



MRC Integrative
Epidemiology
Unit



University of
BRISTOL

School of Psychological Science
12a Priory Road
Bristol BS8 1TU
United Kingdom

kd16662@bristol.ac.uk

www.bristol.ac.uk

STATISTICAL ANALYSIS PLAN

WOULD CONSUMERS BE CONFUSED BY THE TERM 'MILK' ON MILK SUBSTITUTE LABELS? AN ONLINE EXPERIMENTAL STUDY.

Katie De-loyde, Marcus Munafò, Angela Attwood, Oliva Maynard

Definitions

Dairy milk: (fluid) milk from an animal source

Milk substitute: (fluid) milk not from an animal source (from a plant-based source).

Study Design

This study will be an experimental study conducted online.

Participants will be randomised in a 1:1 ratio, to one of two conditions, in a between-subjects parallel-groups design:

No Milk Labelling Condition: Where milk substitutes are not labelled with the term milk, and instead have examples of current UK labelling found in supermarkets.

Milk Labelling Condition. Where milk substitutes are labelled with the term milk.

See Appendix 1 for further details on carton design and labelling.

Study Hypotheses

Primary hypothesis:

Participants in the Milk Labelling Condition will correctly identify more milk substitutes as being a product that could be added to a cup of tea or coffee compared to participants in the No Milk Labelling Condition - using the primary outcome.

Secondary hypothesis:

Participants in the Milk Labelling Condition will misidentify more milk substitutes as animal source products compared to participants in the No Milk Labelling Condition - using the secondary outcome.

Tertiary hypotheses:

Participants in the Milk Labelling Condition will be faster at identifying milk substitutes as products that could be added to a cup of tea or coffee compared to the No Milk Labelling Condition - tertiary outcome 3.1.

Participants in the Milk Labelling Condition will be slower at identifying milk substitutes as non-animal source products compared to the No Milk Labelling Condition - tertiary outcome 3.2.

Participants in the Milk Labelling Condition will be more likely to themselves choose a milk substitute to put in their cup of tea / coffee or a bowl of cereal in the future compared to participants in the No Milk Labelling Condition - tertiary outcomes 3.3 and 3.4.

Study questions

1) Please select whether you think someone would add each of the products to a cup of tea or coffee. This includes those products which would be added to tea or coffee by someone using the product to replace dairy milk. (yes, would be added / no, would not be added / unsure).

-100% correct would be 10 x yes to milk substitute cartons

2) Please select whether you think each product listed below comes from an animal source or not (yes, comes from an animal source / no, does not come from an animal source / unsure).

-100% misidentified would be 10 x yes to milk substitute cartons

3) Please select whether you think, in the future, you may purchase each product below to add to your tea or coffee (or for another such use, for example, to be added to a smoothie or bowl of cereal) (yes, I would purchase / no, I would not purchase / unsure).

-Number of yes answers out of the 10 milk substitutes.

Outcomes

The primary outcome will be:

The number of milk substitutes the participant correctly identifies in question 1 (out of 10).

100% correct would be 10 x yes to milk substitute cartons (i.e., a no or unsure answer would be considered an answer that does not correctly identify milk substitutes as being a product that could be added to a cup of tea or coffee).

The secondary outcome will be:

The number of milk substitutes the participant misidentifies in question 2 (out of 10).

100% misidentified would be 10 x yes or unsure to milk substitute cartons (i.e., a no answer would be considered the only answer that does correctly identify a milk substitute as not being from an animal source).

Tertiary outcomes will be:

3.1 The time it takes to complete question 1 (in seconds).

3.2 The time it takes to complete question 2 (in seconds).

3.3 The number of milk substitutes the participant selects in question 3 (out of 10).

3.4 Were any milk substitutes selected in question 3 (yes or no)

Additional measures

Participant gender, age, highest education level, long-term dietary restrictions (i.e., following a vegan or dairy-free diet) baseline dairy milk consumption (Question: How many days a week do you consume dairy milk in an average week) and baseline milk substitute consumption (Question: How many days a week do you consume a milk substitute (i.e. a plant-based milk) in an average week) will also be collected to characterise the study population and so that balance between conditions can be viewed.

Outliers

Any outliers for each outcome will be identified using range checks, scatter plots and histograms. True outliers will be defined as any value where the median absolute deviation exceeds 3. Any true outliers will be included in the primary and secondary analysis but, if deemed necessary, a sensitivity analysis will be completed without any true outliers to compare results.

Missing data

Data that are not applicable will be coded as -888. Data that are missing due to an administrative or other error will be coded as -999.

Missing data checks

As this is an online study, with forced responses, we anticipate that all outcomes will be complete.

Descriptive statistics

A CONSORT diagram will be produced showing the flow of participants through the study (see example in Protocol Appendix).

A table will be produced showing additional measures (i.e., participant demographics) between the study conditions.

A second table will be produced showing the mean and median number of correct answers given for each of the three different product categories (dairy milk, milk substitutes and other drinks), for each of the three questions, between study conditions.

See below for example:

Question	Correct answers (mean, median, SD)					
	Dairy milk (/5)		Other drink (/5)		Milk substitute (/10)	
	Milk label condition	No milk label condition	Milk label condition	No milk label condition	Milk label condition	No milk label condition
Q1						
Q2						
Q3						

A third table will be produced showing the responses (yes, no and unsure) for each question, presented descriptively for each milk substitute label (i.e., how many participants correctly identified ‘soya drink’ and how many correctly identified ‘soya milk’).

See example below for one question:

Milk substitute	Response					
	Yes		No		Unsure	
	Milk label condition	No milk label condition	Milk label condition	No milk label condition	Milk label condition	No milk label condition
Oatly						
Alpro Soya						
Soya						
Rice						
Almond						
Mylk						
Oat						
Soya						
Cashew						
Dream rice						

Outcome analysis

Primary outcome

A GLM (general linear model), or similar (see **Model types** below), will be used to compare the means of the primary outcome (the number of **yes** answers given to milk substitutes [out of 10]) between the two study conditions. The difference in means with 95% CI (confidence interval) of the mean, t statistics and p-values will be reported.

Secondary and tertiary outcomes

The secondary outcome (the number of yes or unsure answers given to milk substitutes [out of 10]) and outcome 3.3 (the number of yes answers the participants gives to whether they themselves would choose a milk substitute to put in their tea or coffee in the future [out of 10]) will also use a GLM, or similar (see **Model types** below), to compare the means of each outcome between the two study conditions.

Tertiary outcomes 3.1 and 3.2 will be compared between study conditions using an GLM, or similar. For each condition, the mean time it took to answer each question will be reported separately for those who were 100% correct and those who were not 100% correct, alongside the difference in means with 95% CI of the mean, t statistics and p-value.

See below for example table for question 1:

	Secondary outcome 3.1 (time taken to answer Q1)		
Study condition	Mean overall time taken to answer	Mean time to answer for those who were 100% correct	Mean time to answer for those who were not 100% correct
Milk label			
No milk label			

Secondary outcome 3.4 will be compared between study conditions using a logistic regression model to predict whether participants select a milk substitute (or not). The odds ratio (OR), along with a 95% and p-value will be reported.

Model types

The type of model used for each analysis of continuous outcomes, will be based on the scale and distribution of each outcome. We anticipate skewed distributions towards ten for the primary outcome, and towards zero for the secondary outcome, and will therefore consider a transformation of each outcome, before fitting the appropriate model. Non-linear models will be considered, depending on the distribution of model residuals.

If there is any indication of a strong departure from Normality for the residual plots for the continuous outcome models, a p-value and 95% confidence interval (CI) will be calculated using the bootstrap method from 1000 bootstrap samples.

Secondary analyses

All initial models listed above will include the main effect of study condition only. As secondary analyses, pre-specified covariates of baseline dairy milk consumption and baseline milk substitute consumption will be added to each model. For this secondary analysis, the two-way interactions between study condition and each covariate will be added to the model. Unless each interaction term attains evidence of a high degree of statistical significance (which we define as $p < 0.01$), it will be dropped in favour of a model which includes only the three main effects. If there is clear evidence of a 2-way interaction (which we define as $p < 0.01$), we will use the model, including the main effects to plot interaction terms. A p-value will be reported for the interaction term.

Sensitivity analyses

For sensitivity analyses, the primary outcome analyses will be repeated four times: 1. Excluding participants that were deemed to guess the true nature of the study, 2. Excluding those that reported being on a restricted diet (diary-free or vegan), 3. Excluding any participant that said yes to adding an 'other drink' (i.e., a fruit juice) to their tea or coffee', and 4. Excluding anyone who answered zero to the 'number of times the participants consume dairy milk in an average week'.