Pre-analysis plan: Cyber Security Labels

Key dates

Pre-registration on the AEA RCT registry:	13 August 2021
Trial launch:	27 July 2021
Trial closed:	11 August 2021

Policy problem

Smart devices, sometimes referred to as consumer Internet of Things (IoT) devices, are products with extra functionality to connect to the internet. Research from around the world has shown that many smart devices currently lack basic cyber security features. The introduction of cyber security labels may help consumers choose products that are more cyber secure, and increase consumer awareness about the risks of insecure smart devices. The Department of Home Affairs (HA) has asked BETA to build on existing international evidence and design and evaluate the effectiveness of cyber security labels in the Australian context.

Trial aim

The aim of this project is to examine the extent to which Australian consumers will be guided by cyber security labels when purchasing smart devices, and to examine which type of label is the most impactful. We also aim to determine 'willingness to pay' for devices with higher cyber security ratings.

The research will be a combined online randomised controlled trial (RCT) and discrete choice experiment (DCE) – see details below.

Interventions

The intervention is a cyber security label designed by BETA and HA, modelled on existing international examples. There are three different versions of the label. Two communicate the 'expiry date' beyond which cyber security updates for the device are no longer guaranteed – one of these is a plain text label, the other includes an icon. The third label is a 'graded' label, where each level of cyber security is indicated by an increasing number of shields. All labels have four levels (see Table 1 and Figure 1). For the purposes of this trial, *any* label should be preferable to no

label. That is, even the lowest level (Level 1) should provide participants with some guidance about a product's cyber security in the trial, relative to the absence of a label.

We will evaluate the effect of the labels on purchasing decisions through a combined RCT and DCE.

Experimental design

Overview

All participants will "shop" online for a smart device. Participants will be randomly assigned to shop for *one* of four product categories – smart light bulb, smart watch, home hub or smart TV. When participants "shop" for the smart devices, they will be shown three devices (e.g., three TVs, three watches) in a choice set. The devices will vary in price, features, and level of cyber security. Participants will be asked to choose the one they prefer, making trade-offs between the devices' attributes. (They can also respond 'none of these'.) Then they will be shown a new choice set, with three different devices. They will repeat this task ten times.

The RCT

Some of the devices will be displayed with a cyber security label. Participants will be randomly assigned to see *one* of the three cyber security labels. The type of label is a between-subjects variable (see Figure 1 and Table 1). The RCT component of the trial will allow us to test which type of label has the biggest impact on purchasing 'decisions'. (We will aggregate across the four product categories for our primary analyses.)

The DCE

Through having participants make repeated choices between many different combinations of prices, features, and cyber security levels, we will be able to determine which attributes have the greatest impact on purchasing decisions (including the effect of a label vs no label), and how participants are willing to 'trade off' the cyber security rating of a device against its price and features (see Method of Analysis section).

Figure 1: Overview of study design and randomisation

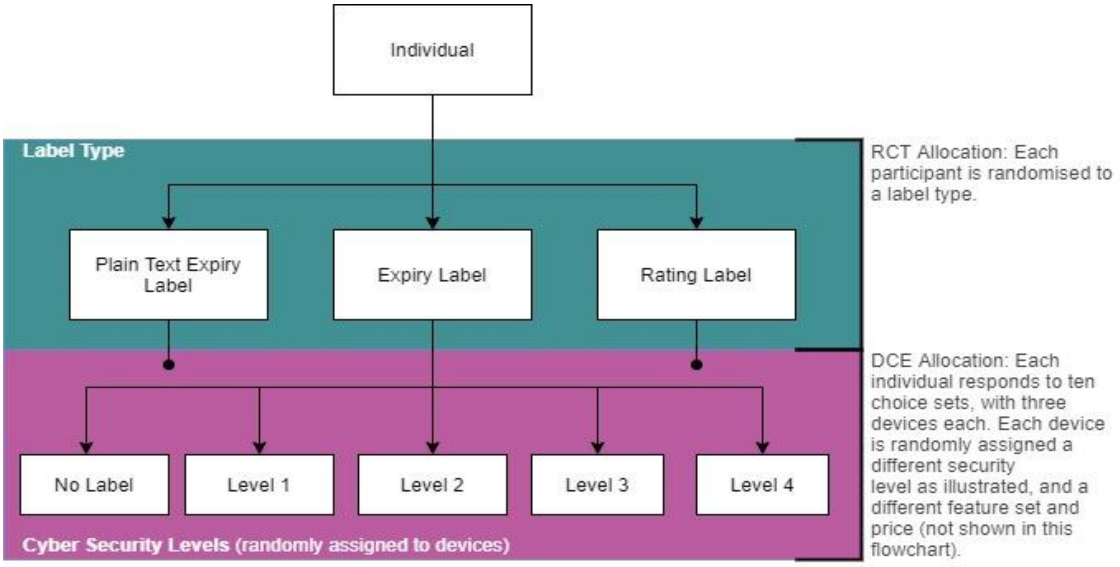

Table 1: Overview of study design.

		Within-subjects, DCE component
		Security levels traded off against price and features
Between-subjects RCT allocation	Group A: plain text	No label
		June 2021
		February 2022
		August 2023
		August 2026
	Group B: expiry label	No label
		June 2021
		February 2022
		August 2023
		August 2026
	Group C: graded label	No label
		Level 1: Baseline
		Level 2: Intermediate
		Level 3: Enhanced
		Level 4: Hardened

Note: In addition to the between-subjects allocation of three different label types, we will randomly allocate participants to 'shop' for one of four different product types (TV, home hub, watch, light bulbs) – creating a total of twelve between-subjects groups. As we plan to pool the data from the four product types for our primary analyses, this factor is not displayed in Figure 1/Table 1.

Outcome measures

Primary outcome measures

The primary outcome measure for **the RCT component** is whether, for each choice set, an individual chose to 'buy' a device *with* a cyber security label (coded as '1') or one *without* a label (coded as '0'). We will calculate sample proportions from this binary measure.

The primary outcome measure for **the DCE component** is the device each individual chooses to 'buy', in each of ten choice sets (each choice set contains three variations on a single device – e.g., watch or TV). Each device is coded as '1' if it was purchased, and as '0' if it was not purchased. We will calculate sample proportions from this binary measure.

Secondary outcome measures

The secondary outcome measure for **the RCT component** is, for each choice set, the *level of cyber security* of the device the individual chose to 'buy'. The level will be treated as a continuous variable from 0 (no label) to 4 (Level 4 label). That is, if an individual chooses a device with no label, they will get a 'score' of 0 for that choice set. If they choose a device with a Level 1 label, they will get a 'score' of 1, and so on.

For **the DCE component** we will also calculate 'willingness to pay' for cyber security labels (see Method of Analysis section).

Hypotheses

Primary hypotheses

Randomised controlled trial

H1a: People in the icon expiry label group will choose a greater proportion of devices with labels than people in the expiry plain text group (B > A, one-tailed test).

H1b: People in the graded shield label group will choose a greater proportion of devices with labels than will people in the icon expiry label group (C > B, one-tailed test).

For H1 we will pool the data from the four product categories.

Discrete choice experiment

H2: The presence of a cyber security label (versus no cyber security label) will increase people's likelihood of purchasing a given device (one-tailed test).

We will conduct this (conjoint) analysis separately for the three label types (A, B, C), pooling the four product categories.

Secondary hypotheses

Our secondary hypothesis SH3 is the same as H1, but using the (secondary) continuous rather than the binary (primary) outcome measure.

SH3a: People in the icon expiry label group will choose a higher level of cyber security than people in the expiry plain text group (B > A, one-tailed test).

SH3b: People in the graded shield label group will choose a higher level of cyber security than will people in the icon expiry label group (C > B, one-tailed test).

For SH3, we will pool the data from the four product categories.

SH4: Labels indicating higher cyber security ratings will have a greater impact than those with lower security ratings. For this analysis we will pool the data from the four devices, and conduct a (conjoint) analysis separately for the three label types (A, B, C), as for H2.

For all hypotheses we will undertake the same analyses separately for the four product types as well, as exploratory follow-up analyses.

Population and sample selection

Our population of interest is the general population of Australian adults. Participants will be 6,000 people recruited by Dynata from their participant pool, and will be over 18 years old. They will otherwise be representative of the Australian population on gender and age (three bands). We will also aim for location (state) statistics to match the general population (soft quotas) or as close as possible given the constraints of Dynata's panel and our time frames. Dynata will require that participants complete the study on non-mobile devices. We do not have any further exclusion criteria.

Randomisation

The RCT component of this trial is an individually randomised online experiment, with repeated measures. Participants will be randomised to 1 of 3 cells (corresponding to three different label types, see Figure 1). Randomisation will be done by Qualtrics (the survey software), by giving each participant a 1/3 probability of being assigned to each trial arm. We will use an option that Qualtrics provides, to prevent cell sizes becoming too uneven.

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

- each participant responds to ten choice sets;
- each choice set contains three devices;

- there are three attributes for each device
  - *features:* standard vs premium;
  - *price:* five levels, starting at average price and increasing by 20% at each level;
  - *label:* five levels, as specified in Figure 1 and Table 1

Qualtrics then calculates a D-efficient (orthogonal and balanced) experimental design on the basis of these specifications. This means that participants see a "random" set of devices, with a "random" set of attributes, and the net effect is that we gain the maximum amount of information about the influence of each attribute on participants' choices.

Sample size and power calculations

Here, we present power calculations for the RCT component of the research project only. Due to resource and timing constraints, our sample is fixed at around 6,000 individuals, which will provide 2,000 individuals per group. Each individual will respond to 10 choice sets. To account for repeated measures, we will cluster our standard errors by individual. Individual preference for selecting a labelled device over a non-labelled device is likely to be highly correlated across an individual's 10 choice sets. For the sake of these power calculations, we assume an ICC of 1. This is very conservative with any reduction in this correlation reducing the minimum detectable effect at a given power level/sample size. For this study, alpha is set to 0.05, and beta to 0.2, and hypothesis tests will be one-sided.

With these assumptions in place, for H1a and H1b, which both relate to the RCT component, we estimate that our design can detect a standardised effect of 0.08 (Cohen's *h*) with 80% power, this corresponds to approximately a 4 percentage point increase from a conservative 50% baseline.

Method of analysis

For H1, which stems from the RCT component of this research, we will fit an OLS regression with cluster-robust standard errors. In this case, choice of a device *with* vs *without* a label (1 vs 0) will be regressed on a binary treatment indicator (type of label coded depending on which comparison we are making). There will be no covariates. We will fit this model only to the subset of the data that is relevant to the comparison we are making (A & B for H1a, B & C for H1b). We will pool product categories. For SH3, we will fit the same models but using the continuous secondary outcome measure.

For H2, which stems from the DCE component of this research, we will fit a mixed-effects linear regression. Choice of device (0 vs 1) will be regressed on a binary treatment indicator (0 = no label, 1 = any label), a binary indicator for feature level (0 = standard, 1 = premium), and a continuous variable for price, mean-centred. We

will specify random slopes for treatment (label) and feature level, by individual. The random slopes will be modelled as uncorrelated. We will fit this model separately for each label type (A, B, C), but pool product categories. For SH4, we will fit the same models but use dummy codes for each level of the labels.

Trial threats

Missing data

We expect that some individuals will drop out of the DCE before completing it. This will result in missing data. This data will be missing at random relative to treatment assignment and we will exclude these records from our analysis.

The main potential threat to internal validity in the RCT component is missingness related to treatment assignment. This will only be a concern if an individual's propensity to drop out of the survey is related to the label variation they are randomly assigned to. We don't expect this to occur to an extent that would impact our estimates, however, we will assess our dataset for this issue. If we find evidence for this we will report it and take it into account in our interpretation.

There will also be some missing data for the conjoint analysis (H2 and SH4), when participants select 'none of these' in a given choice set (rather than selecting one of the three devices). This data will be missing at random, and will be excluded from the conjoint analyses. For the RCT component (H1 and SH3), we will code these responses as '0', and include them in the analyses.

Blinding

Individuals will be aware that they are participating in a research project, however, they will be unaware that they have been randomised to see a particular label over another. Random assignment will occur on an online platform as individuals enrol and we will not have the ability to influence this procedure. Treatment assignments will be visible to our researchers in the final dataset.

Interpretation of results

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 and design limitations in order to assess the strength of a finding and our recommendations.

At the end of the study participants will complete a number of subjective survey questions, which we will explore to help understand purchasing decisions and participants' understanding of the cyber security labels.

Pre-analysis plan commitments

We conducted data quality checks and tested our conjoint models for the DCE component (H2 and SH4) when we had recruited the first 2,000 participants in our sample. We also tested our analysis for the RCT component, on the same data set, but after we had altered it by manually scrambling the condition assignment and outcome measure. No further analysis has been undertaken on the RCT component 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.