonth)
- Derived from AIHW frequency quantity items: average standard drinks per day (derived variable: alc_perday)
- AIHW frequency: In the past month, how often did you have an alcoholic drink of any kind? (alc_freq)
- AIHW quantity: quantity of alcohol consumed on a typical day calculated in standard drinks (alc_quant_sd)

Secondary outcome 4 – Change in alcohol consumption at 4-weeks and 12-weeks post-randomisation, among only participants who drink more than 10 standard drinks per week at baseline

Subgroup defined as either i. women who consumed >10 standard drinks in week one AND/OR >10 standard drinks in week two (tlfb_nhmrc_both_comprh) or ii. women who consumed >10 standard drinks in both week one AND week two (tlfb_nhmrc_bothwks).

Outcome derived from:

- TLFB: Drinking Days (tlfb_drinkdays)
- TLFB: Drinking Days with > 2 Standard Drinks (tlfb_days_2stand)
- TLFB: Drinking Days with > 4 Standard Drinks (tlfb_days_4stand)
- TLFB: Total Standard Drinks in the Past 14 days (tlfb_totalmonth)
- Derived from AIHW frequency quantity items: average standard drinks per day (alc_perday)
- AIHW frequency: In the past month, how often did you have an alcoholic drink of any kind? (alc_freq)
- AIHW quantity: quantity of alcohol consumed on a typical day calculated in standard drinks (alc_quant_sd)

Alcohol health literacy

Secondary outcome 5 – Change in participants' attitudes regarding alcohol and breast cancer risk from baseline to 4-weeks post-randomisation. A linear mixed model (for repeated measures) will be used to analyse the Likert scale response (derived from atttd_alcrisk). Transformations of response variables will be used as appropriate, such as in GLMMs or GAMMs.

Secondary outcome 6 – Proportion of participants accurately identifying i) the amount of alcohol in an Australian standard drink; ii) the number of standard drinks in an average restaurant serve of red wine; iii) the maximum number of standard drinks per week recommended by current Australian alcohol guidelines, at 4-weeks post-randomisation).

Analyses will be the comparison, between treatment groups, of change from baseline to 4-weeks post-randomisation using a Cochran-Mantel-Haenszel test. Relevant covariates (from the list of potential covariates below) may also be used in further analysis, using generalised linear models (logistic regression) to make comparisons. Additional analyses may use covariates for post-stratification. A binomial test (that is, a test of proportions) may be used as an ancillary analysis of the outcome (i.e. the difference in proportions between the two groups at 4-weeks post-randomisation), with logistic regression as an option if covariates are employed.

Outcome derived from:

- Proportion of participants who accurately respond to “In Australia, a standard drink contains ___ grams of alcohol” knowl_sd = 1 (10 grams)
    - recode to binary variable (1 = accurate | 2 = not accurate)

- Proportion of participants who accurately respond to “An average restaurant serve of red wine contains...” knowl_redwine = 3 (1.6 standard drinks)
    - recode to binary variable (1 = accurate | 2 = not accurate)

- Proportion of participants who accurately respond to “According to the latest Australian alcohol guidelines, it is recommended that healthy women should drink no more than ____ standard drinks per week.” knowl_alcguideline = free-text response of "10"
    - recode to binary variable (1 = accurate | 2 = not accurate)

Secondary outcome 7 – Proportion of participants who have accessed health information on i) alcohol harms, ii) alcohol and breast cancer risk, and iii) alcohol harm-reduction.

A binomial test (that is, a test of proportions) will be used (i.e. the difference in proportions between the two groups at 12-weeks post-randomisation), with logistic regression as an option if covariates are employed.

Outcome derived from info_sought_catg = 1 (alcohol) and/or 2 (alcohol and breast cancer risk) and/or 3 (alcohol and other health risks) and/or 4 (alcohol and women's health)

- recode to binary variable (1 = accessed | 2 = did not access)

(NB. This outcome is described in the ClinicalTrials.gov trial registration as assessed at 4-weeks post-randomisation, however to maintain blinding of Research Assistant these data were collected at 12-weeks post-randomisation.)

Quality of Life / health-related quality of life

Secondary outcome 8 - Change in general health/health-related quality of life at 4- and 12-weeks post-randomisation will be assessed using the SF-12.

Physical and mental component summary scores will be derived from SF-12 variables (sf12_health, sf12_activmod, sf12_activstair, sf12_physless, sf12_physlimit, sf12_emoless, sf12_emocaref, sf12_pain, sf12_calm, sf12_energy, sf12_down, sf12_interf).

Weighting coefficients for computing the scores will be those specified in Tucker et al (2010)[i].

Linear mixed models (for repeated measures) will be used to analyse the continuous responses (derived from the two SF-12 scores). Transformations of response variables will be used as appropriate.

Secondary outcome 9 – Change in quality of life at 4- and 12-weeks post-randomisation will be assessed using the EUROHIS-QOL single item. Linear mixed models (for repeated measures) will be used to analyse the Likert scale response (derived from eurohis_qol1). Transformations of response variables will be used as appropriate.

Potential additional exploratory analyses

As above, for repeated measures outcomes, linear mixed models or GLMMs or GAMMs will be used (or test of change in proportions for categorical outcomes, e.g. Cochran-Mantel-Haenszel test). When examining differences between groups at one follow-up time point (i.e. non-repeated measures), a binomial test will be used to test the difference in proportions between the two groups.

i. Change in the proportion of women who identify alcohol as a risk factor for breast cancer without prompt (0 and 4-weeks post-randomisation)
    - derived from knowl_alc
    - binary variable, Y | N

ii. Change in the proportion of women who identify inactivity as a risk factor for breast cancer without prompt (0 and 4-weeks post-randomisation)
    - derived from knowl_inactiv
    - binary variable, Y | N

iii. Change in the proportion of women who identify excess weight as a risk factor for breast cancer without prompt (0 and 4-weeks post-randomisation)
   - derived from knowl_weight
   - binary variable, Y | N

iv. Change in total minutes of physical activity in the past week (0, 4-week and 12-week post-randomisation)
   - derived from phys_activetotal
   - continuous variable

v. Proportion of women reporting a perceived change in physical activity (4-week and 12-week variable only, no baseline)
   - derived from phys_change
   - categorical variable, 1 = more physically active | 2 = less physically active | 3 = about the same

vi. Proportion of women reporting a perceived improvement in physical activity (4-week and 12-week variable only, no baseline)
   - derived from phys_change
   - recode categorical variable, 1 = more physically active | 0 = not more physically active

vii. Change in past-week vegetable intake (0, 4-week and 12-week post-randomisation)
   - derived from veg_intake
   - continuous variable, days out of 7 days

viii. Change in past-week fruit intake (0, 4-week and 12-week post-randomisation)
   - derived from fruit_intake
   - continuous variable, days out of 7 days

ix. Proportion of participants accurately identifying inactivity as a clear risk factor for breast cancer (0 and 4-weeks post-randomisation)
   - derived from bcriskfact_inactive
   - recode to binary variable (1 = identified as a clear risk factor | 2 = not identified as a clear risk factor)

x. Proportion of participants accurately identifying alcohol as a clear risk factor for breast cancer (0 and 4-weeks post-randomisation)
   - derived from bcriskfact_alc
   - 4-level categorical variable (1 = clear risk factor | 2 = possible risk factor | 3 = not a proven risk factor | 4 = unsure)

xi. Proportion of participants accurately identifying overweight as a clear risk factor for breast cancer (0 and 4-weeks post-randomisation)
   - derived from bcriskfact_ovweight

- recode to binary variable (1 = identified as a clear risk factor | 2 = not identified as a clear risk factor)

xii. Change in knowledge of exercise as a bc risk factor (0 and 4-weeks post-randomisation)
- derived from knowl_exercise
- accurate response knowl_exercise = 2 (at least 2.5 hours)
- recode to binary variable (1 = accurate | 2 = not accurate)

xiii. Change in knowledge of diet as a bc risk factor (0 and 4-weeks post-randomisation)
- derived from knowl_diet
- accurate response knowl_diet = 1 (dairy)
- recode to binary variable (1 = accurate | 2 = not accurate)

xiv. Change in attitude "It is important that I understand the risk factors for breast cancer" (0 and 4-weeks post-randomisation)
- derived from atttd_riskundrstnd
- Likert response

xv. Change in attitude "I believe I am able to change my risk of breast cancer" (0 and 4-weeks post-randomisation)
- derived from atttd_changerisk
- Likert response

Potential covariates for exploratory analyses of primary and secondary outcomes

- age (continuous; categorical 40-49, 50-59, 60-69, 70+ years)
- lgbtq+
- education
- born in Australia
- culturally and linguistically diverse background
- household composition
- disability status
- alcohol consumption at baseline (drinking above/within current NHMRC guidelines, 14-day TLFB subscales scores, AIHW frequency quantity scores)
- QoL/health-related QoL at baseline (SF-12 mental and physical component scores, EUROHIS-QOL single item)
- health literacy at baseline (attitude and knowledge scores)

[i]Graeme Tucker, Robert Adams, David Wilson. New Australian population scoring coefficients for the old version of the SF-36 and SF-12 health status questionnaires, Qual Life Res (2010) 19:1069-1076, DO 10.1007/s11136-010-9658-9