behaviour change by design

Generating and Implementing Evidence to Improve Health for All

STATISTICAL ANALYSIS PLAN

Cigarette Pack Size and Consumption: a randomised crossover trial

Study design

A randomised crossover design in which participants will smoke from a single pack size (i. 20, ii. 25) throughout two two-week periods.

Primary research question

Does reducing cigarette pack sizes from 25 to 20 reduce average consumption by 1.5 cigarettes per day?

Study hypothesis

Reducing cigarette pack sizes from 25 to 20 reduces average consumption by 1.5 cigarettes per day.

Outcomes

Primary outcome

The primary outcome is the average (mean) number of cigarettes smoked per day during each intervention period.

This will be calculated by adding up all of the cigarettes smoked within each intervention period and dividing by 14 to obtain a measure of daily consumption.

The number of cigarettes smoked will be calculated from photographs participants take of their empty cigarette packs labelled with stickers with the following information:

i. Date pack finished

ii. Number of cigarettes smoked by participant from this pack (excluding those given away or not consumed by the participant for any other reason)

iii. Number of cigarettes smoked by participant not from that pack while that pack has been open (e.g. given to them by a friend)

iv. Rating of warning message (to align with the study cover story): “The warning on this pack is effective” Scale range: 0, No to 7, Yes.

The total number of cigarettes smoked in each intervention period is calculated by summing i. the number of cigarettes participants smoked from each pack for which they returned

photographs (variable I above) and ii. the number of cigarettes participants reported smoking from packs owned by someone else (variable ii. above).

There is an additional sticker for participants to stick to any packs that they are partway through at the end of an intervention period to indicate how many cigarettes remain in the pack.

Secondary outcome
The secondary outcome is motivation to stop smoking. This will be measured by the single-item 'Motivation to Stop Scale' (MTSS; Kotz et al 2013) with responses to the question: Which of the following describes you? Responses range from (1) I don't want to stop smoking to (7) I REALLY want to stop smoking and intend to in the next month. Participants answer this question at the end of each intervention period.

Demographic characteristics
The following demographic variables were recorded for each participant:

1. Age
2. Annual household income
3. Highest level of formal education
4. Perceived income
5. Sex
6. Aboriginal status
7. Racial and cultural group

Baseline measures
The following measures of smoking behaviour and attitudes are assessed at enrolment via an online questionnaire, once participants have consented to take part in the study.

i. Heaviness of smoking: this will be assessed using the Heaviness of Smoking Index (HSI) assessing cigarettes smoked per day and time to first cigarette (Heatherton 1989).

ii. Motivation to stop smoking (see above).

iii. Self-reported number of cigarettes smoked per day at enrolment

End of intervention period surveys
We collected additional information about participants' smoking and cigarette purchasing behaviour during each intervention period (Please refer to registered protocol for questions: https://osf.io/7428u/). This information will be described and tabulated using a mixture of content analysis, inductive and deductive qualitative analyses.

Data collection
A clean data set will be sent directly to the researcher completing the data analysis at the end of the study in an Excel spreadsheet. A data dictionary will also be sent which includes all coding and ranges. A strategy for ensuring the fidelity of the data (double-entry) has been planned.

Outliers
Any outliers will be identified using range checks, scatter plots and histograms. If any outliers are identified (defined by any values that differs from the median by more than 3 using median absolute deviation) further checks will be performed by the research team to ensure they are not the result of data entry errors.

Any outliers (as defined above) will be included in the primary 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 (for example, if a participant only smoked 10 packs of cigarettes, cigarette pack number 11 would be not applicable for this participant). Data that are missing due to an administrative or other error will be coded as 999.

Primary outcome measure
The primary outcome (the average (mean) number of cigarettes smoked per day during each intervention period) is calculated using information participants provide on printed stickers and attach to each cigarette pack they smoke during each intervention period, as described above.

A cigarette pack will be classed as "Primary-missing" if responses to either of the following two items are missing on the sticker:

i) How many of these cigarettes did you smoke yourself?

ii) How many cigarettes from other packs did you smoke while this pack was open?

If only one pack in any one intervention period is classified as "Primary-missing", imputation will be used to replace the missing value by the mean value from the other packs for the participant in this period (for variables i & ii separately). These imputed consumption values will be included in the primary analysis.

If more than one pack for a participant in any one intervention period is classified as "Primary-missing", consumption values will not be imputed for those missing packs in the primary analysis.

The strategy for imputing missing values for all of the possible fields on each sticker is described in Appendix A 'Imputations for missing data'.

Secondary outcome measures
All missing data for the secondary outcomes will be analysed on a complete case basis as missing data should be minimal for these variables due to the nature of the study.

Missing data checks
If more than 10% of data are missing for any single primary or secondary outcome variable, a Table will be produced for that variable comparing participants' characteristics between study arms for participants who have, and those who have not provided outcome data (e.g. missing [Y/N] vs gender [male/female]) – see Groenwold, Moons & Vandenbroucke, 2014.

Violations of normality
It is expected that the primary outcome would follow a Normal distribution in the underlying population, and parametric analyses will be carried out accordingly. Nevertheless, whether the study data appears to follow a Normal distribution or not, will be assessed using normality plots on residuals from the final ANCOVA/regression models specified below.

If there is any indication of a strong departure from Normality for any of the residuals then alternative distributions may also be considered, and the p-value and 95% confidence

interval will be calculated using the bootstrap method (using 1000 bootstrap samples) with bias correction.

Descriptive statistics

A CONSORT flow chart will be constructed to show the numbers of individuals assessed for eligibility, recruited for baseline study, randomised and followed up.

A table will compare demographic characteristics of participants allocated to the two possible treatment orders: means and SDs will be shown for continuous variables, with numbers and percentages within each category of nominal or ordered categorical variables.

Outcome analysis

All analysis will be done in IBM SPSS version 27 or similar. Analysis will be coded in syntax and this will be added as Appendix B after the analysis is complete.

Primary outcome analysis

1. Primary analysis

Intention-to-treat analysis will be reported as the primary analysis including all data from randomised participants, analysed as randomised.

This is particularly important for the reporting of randomised cross-over studies to avoid the introduction of bias caused by non-random attrition of participants after randomisation (https://bmjopen.bmj.com/content/3/11/e003464.long).

Mixed effects models will be used (or similar depending on model diagnostics) based on the observed data with minimal imputation and assumptions.

This class of models uses all available explanatory variable data even if some are missing. However, such models cannot include data when the primary outcome is also missing.

See the 'Missing data' section for a description of the imputations that will be made in the dataset for the primary analysis.

Comparisons of the primary outcome will be made by estimating the mean difference according to pack size condition, with 95% confidence interval and p-value obtained from a generalized linear model of the primary outcome. This will involve a repeated measures analysis including terms for the treatment effect, period effect and order effect. Evidence for a treatment x order interaction will be examined, but this is not expected to be required in the final model given the use of a washout period.

The following covariates will be considered for inclusion in the model depending on modelling diagnostics (i.e. multi-collinearity, overfitting, numerical convergence stability):

1. Self-reported number of cigarettes smoked per day at enrolment
2. Heaviness of Smoking Index, as measured at enrolment
3. Motivation to stop smoking, as measured at enrolment
4. Price per cigarette
5. Wash-out time
6. Non-study cigarettes smoked

2. Sensitivity analysis 1: No imputation

The data will be analysed without any imputation for missing values (and without any of the assumptions required for this).

3. Sensitivity analysis 2: Imputation strategy 2
This will be the same as the primary analysis, but information for up to two packs classed as 'Primary-missing' will be imputed.

4. Sensitivity analysis 3: Per protocol analysis
The primary analysis (i.e. with up to one pack imputed) will be repeated using the per-protocol dataset. The per-protocol dataset excludes participants not meeting the criteria for adherence to the protocol as defined in the 'Non-adherence' section (below).

5. Sensitivity analysis 4: Self-reported mitigating factors
The primary analysis (i.e. with up to one pack imputed) will be repeated with an additional covariate (coded from -1 to 1) for self-reported mitigating factors for each intervention period impacting on cigarette consumption.

Secondary outcomes
The secondary outcome, motivation to stop smoking, will be analysed in the same way as the primary outcome.

Checking assumptions
Residuals from models before and after adjustment for baseline variables will be assessed using a normal probability plot.

Subgroup analysis
The study is unlikely to have sufficient power to detect any modest differences between demographic groups. However our earlier study (Lee et al, in press), suggested that heavier smoking may be associated with greater reductions in cigarette consumption as a result of the intervention. We will investigate this by testing for an interaction between "treatment" and baseline heaviness of smoking or some suitable transformation.

Non-adherence
We will repeat key analyses for the primary outcome, excluding participants deemed to have been non-adherent to the intervention and study instructions based on the photos of cigarette packs they send during the intervention periods. Non-adherence will be operationalised as:

1) participants smoking from the pack size to which they had not been randomly allocated to smoke from in that intervention period

2) participants using a different brand variant of cigarettes (including cigarette length) than that which had been agreed to prior to randomisation

3) participants deviating significantly from study instructions as judged by the investigators

Participants will be deemed to be adherent if at least 90% of the cigarette packs from which they have smoked during each of the two intervention periods do not meet criteria 1 and 2 for non-adherence defined above and they do not meet criterion 3.

The pre-registered protocol defined adherence only in relation to the pack size of cigarettes that participants smoked in each intervention period. Participants were however also instructed as per protocol to pre-select a cigarette brand variant and cigarette length (available in both pack sizes) to be smoked in each intervention period.