Pre-analysis plan: Testing incentives to encourage asbestos removal

This pre-analysis plan includes mathematical notation that may not be fully accessible for those using assistive technology. Please email BETA@pmc.gov.au if you require assistance engaging with this material.

Policy problem

As part of the Phase 3: Asbestos National Strategic Plan 2024-2030, the Asbestos and Silica Safety and Eradication Agency (ASSEA) aims to encourage the safe removal of asbestos from Australian residences.

For residential buildings, the cost of asbestos removal, disposal and replacement is the main impediment to removal (Ipsos 2018). Hypothetical government initiatives that reduced this cost for homeowners (such as subsidised removal or disposal, low or no interest loans, and tax concessions) have been associated with higher reported willingness to remove asbestos (Ipsos 2018).

Trial aims

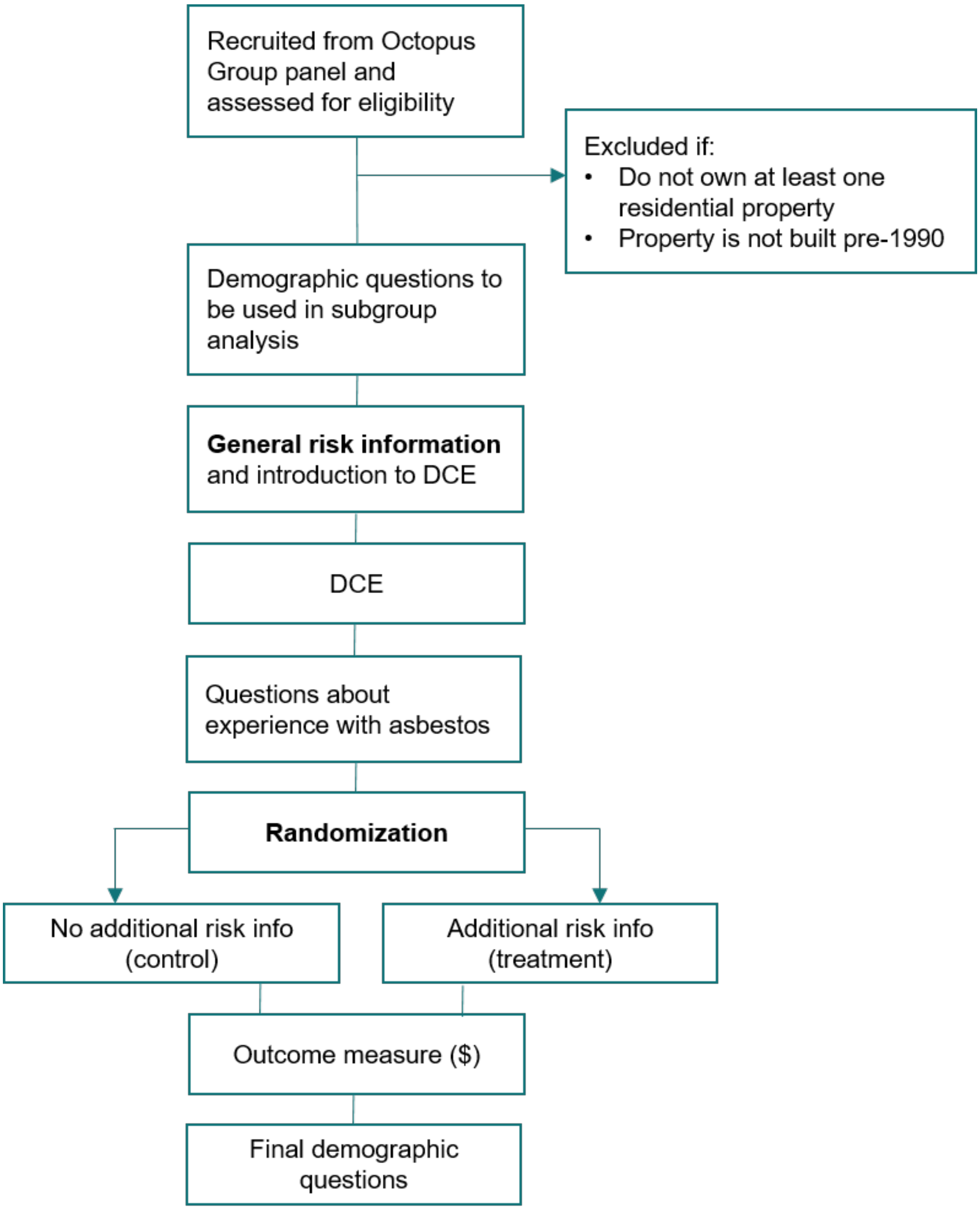
ASSEA has commissioned BETA to help identify how best to optimise any incentives offered to homeowners to increase their willingness to pay for asbestos removal, increasing the likelihood that they will act to have asbestos safely removed from their owner-occupied and investment properties. The aim of this trial is to:

- Estimate the impact of different baseline asbestos removal costs without incentives, and different *incentive types (and levels)* on participants' willingness to remove asbestos (Discrete Choice Experiment, DCE)
- Estimate the impact that information about *the increasing risk of asbestos* (due to its increasing age) can have on the amount participants are willing to pay to remove asbestos (Randomised Controlled Trial, RCT)

Trial design

Figure 1 below illustrates the trial design, including sequencing of initial risk information, the DCE, and the additional risk information (as part of the RCT).

Figure 1. CONSORT diagram illustrating trial design



Population and sample selection

The population of interest is adults in Australia who own at least one residential property built before 1990. We will monitor a number of characteristics to ensure the sample includes good coverage of the characteristics of interest, including: owner-occupiers vs landlords, those with mortgages vs own outright, those who live in regional/remote areas vs those who live in urban centres, those from low vs high income brackets. We will also monitor age, gender, state or territory of residence, and culturally and linguistically diverse (CALD) status. For this project a participant will be classified as CALD if they either mainly speak a language other than English at home or were born overseas (or both).

The target sample size for the overall study is 4,500. The sample will be recruited by Octopus Surveys, from their panel of participants. The study will be conducted online. We have planned some subgroup analyses as outlined below.

Screening and replacement participants

We will start the survey by asking respondents a few questions to check their eligibility (see Consort diagram, Figure 1). To participate in the study, respondents must own one or more residential properties that were built prior to 1990 (and be over 18 years old and live in Australia). Buildings constructed pre-1990 are very likely to contain asbestos, and this cohort is therefore most relevant to ASSEA.

After data collection, we will check the quality of the data. Any *duplicate* nonsense open-ended responses will be excluded from the data set (as they are likely to be bots) and replaced by the recruitment provider. Otherwise, we do not have any automatic exclusions planned.

However, we will conduct two robustness checks. Our main analysis will be on the data set minus *duplicate* nonsense responses, as described above. Our first robustness check will be with the *full* sample (i.e. including duplicate nonsense responses, even though they are likely to be bots).

For the second robustness check, we will examine responses that receive a score less than one on Qualtrics' bot score, or who provide nonsense/suspicious open-ended responses, or who are very fast (less than 1/3 of the median response time). While these responses will not be excluded automatically, we will scrutinise their responses more closely and exclude them if there are multiple indicators that they are non-serious/nonsense responses. This sample – with *possible* bots removed – will be our second robustness check.

Interventions and randomisation

The DCE component of this trial is randomised at the level of attributes and options. We have specified the following:

- each participant responds to seven choice sets (the second and last will be the same to calculate intra-respondent reliability, see further details below)
- each choice set contains two options (and an 'I would not remove the asbestos' option)
- there are five attributes for each option
  - loan
  - grant
  - tax offset
  - lottery
  - quoted cost
- the levels for each attribute are specified in Table 1.

We have calculated a D-efficient experimental design based on these specifications, using Qualtrics' "Conjoint Analysis" feature. We will implement a conditional constraint on the grant and quoted cost attributes to ensure the grant amount never exceeds the quoted cost. While this may affect perfect design balance, it maintains scenario realism and enhances response validity. While the remaining combinations vary in how ‘realistic’ they are, we don’t anticipate that this will create problems for the analyses (and we have planned a number of validity checks, see below).

Table 1: Attributes and levels used in the present study

Attribute	Number of levels	Value of levels
Quoted cost	4	$5,000; $10,000; $20,000; $30,000
Grant	4	None; $5,000; $10,000; $15,000
Interest-free loan	4	None; 5 year; 10 year; 15 year
Tax offset	2	Not available; available
Lottery	2	Not available; available

The RCT component of this trial is an individually randomised online experiment. Participants will be randomised to 1 of 2 arms (no additional risk information vs additional risk information). Randomisation will be done by Qualtrics (the survey software), by giving each participant a 1/2 probability of being assigned to each trial arm.

Outcome measures

Primary outcome measures

The primary outcome measure for the **DCE component** is the option each individual chooses, in each of six (plus one repeated) choice sets. Each choice set contains two options with varying incentives and cost for removing asbestos. Each decision is coded as ‘1’ if option was chosen, and as ‘0’ if it was not chosen. If a participant chooses ‘I would not remove’, both options in that choice set will be coded as 0. Using this outcome, we will calculate the influence each attribute (incentives, cost) has on people’s choices.

For the **RCT component,** the primary outcome measure is the price (in dollars) participants say they could afford to pay if they had to remove asbestos. At the individual level this is asked as a single question, and participants enter a dollar value in response. At a group level this will be the mean dollar value in each group.

- If a participant skips this question, their response will be coded as ‘missing’.
- If they enter $0, this will count as a legitimate response.

Hypotheses

For the RCT, we hypothesise that people who do not receive additional risk information will indicate a different maximum price (in dollars) than those who receive additional risk information (B ≠ A two-tailed test).

We do not have specific confirmatory hypotheses for the DCE.

Power calculations

Here, we present power calculations for the RCT component of the research project only. Due to resource constraints, our sample is fixed at around 4,500 individuals, which will provide 2,000 individuals per group (conservatively, after exclusions). For this study, alpha is set to 0.05, and hypothesis tests will be two-sided.

ASSEA indicated the smallest effect size of interest would be around $500.

Initially, we powered our study assuming a SD of $2000, based on expected variability in the outcome measure (max price). This allowed us to calculate sample sizes with 90% power, targeting a $350 difference between groups (Cohen's d = 0.175). Given these assumptions, our original sample size of 4000 participants was sufficient, and we planned to conduct sensitivity analyses with smaller subsamples (e.g., 2000 per participants) to examine whether the effect remains practically significant and robust under more typical, not overpowered study conditions.

However, our pilot study revealed a much larger SD in the outcome variable than initially assumed (SD = $28,599), which significantly alters our power calculations. With this variability incorporated, we calculated the design has 80% power to detect an effect size of $1740 (Cohen's d = 0.061).

Method of analysis: RCT

The principal analysis of the effect of the intervention will consist of a covariate-adjusted comparison of our primary outcome. This estimate, confidence intervals and p-values will be derived from a linear regression model using robust (HC2) standard errors and with the following specification:

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 X_i + \beta_3 Z_i X_i + \epsilon_i$$

Where *i* is an index for each individual in the trial, *Y* is the individual's dollar response, β_0 is the intercept, *Z* is treatment assignment indicator, β_1 is the coefficient representing the average treatment effect for the intervention relative to control, *X* is a mean-centred covariate (see Covariates section below), *ZX* is the interaction of the treatment indicator with the mean-centred covariate indicator, and ε is the individual error term.

Covariates

We will include participants' household income to the model as a baseline dummy-coded, mean-centred covariate. It will be recoded into low, medium, and high with low as the reference group. The question asks for household income in 8 brackets: <$30,000; 30,000 - 50,000; 50,001 - 70,000; 70,001 - 100,000; 100,001 - 130,000; 130,001 - 150,000; 150,001 - 200,000; over 200,001. We will recode incomes below 70,001 as low, incomes between 70,001 and 130,000 as medium and those 130,001 and above as high.

Robustness checks

We will check for robustness by winsorising the outcome variable to reduce the influence of extreme outliers. This approach involves capping the most extreme values (99th percentile), allowing us to assess whether the main analysis results remain consistent when the variability caused by outliers is minimised.

Missing data approach

Since participants can skip the outcome question, there is the possibility of differential attrition. To address this, we will take a multi-stage approach. At the end of data collection, we will examine the rate of missingness in our primary outcome variables. If this is below 10% in both arms we will conduct a complete case analysis. If the rate of missingness is above 10% in at least one arm, we will examine the pattern of missingness by first checking if there is differential attrition. If there is a significant difference in rates of missingness between arms, we will conduct Lee bounds and present both complete case analysis and Lee bounds (Lee 2002), otherwise we will conduct a complete case analysis.

Exploratory analyses

We plan to treat the outcome as a continuous measure, however depending on the distribution of responses we may treat it as a proportion (0 vs any amount). These analyses will be clearly marked as exploratory.

Method of analysis: DCE

Due to the complexity of the design and uncertainty about how participants will respond, we have made the following high-level plan for the analysis of the DCE (see further details about each step below):

1. Fit the 'full' model first (all levels of all incentives, random slopes except on cost).
2. If it fails to converge or we have problems with sparseness, fit a simplified model (all levels of all incentives, fixed effects only).
3. If there are still problems, recode all incentives to binary (no incentive vs any incentive, fixed effects only).
4. If Step 1 worked fine, we will also recode all incentives to binary as in Step 3 but use random slopes (except on cost). This is because we'd like an 'average' assessment of the impact of each incentive (vs no incentive) as well. It will also allow us to see how much individual participants vary in their sensitivity to the presence of the incentives (we will quantify this variability through the SD of the random slopes).

Step 1 model

Choice of option (0 vs 1) will be regressed on the following indicators:

- a continuous variable for quoted cost, fixed effect
- a binary indicator for lottery (0 = no lottery, 1 = lottery), random effect
- a continuous variable for grant, random effect
- an indicator for loan, dummy-coded (0 = no loan, three levels of loan), random effect
- a binary indicator for tax incentive (0 = no tax incentive, 1 = tax incentive), random effect

The random effects will be coded as uncorrelated. To aid the mixed effect model in convergence we will normalise the cost and grant variables by the maximum value for cost (so that they are on the same scale). We do not centre the variables as when we rescale the grant and cost variables they are bounded between zero and 1. Thus we set:

$$X_i^{norm} = \frac{X_i}{P_i^{\max}}$$

Where *X* (subscript *i*) is the value for either the cost or grant attribute and *Pmax* is the maximum value for the cost attribute.

We will back-convert the normalised and centred value for reporting.

Step 2 model

As Step 1, but all fixed effects.

Step 3 model

Choice of option (0 vs 1) will be regressed on the following indicators (all fixed effects):

- a continuous variable for quoted cost – normalised and mean centred as above
- a binary indicator for lottery (0 = no lottery, 1 = lottery)
- a binary indicator for grant (0 = no grant, 1 = any grant) - recoded
- a binary indicator for loan (0 = no loan, 1 = any loan) - recoded
- a binary indicator for tax incentive (0 = no tax incentive, 1 = tax incentive)

Intra-respondent reliability correction

For all these models, we will also calculate and correct for the intra-respondent reliability (IRR) to control for potential measurement error. The IRR will be calculated by comparing participants' responses to choice set 2 and a repeated choice set 7, which will be identical except for the reversal of the left and right positions of the alternatives. This approach will enable us to measure the consistency of participants' choices. In this study, we follow the approach outlined by Clayton et al. (2023).

The IRR will be calculated as the proportion of consistent responses between the original and repeated choice sets. The IRR will then be used to adjust the observed Marginal Means (MM) and Average Marginal Component Effects (AMCE). We will do this by applying the following correction formula:

$$MM_{corrected} = \frac{MM_{observed\,-\tau}}{1-2\tau}$$

where τ is the swapping error rate derived from the IRR. The swapping error rate refers to the probability that a participant's response to a choice set does not reflect their true preference but is instead randomly "swapped" or chosen incorrectly.

$$AMCE_{corrected} = \frac{AMCE_{observed\,-\tau}}{1-2\tau}$$

To assess the impact of these corrections, we will calculate the Root Mean Square Error (RMSE) for both observed and corrected MM and AMCE values.

Subgroup analyses

We will fit the DCE model to a number of subsets of the data, provided each subset is larger than n = 500 (after excluding 'would not remove' responses). These subsets will be:

- Owner-occupiers vs people who are renting out their property

- People whose property is mortgaged vs owned outright
- People in different states (combining ACT and NSW and excluding smaller jurisdictions)
- People in different income brackets (lower vs higher – the exact cut-off to be determined based on sample size)
- Urban/suburban vs regional/remote
- Strata/community owned vs non-strata

For each subgroup, we will calculate IRR separately to account for potential measurement error across these groups.

Interpretation and reporting

For the RCT component, although we will use p-values, with a pre-registered rejection threshold to test our hypotheses, we will consider the outcome of our hypothesis tests alongside prior evidence, effect size, outcome variability, exploratory analyses and design limitations in order to assess the strength of a finding and our recommendations.

For the DCE component, we will primarily rely on the coefficients and standard errors for each incentive in order to assess the influence of each incentive on willingness to remove asbestos.

Pilot study

We conducted a pilot study to help determine the details of the design and analysis. The pilot study included about 200 participants each responding to 6 unique choice sets, totalling 1200 observations. The purpose of the pilot was to test that the DCE was giving us expected results, and that participants were correctly interpreting our questions. We also used the pilot to estimate variability in the outcome measure for the RCT to adjust the power calculations.

DCE

We made no changes to the DCE as a result of the pilot study.

Survey questions

No changes were made to the survey questions as a result of the pilot study.

RCT

The pilot study revealed a much larger SD in the outcome variable than initially assumed (SD = $28,599.28), which significantly altered our power calculations and approach (see above).

The pilot data will not be included in the RCT analysis.

Pre-analysis plan commitments

No analysis has been undertaken prior to the completion of this pre-analysis plan. We will be transparent about, and provide justification for, any deviations (additions or omissions) from this plan. We updated this pre-analysis plan after the pilot study to explain the changes that we made.

References

Clayton, K, Horiuchi, Y, Kaufman, AR, King, G and Komisarchik, M (2023) Correcting measurement error bias in conjoint survey experiments. Harvard University Working Paper.

Ipsos (2018) *Barriers, motivations and options for increasing asbestos removal in the residential and commercial sectors*. Ipsos Social Research Institute.

Lee, DS (2002) Trimming for bounds on treatment effects with missing outcomes. Centre for Labor Economics, Working Paper 51.